Exploring Methodologies for Utilizing Click-Track Data Using Educational Data Mining and Evidence Centered Design in Online Professional Development Environments

---

A dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Caner Uguz

August, 2016

Abstract

The opportunities for delivering effective PD has dramatically increased with the recent growth of web application capabilities. However, our methodologies for making meaningful inferences about these digital learning environments have remained limited. This dissertation argues that valid measurements of research constructs in online PD environments require three components to be successful: 1) a methodology for reliably recording user behavior, 2) an assessment framework for building the connection between constructs and observable behavior and 3) a statistical analysis approach that provides post-hoc methodologies for recognizing patterns in the observable behavior. The first manuscript in this dissertation conducts a review of the literature on methods used in a sample of existing studies and suggests that a combination of click-track data as the methodology for recording user behavior, Evidence Centered Design (ECD) as the assessment framework and Educational Data Mining (EDM) as the statistical analysis approach has the potential to provide insights in online PD environments. The second manuscript uses EDM methodologies to investigate construct assumptions and user behavior patterns through click-track data. The third manuscript uses a combination of ECD and EDM methodologies to build a measure for evaluating teacher engagement in an online PD environment. The dissertation provides case studies for the use of these combined methodologies, which show promise as a viable strategy for researching and understanding online PD environments. Insights and limitations of using click-track data

and directions for further research Domain Modeling For The Main Aspects of the Domain Analysis are also discussed.

Curriculum, Instruction & Special Education
Curry School of Education
University of Virginia
Charlottesville, Virginia

APPROVAL OF THE DISSERTATION

This dissertation, ("Exploring Methodologies for Utilizing Click-Track Data Using Educational Data Mining and Evidence Centered Design in Online Professional Development Environments"), has been approved by the Graduate Faculty of the Curry School of Education in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____
Sara Dexter, Chair

____
Jennie Chiu

____
Patrick Meyer

____
Joe Garofalo

______________________ Date

Dedicated to my son Rémi William Shaughnessy,
the most beautiful motivation to finish a dissertation anyone could ask for.

Acknowledgement

This dissertation could not have been possible without the significant contributions of many fellow researchers. Work carried out by previous researchers with strong commitment to furthering knowledge and improving educational experience has helped to shape this thesis and built a foundation of support.

The efforts of the ETIPS and CANLEAD research teams from conceptualization, management and especially collection of the online data have been critical in completing the analyses in this dissertation. The administration of Curry School of education has also been very kind and helpful in tackling the deadlines and formalities of the writing and submission process.

I have also been very fortunate to have the continuing support and wisdom from my dissertation committee, who have at every meeting provided guidance, context and many enhancements.

Lastly, I would like to thank my advisor Sara Dexter. Without her patient guidance and wealth of expertise this research may not have come to light at all.

TABLE OF CONTENTS

LIST OF TABLES

Manuscript Three: Using Evidence Centered Design and Educational Data Mining Methodologies With Click-Track Data To Explore Teacher Engagement In An Online PD Environment

LIST OF FIGURES

Manuscript Three: Using Evidence Centered Design and Educational Data Mining Methodologies With Click-Track Data To Explore Teacher Engagement In An Online PD Environment

Overview of Manuscript Papers

(Linking Document)

Caner Uguz

University of Virginia

Overview of Manuscript Papers

The last two decades have presented a staggering increase in the amount of services and tools provided through the web as access to the internet has become almost ubiquitous. At the same time mobile tools allow users to spend more time on online environments. The decade following the dot-com bubble showed that internet services are here to stay and provide both opportunities and issues in the great wave of change that they unleashed.

The field of education has not been immune to these developments with initial interest in discussion forums and online learning environments with websites such as WebCT, Blackboard and open source alternatives such as Moodle. In the last five years however, fuelled with success stories from Silicon Valley, young and experienced entrepreneurs have been rushing to tackle the diverse problems of the education industry. Areas of development have also included aspects of professional development from projects earlier in history of the internet such as TappedIn to today's networking and sharing tools such as Twitter and Pinterest.

Virtually every aspect of education has a digital component. Teachers heavily make use of online applications and apps in mobile environments for the purpose of researching content, preparing handouts, or organizing student information. They also provide online tools that include quizzes, media, blogs and wikis that are helping the movement to "flip" classrooms. In the meantime, online learning is carving a place in the formal and information education industry. Already in Virginia, many K-12 students must take an online course to graduate.

In accordance educational researchers have seen that their research environments have often moved beyond the physical environs of schools. In comparison to these online environments, the new online environments require technical expertise many researchers may not yet have, from the design of online interventions to the possibilities of customizing or creating new online tools. Despite this challenge, there has been a growing interest among researchers in utilizing affordances that the online tools provide.

Because of their digital nature, online environments can track user information and make just-in-time decisions based on this information. When users interact with digital tools they leave traces of their actions in a series of click-track information. This information can be collected in real time and acted on instantly if the digital tool has been programmed to do so. More importantly, this information presents a time-stamped transcript of all user actions. It is possible for researchers to know where the user clicked, when the user entered the system, and what path and duration they took.

While many researchers are becoming more aware of how to collect click-track data and its potential for research, few published research studies show utilization of it. A contributing factor may be lack of academic work providing frameworks that can be applied to click-track data. Such a framework would combine technical knowledge on how click-track data can be used, and knowledge of advanced statistical analysis, as well as guidelines for assessment of large-scale data.

To add to this area of educational research, this dissertation illustrates approached researchers can use to appropriately embed click-track data into their research design. The overarching question that this dissertation will address is "What are the relative advantages different methodologies offer for using click-track data in online educational research?".

This question provides the general guidance and motivation for this study. Specifically, focus will be on how educational data mining and evidence-centered design can be best used with click-track data gleaned from online professional development environments.

Background

This dissertation focuses on the research method of collecting users' click-track data to make sense of their behavior in online learning environments. Making sense of these data however, requires two additional components. One is a method of analyzing the data that fits well to the calculations and inferences that the data allows and the other is a theoretical assessment framework and methodology to guide the use of the data and analysis (see Figure 1). I provide a rationale for why these systems work well together in the following section.

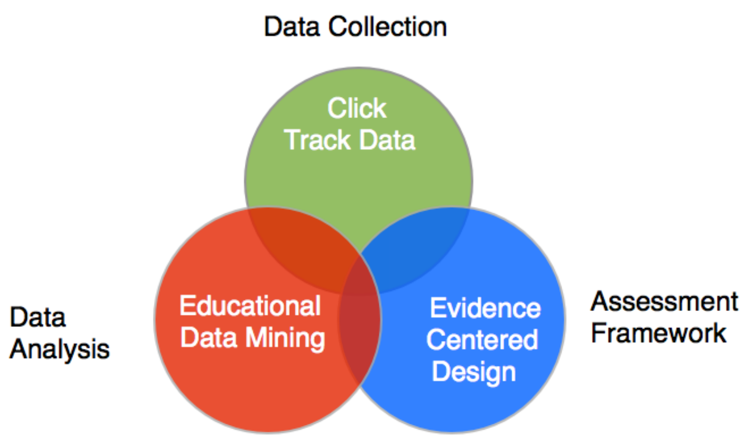


Figure 1. Three components for analyzing click-track data.

Click-track Data Analysis

In online environments users interact with a digital tool by using the mouse or other input devices such as a keyboard for a variety of tasks from clicking on links or scrolling the page to interacting with games and simulations. Almost all user activity completed on these online environments can be recorded, given the appropriate tracking tools.

In its early uses tracking user behavior was limited to logs that servers provided when users requested a page from the server. These are called server logs and in the research literature to date these defined tracking. However, they provide a very restricted amount of information, namely the pages that were visited and the associated times. In the last decade as web development improved, users were given many different methods of interacting with the web pages without necessarily requesting a new page from the server. Such information would be lost using only simple server logs and therefore more advanced user tracking systems such as Google Analytics were developed.

For online educational interventions, however, because the website is designed by the research team, however, the developers can design custom recording that can capture any information including mouse movements. The data collected in this fashion makes the terms "server" or "page" logs obsolete. Instead this activity can best be described as click-tracking, since this covers events that include new methods of tracking the clicks and other mouse behaviors as well.

Click-track data therefore is the information that is gained from a user's interaction on any website where the mouse is used to move, click or drag as well as any input completed by the keyword. This covers loading pages, viewing pages, showing/hiding page components, dragging page components and moving the mouse within the screen. Because of the detailed recording of behavior click-track data results in large sets of rows representing the users' every recorded action.

The advantage of using click-track data is its ability to allow for highly detailed records. In an analogy to observation in the physical world, a click-track data file is similar to a video recording of each participant where the recording has been coded in a quantitative manner to reflect the actions as they happen every second. In other words, a click-track data recording provides continuous observation of online behavior. Online environments also provide more opportunities to design the environment and focus behavior compared to face-to-face interaction by constraining what users can do on the page, which allows for more targeted data analysis. For example in a classroom it may not be feasible to redesign the layout of the classroom, change the items hanging on the wall etc but an online page is a blank canvas that can be designed for specific needs at a low cost.

For these reasons, click-track data can be a potentially highly useful tool when researching online environments. However, as in any data analysis methodology, theoretical and methodological guidance are essential to using click-track data appropriately.

Educational Data Mining

The first requirement for using click-track data effectively is a set of statistical methods that allow for these type of large data sets to be analyzed. Educational data mining (EDM) proves to be an important resource for this purpose.

EDM is an extension of the general field of data mining as it applies to education related data. Data Mining is a collection of statistical methods for making inferences about large data sets. One of the aspects that sets data mining apart is that it is mainly used for identifying meaningful relationships without a priori assumptions or research on these relationships. EDM is a growing field, with significant contributions in the last five years. The International Educational Data Mining Society was formed in July 2011 and the Journal Educational Data Mining has been in publication since 2009. Educational data mining can work with any data set however click-track data is well suited for analysis using EDM methods for the following reasons:

1. **Large data sets:** Educational data mining works with the relatively large data sets generated by click-track data. Before digital data collection tools were widely available, collecting large amounts of data was limited only to research projects with significant funding. However the increase in the digital educational tools in the past two decades has provided more sets of data for educators to analyze. It is therefore no surprise that the growth of the educational data mining field corresponds to the growing amount of available digital data from games and online learning environments. Click-track data can generate large amounts of information depending on the granularity of the tracking and how much interaction is possible, but usually several different activities are recorded each minute the user is online. This leads to thousands of interaction cases over the use of the click-tracking system. Even when the online environment is used only by a small group of individuals, the total interactions provide large data sets.

2. **Data collected requires a quantitative approach:** Data mining is essentially a collection of statistical analyses conducted with large data sets that help with finding patterns that may lead to prediction or classification. Some common analyses used are multiple regression, cluster analysis, decision tree, and discriminant function analysis, among others. These analyses rely on quantitative information that is often in the form of a ratio scale, or coded into ratio scales for calculations. This fits well with click-track data which mostly includes variables that are either a) number of occurrences, b) Boolean information about a certain event took place or c) time variables in seconds or minutes.

3. **Post hoc generation of variables:** Click-track data are the raw data from a user's interactions. From these data many additional variables of interest can be generated. For

example, from raw click track data on page visits users leave a timestamp of which pages were visited when. From these data we can create a variable called unique pages per visit, which includes the total of number of pages that the user visited that were unique. If the user visited 5 pages but visited 2 pages twice, the total unique pages for that session would be 3. From these data we can calculate a unique page ratio to have a better sense of what percentage of this person's visit is to unique pages by calculating *unique visits / total visits*. This would give a number between 0 and 1; closer to 0 would mean the user has been to only one page multiple times while 1 would mean all pages the user visited were unique. In our example the user's unique page score for this session is 0.6. It is up to the researcher to determine whether this information is useful and how to use it in statistical analysis to interpret its relation to other users and variables. In this case one may hypothesize that the more unique pages participants visit the more successful their interaction would be as indicated by an external variable. Thus, for example, click-track data allows for post-hoc calculations of variables from raw data, which dramatically expands the possibilities for variables that can be used. As long as the researchers design the tracking process to collect granular information they will end up with rich data that allow for post-hoc analyses to find patterns that were not anticipated at the beginning of the study.

Educational data mining offers great potential in terms of data analysis and preparation for click-track data. However this process would benefit from further guidance that connects educational theory to this methodology,to guide the creation of a variable, such as unique visit score, and whether it makes sense for the purpose of the research. Evidence Centered Design is a framework that can provide guidance for this process.

Evidence Centered Design

Evidence Centered Design (ECD) is an assessment framework developed by Robert Mislevy and colleagues to provide guidelines and specific methodologies for assessment where complex constructs need to be measured. The strength of the ECD model is that it goes beyond the creation of a simple rubric for assessments and suggests a systematic methodology for how test behavior should be linked to the constructs that are being measured.

The specific sub-component of ECD that is of main interest to this research is the Conceptual Assessment Framework (CAF), which offers guidance for the steps required to deliver assessment in educational settings from the conceptual development to the production of the assessment itself. CAF starts by dissecting the constructs of interest into smaller parts to build a “student model”. For instance "engagement" as a construct is general and although there is a significant amount of research measuring and working with engagement, the specific meaning of the construct depends on what engagement entails in the research project. In a Student Model researchers might separate engagement into better defined details such as "demonstrating interest in the subject matter", "interacting with other users", "making use of website functionalities", and so on. The next part of the CAF is the Evidence Model, where researchers translate aspects of the Student Model into observable behaviors. For instance, interacting with other users would be operationalized into what typical behaviors constitute interaction, and what amount of interaction is the threshold to lavel this behavior as interaction. A third component of CAF is the “task model”, which guides design decisions for how users perform these observable behaviors. The designed environment elicits behaviors to measure the construct of interest, which might range from a test environment like an actual quiz or test, to a game itself as the user works through its challenges. In an online environment the task model might range from an online test, to all of the interactions of tracked during the entire visit of a user to a website.

The way CAF components work together as described above highlights the fit of ECD with click-track data. One of the first areas of research in online environments using ECD included online simulations and later games because these two media types allowed for observations during and within the environment, in contrast with a quiz after the learning event has taken place. It therefore allowed for a more nuanced task model in which to operationalize and elicit the constructs of the student model. Click-track information brings the same opportunities to create a task model within any online learning environment.

At its core ECD is a framework that provides guidance to design an assessment from a conceptual basis, to ensure the evidence collected will inform the constructs. ECD does not dictate or provide a packaged data analysis methodology since the needs of the assessments vary. In one of the studies that comprises this dissertation I use the analysis methods from data mining for building the operationalized rubrics that will become the student model and inform the evidence model and the online environment will define the interactions that will constitute the task model.

Online Professional Development

Online learning is a growing field as almost every component of the educational experience is now being conducted online to some extent. Professional Development in education has also adopted online methodologies that started as simply providing the handouts and presentations in digital format online. Websites like TappedIn began to use concepts from networking and workshops that provided virtual rooms and interactions among users of the site, which increased the engagement of teachers. The last decade has seen online PD use different tools such as case studies, mock websites, and interactive environments. In its development online PD environments have moved from being static websites to interactive environments where teachers don't simply read material but engage with the content. This has allowed for opportunities to analyze what the teachers are doing online as a potential insight into how they learn from the PD environment; opening the door for potentially collecting and analyzing this data. Because of the significant breadth of all online learning and the recent opportunities online professional environments provide, this dissertation will focus on the analysis of data specifically from online PD. There are two specific reasons why professional development is a good area to study:

1. Online professional development is an area that can make use of the advantages of click-track data analysis to guide the field in achieving the often discussed goals of contextualizing and personalizing information for teacher learners, as well as measuring their learning and engagement. These interactions will be discussed thoroughly in the second manuscript's literature review, which uses data from an online professional development environment.

2. On a more practical note, considering the difficulty of getting access to online click-track data, the availability of existing data from previous and ongoing research at the Curry School has factored into the narrowing of this study in the area of professional development, as opposed to other fields.

Research Questions

In the section above, click-track data were described and introduced as a potentially powerful source of information in online learning environments but one which only provides raw data that can not be meaningfully used alone but in conjunction with data analysis methods and a framework for conceptualizing how it aids assessment. Data mining methodologies for education provide the analysis tools and research precedence to make meaningful observations from the click-track data, and ECD provides the methodologies to build reliable assessment design.

The purpose of this dissertation is to investigate the feasibility and relative advantages of approaching research in online learning environments with a combination of click-track data, EMD, and ECD, investigate methodological options for research, rather than working on a single aspect of online learning such as engagement, through the following three manuscripts.

Manuscript 1: A Review of the Literature

The first manuscript, "Does Educational Data Mining and Evidence Centered Design provide better use of click-track data in Online Educational Research? A Review Of The Literature" looks at existing research in education to get a better sense of how researchers have been using click-track data in online environments and whether there is evidence in research to show that the use of ECD and EDM can provide improvements over other existing methodologies. It provides an overview of the field and informs the remaining studies by addressing the question "What are the methods currently used by education researchers to analyze online click-track data?"

The methodology used in this literature review was to look at existing educational research literature published in between 2000 and 2013 for the ways that click-track data have been used, categories that emerge from the literature review, and then classify illustrative articles to review. ECD and EDM were used as search terms to make sure that literature that already uses these components could be included but the search was kept general to include any research paper that utilized to some extent click-track data in an online educational setting.

Manuscript one presents results and discussion of the literature review. The analysis of the literature suggested that the rigor of statistical analysis and the existence of operationalizations applied to the click track data were the two defining factors for classifying the existing research. Many studies did not make significant use of click-track data and used them only as supplemental evidence to triangulate other mostly qualitative measures. Researchers either misused or underutilized click-track data by not using stringent methodologies for either operationalizing variables or analysis. For instance, a common use of click-track data was to look at visits to the online learning environment where visits were interpreted as engagement and simple frequencies were used as analysis. On the other hand, there were several examples where EDM methods were being used effectively for online studies, but many of these studies did not use an assessment framework. Finally, only a handful of studies applied EDM and ECD together. The review ends with the discussion of the need of using ECD and EDM in more studies would be necessary to understand the ways in which these components help researchers make better sense of click-track data.

Following from this analysis of the literature, the next two manuscripts each use different data set to first investigate the benefits of EDM, and secondly combine together EDM with ECD.

Manuscript 2: Using EDM as a Post-hoc Tool

The next study in my dissertation applies educational data mining approaches in a post hoc method to see what type of value such analysis can add to online educational research after the outcome data have already been collected. The research question that is addressed is “How can data mining of click-track data add value to online professional development research?”.

The data for this paper draws upon results from a study done within ETIPS (Educational Theory into Practice Software); the data were collected in 2008-09 by Sara Dexter and Pamela D. Tucker, both of the Curry School of Education. ETIPS is a case-based learning platform for educators to learn and practice decision making. ETIPS simulates a school environment where the users imagine themselves working. The school is portrayed through a public website and intranet with a wide variety of information pertinent to the school's operation. Each case introduction provides a reason for the users to explore the school's websites and is written to reflect a realistic situation that poses opportunities and challenges requiring a decision. Students are asked to look for information on the school's website and intranet to glean information to respond to case questions. These short essay responses as a measure of students' understanding and analysis of the school's situation, and for the quality and accuracy of the decision making about the leadership challenge posed in the case introduction.

The ETIPS study lends itself well to a post-hoc analysis of click-track data because raw data from users' interaction with the school websites were recorded and saved. However, the click-track data have not yet been analyzed because the original study instead relied on students' essay scores. The use of post-hoc analysis is important here because it investigates how additional important information from EDM methods with click-track data might add value to the previously completed analysis. One of the claims of EDM is that even without targeted data analysis patterns can be found and since this analysis is distinct from the original study this claim can be tested.

Manuscript 3: Using EDM with ECD

The next study in this dissertation aims to investigate the combined use of ECD and EDM in an online professional development study. Whereas manuscript two aims to implement EDM as a post hoc measure, here I investigate how providing an assessment framework can provide a more cohesive data collection and analysis. A conclusion I drew in the literature review of manuscript one is that using click-track data requires a solid link between the constructs being measured and the specific click-track actions. The ECD assessment framework aids researchers in making this connection by their establishing the construct (as a student model) and then operationalizing it (in the task and evidence models). This process, however, requires the researchers build the assessment framework into the research design and for this reason ECD can't be applied post hoc. By definition it must be applied from the beginning of the design and research of an intervention. Data from the CANLEAD project allows for this study because it is a research and development project for an online learning environment and from the beginning applied the ECD framework in its data analysis. I worked on the development of measures from the beginning of the research design. The research question that is addressed in this manuscript is "How can ECD and EDM methodologies be used together to evaluate teacher participation in online professional development environments?"

CANLEAD is a multi-year study funded by Institute of Education Sciences to develop and research a leadership intervention that supports a leadership team's work with middle school math and science teachers. CANLEAD provides materials via a website for both leaders' and teachers' use. At this website in-service teachers can access information, resources and examples for integrating technology with specific math or science content; they also get an online environment for informal collaboration and learning. This software environment records user behavior online, but in this research I was able to provide input for what type of interactions should be recorded and as a result the click-track data is sufficiently detailed regarding interactions so as to allow for EDM.

This manuscript outlines the process of developing the ECD model for this project and tying the click-track data collection into it. It uses data recorded from user's interactions within CANLEAD and analyze the data using EDM methodologies. The specific research question with a combination of ECD and EDM is “how can we measure a teacher’s online engagement effectively in a custom built online professional development environment with unique interactions?” Here engagement is not based on an existing definition from the literature but rather on parameters that are specific to the research goals of CANLEAD and the affordances of its online environment. This illustrates the point that ECD is important because we often need constructs that are specific to the use case and need to rebuild our assessment models each time.

It is important to note that in both the second and third manuscript the research questions are related to whether the methodologies provide additional and valuable information. The goal of this dissertation is not to answer the research questions of the projects themselves but to tackle a methodological issue regarding how to analyze the learning environments used in the two different studies.

Conclusion

Online learning is an increasingly important field for educational researchers and more methodological studies are required to help researchers make better sense of what is going on in digital learning environments. Educational research has had decades to build and improve skills that are used for research on physical classrooms—from survey methodology, to observation and coding protocols, to data analysis. Online environments bring new challenges to these techniques but also promising new data sources and methodologies for novel approaches. This dissertation is therefore a small but timely contribution to the ongoing discussion about best practices and potentials for the research of online learning environments.

This dissertation also fits well to the structure of the three-paper manuscript because it synthesizes theory with examples of implementation, which new methodologies require. A fully theoretical approach to the potential of click-track data would not have been useful without practical research applications. Also, using two instead of a single research project as an avenue for testing these methodologies allows for the much-needed re-application of techniques in different settings.

Finally, this topic fits my personal and career goals. As a programmer I have been building web applications and analytics scripts that provide a good picture of the online interaction. Pursuing this research is helping me advance my thinking and training by also looking at data analysis methods and research. My belief is that knowledge of the data generation tools cannot help us solve educational problems online unless they are coupled with understanding of instructional technology principles. My goal in the future is to continue developing online educational solutions and I believe this dissertation will significantly inform and be a consistent part of my professional development. I hope at the same time it will provide guidance and further perspective to fellow researchers and make a meaningful addition to the corpus of our knowledge in this specialized field.

How Does Educational Data Mining And Evidence Centered Design Make Use Of Click-Track Data In Online Educational Research? A Review Of The Literature

Caner Uguz

University of Virginia

Introduction: Trends and Need

With the advances in web-based resources and tools, many services are being implemented in purely online or hybrid environments, and education has been one of these areas that is strongly affected by this development. Online learning provides significant affordances to the learner, such as freedom to allocate time at asynchronous participation, saving commute time, removing geographical barriers to involvement, and continuous access to information. The increasing adoption of online learning in both traditional institutions and new ventures has shown that the trend of providing more of the learning in online environments is only going to continue (Allen & Seaman, 2013).

This trend, however, brings a challenge to educational researchers working with online or hybrid environments. Traditional measurement methods of qualitative observation, such as classroom observations, focus groups, think-aloud methods, or interviews, are powerful tools that are available in hybrid and even some online environments but they have limitations when online classes have large enrollment, geographic dispersion of students, or lack of options for face-to-face class meetings. Conducting interviews with 20,000 online class members would not be feasible, nor would doing exit interviews with students from across the country. Yet, online environments create observational opportunities of their own. For instance, one can track user behavior by keeping a log of the user interactions online and thereby generate prolific observations that provide data on the activities of the student to a second-by-second level of precision. These methods have thus far been used only sparingly in existing studies most likely because of the difficulties in operationalizing large sets of log data into measurable constructs that make sense for the research purpose. Because of the different needs of each research project, a cookie cutter process cannot be established. However, there have been increasing use of specific methodologies in the past few years to suggest techniques for best practice.

Within the professional organizations for educational research two methodologies have garnered increasing attention. Evidence centered design has grown as an area of interest for building assessment models while educational data mining has concentrated efforts on using data mining methods in educational (mostly digital) settings.

This paper aims to answer key questions that any researcher will be investigating as they think about the methodological aspects of researching their online learning environment: What are some of the ways to make good use of the data available through logs online and whether specifically ECD or EDM have demonstrated their effectiveness in existing studies? In order to get a better sense of the field it is important to step back and analyze as many examples of online educational research as possible and evaluate how the ECD and EDM methods have been used in the last decade within the landscape of online educational research.

Background Information

Click-track Data

Click-track analysis goes back to the 1980s with the very first records created on web servers as users visited the websites these servers hosted (Agosti, Crivellari, & Di Nunzio, 2011). Today developers of any online environment might program servers to log basic access from any user, which includes page visits, along with any additional logs to get information about other behaviors such as mouse clicks within the page. While logging user interaction is very common, there is no standardized term used to describe this process.

The technical product that is created by a system where the system records all its activities is called a "*log file*" and the shorter term "*log*" has been used in the industry and literature. A "*web log*" refers to the logs on the web but it is often confused with a "*weblog*", which means blogging. This paper uses the term "*click-track data*" to emphasize that the recording is tracking the clicks of the user, which is a predominant method of user interaction online (as opposed to, for example, counting artifacts made online, such as discussion posts). Click-track data is generated when the user clicks links or elements on the page that lead to actions such as opening or closing an information box or loading a new page. Used in conjunction with timestamps, click-track data can provide a wide array of variables such as length and time of visit, path followed in the system, and logins. This review looked at literature that included any form of click-track data, even if authors used different terminology to describe what is included in the definition above.

By its nature click-track data is a method for continuous collection of user interaction without a priori restrictions on the time of data collection or the specific actions collected; therefore click-track data produces a large set of undifferentiated information. Researchers using click-track data to answer research questions still need to go one step further by identifying markers in the click-track data and specifying the inference they are making from those markers. Click-track data in that sense is similar to a continuous recording of video, which still requires researchers to use as a coding process to make sense of the data. Just as coding is used in the qualitative research of real world environments, researchers using click-track data need to systematically define and operationalize the constructs that they are studying to make sense of the data. Some typical approaches include (a) using existing operationalizations based on prior research; (b) creating rubrics that show the connection of constructs with multiple metrics in data; or (c) applying an existing framework for operationalizing assessments to the unique environment from which data are taken. One prominent example of the third category is evidence-centered design (ECD), which provides tested methods for operationalizing large sets of data with granular information.

Evidence-Centered Design

ECD is an assessment framework developed by Robert Mislevy and colleagues for large-scale assessments in which the main goal is to use a systematic model to develop constructs into assessment or measurement items (Almond, Steinberg, & Mislevy, 2002).

The emphasis on the construct in the ECD framework stems from an earlier theoretical assessment framework presented by Messick (1989), where he argues for the importance of understanding the underlying aspect that is being measured and not focusing solely on the assessment task . In keeping with Messick's suggested focus, the ECD framework requires the assessor to consider the links between the construct and what the construct means (which ECD calls the Student Model), what behaviors will indicate that the student is showing required behavior as defined in Student Model (which ECD calls the Evidence Model) and what tasks will give the student the appropriate opportunities to show that behavior (which ECD calls the Task Model) (Mislevy, Steinberg, & Almond, 2003).

ECD has been widely used in large-scale testing with the Advanced Placement program assessment modules (Brennan, 2010), but it has also been used in online learning environments to measure learning outcomes to teach digital networking, where the authors found ECD to be flexible and well suited for large scale data from online systems (Behrens, Mislevy, Bauer, Williamson, & Levy, 2004; Behrens, Levy, & DiCerbo, 2010).

Because online learning and research in online environments is relatively new, researchers need to form their own operationalizations for assessing constructs using click-track data. There are several reasons why ECD is well suited for analyzing click-track data for this purpose. Firstly, ECD encourages in-depth, construct-focused thinking about what to measure and therefore works well as an a priori systematic model for making decisions about what data to collect. Secondly, ECD provides step-by-step guidance for deciding what constitutes evidence and how the evidence can be observed. Thirdly, ECD has already been successfully applied to large sets of data and with click-track information where complex constructs have been adapted to unique learning environments.

Educational Data Mining

Another strategy in analyzing large amounts of data sets is looking for patterns in the data after it has been collected. *Data Mining* is a growing field that provides methodologies of finding meaningful relationships in large data sets, an approach which is increasingly being adopted by educational researchers investigating online environments.

Data Mining is the process of discovering structural patterns in large data (Witten, Frank, & Hall, 2011). Any set of data can be analyzed with data mining procedures, and, in essence, data mining is nothing more than a collection of statistical methods. The main difference, however, is that data mining is generally exploratory and does not necessarily conduct analyses based on existing frameworks or research outcomes. For instance, a traditional research paper would define variables that the study investigates and may run regression analysis to see whether the effects are statistically significant. Data mining however looks at a data set to *discover* if any of the measured variables show significant relationship to the outcome variable. Because of its exploratory use data mining is considered a post-hoc method and does not require an established set of relationships between the construct and the data.

Data mining is also a collection of statistical methods that allow researchers to uncover patterns in data. When we talk about data mining methods we generally mean inferential statistics and advanced analysis methods such as factor analysis, cluster analysis, multiple regression, and classification methods. One of the main differences of data mining in comparison to the standard ways these methods are used is the increasing use of unsupervised techniques. Supervised learning provides labels, parameters or the relationship that is expected while unsupervised learning uncovers these patterns from the data without assumptions about relationships.

Educational data mining (EDM) comprises both a movement and a methodological framework that applies the data mining principles into the education context, borrowing from the literature in statistics, psychometrics and education (Baker & Yacef, 2009). The development of EDM studies in the last decade correspond with the increased interest in online or digital learning environments (Romero & Ventura, 2007). This is no coincidence as EDM is well suited to deal with the challenges presented by click-track data, because EDM methods work very well with the large sets of data that most click-track systems provide. Click-track data is in the form of continuous raw data that is not grouped or filtered a priori for a specific construct. The affordance of this structure is that it allows researchers to go back and look into relationships after the data is collected. In observation-based research this would be similar to video recording the environment the entire time and going back to analyze the video after the research to look for meaningful patterns.

At the same time that EDM has grown into a sub-field of education there has been a similar development in the field organized around the term “learning analytics”. Learning analytics (LA) focuses on the larger question of best practices of using data to improve the learning experiences; it therefore covers a larger group of data analysis methods (Haythornthwaite, de Laat, & Dawson, 2013). LA also uses several of the same approaches and statistical methods used in EDM. However according to George Siemens, one of the pioneering researchers in the learning analytics field, “Although the techniques used are similar in both fields, EDM has a more specific focus on reductionist analysis” (Siemens, 2013, p. 1382). Despite some differences in the philosophies and methodologies, specifically for click-track related data, both LA and EDM use highly overlapping methodologies and in several of the papers analyzed in this study the phrase “data mining” was used with “learning analytics”. EDM generally provides a more specific treatment of data than LA and therefore EDM is the preferred terminology used for the literature review and analysis in this study.

Summary

In summary, the growing use of online environments for educational purposes offers an opportunity to gain more insight about the learner through the use of click-track data. The challenge this poses, however, is how to make sense of such large volumes of data. The literature shows two main approaches that have mostly been used separately. One is the use of systematic models such as ECD that help researchers develop structured measurements that closely fit constructs that are being measured. Yet systematic models that require a priori operationalizations are not well suited for exploring potential new connections or novel avenues of measurement. On the other hand post-hoc EDM highlights statistical methods for finding patterns and relationships that prior research or operationalization could not account for and therefore allows the researchers to explore the data from different perspectives. EDM alone however lacks the guided analysis that a systematic method would require. Finding patterns and relationships is inevitable through EDM methods, especially with large amounts of data, but understanding which relationships and patterns are meaningful in the context of the research requires some a priori development of the construct. This overview of ECD and EDM suggests that these existing approaches might yield the best results if used together.

In this paper the theoretically advantageous combination of EDM and ECD is investigated within the educational research literature to see whether such a combined methodology actually provides a viable solution that addresses the opportunities and challenges of the click-track data in online learning environments. Existing literature on using either ECD and/or EDM with online click-track measurements are reviewed in order to address the question, “What are the best practices for making use of data available through logs online and what is the added value of using ECD and EDM methods?”

Methodology

The review of the literature in this paper was done through a search that included the online databases Academic Research Complete, Education Research Complete, ERIC Psychology and Behavioral Science Collection, Education Full Text, and ScienceDirect. The search covered articles published between 2000 and 2013. The search terms included “online learning” as one of the search parameters and one of the following for the other parameter using the AND operator: "click-track data," "web log analysis," “data analysis," "operationalizing," "educational data mining," "evidence centered design." Initial searches yielded more than 20,000 results. To reduce the number to a manageable selection, some keywords were searched for in the article's abstract (e.g. online learning, click-track data, web log analysis), which found a total of about 2,000 papers through different combinations that were combed through for further refinement, as described below.

The next stage of the analysis eliminated papers that did not meet the following criteria:

- includes an online educational environment;
- includes analysis of click-track data; and
- not highly similar in approach to other papers.

The steps above ensured that studies actually involved both online and click-track methods, since many of the studies found through keyword search made references to both but did not use them in their actual study. Because this paper is a literature review that represents key themes, rather than an exhaustive meta-analysis of outcomes, paper with results similar to others in their approach were not included if they were deemed to not provide additional insights. Since the types of generated variables from click-track data are limited, several papers described using very similar approaches; for instance, using triangulation has been a common theme in many papers.

After applying these criteria there were a total of 36 papers that either directly used click-track data in their analysis or provided reviews of papers that used click-track data (see Table 1). These papers were analyzed for emerging themes, which are reported upon in the discussion section.

An important limitation in this study was the difficulty in finding methods-specific articles. Online learning is a growing field, but many papers do not clearly define their methodology in a way that was easy to find in the databases. Adding to this problem is the highly divergent naming conventions used for click-track and log data, which was often referred to with terms as vague as "interaction" or "engagement" data. For these reasons it would be difficult to claim that this review provides the most comprehensive analysis of all click-track data based studies; however, it does present a representative sample of existing research with click-track data in online learning environments.

Table 1

Distribution Of Papers Based On Their Use Of Systematic Models Or Data Mining with Click-Track Data.

		Uses Systematic Models (including ECD)	
		No	Yes
Uses Data Mining	No	10 papers: Ben-Zadok, Leiba, & Nachmias, 2010, 2011; Brook, 2003; Hrastinski, 2008, 2009; Keramidas, 2012; Michinov, Brunot, Le Bohec, Juhel, & Delaval, 2011; Rafaeli & Ravid, 1997; Warwick, Terras, Huntington, & Pappa, 2007; Xie, 2013	6 papers: Conejo, Guzmán, Perez-de-la-Cruz, & Barros, 2013; Frezzo, Behrens, & Mislevy, 2009; Jafari, SoleymaniSabzchi, & Jamali, 2013; Kupczynski, Gibson, Ice, Richardson, & Challoo, 2011; Morris, Finnegan, & Wu, 2005; Romero & Barberà, 2011
	Yes	17 papers: AlShammari, Aldhafiri, & Al-Shammari, 2013; Antonenko, Toy, & Niederhauser, 2012; Cocea & Weibelzahl, 2006, 2007; Fisch, Lesh, Motoki, Crespo, & Melfi, 2011; He, 2012; Hershkovitz & Nachmias, 2009, 2011; Hung, Hsu, & Rice, 2012; Ke & Hoadley, 2009; Kerr & Chung, 2012; Merceron & Yacef, 2007; Perera, Kay, Koprinska, Yacef, & Zaiane, 2009; C. Romero & Ventura, 2007; Romero, Ventura, & García, 2008; Schlager, Farooq, Fusco, Schank,	3 papers: Gobert & Pedro, 2012; Rupp, Levy, & DiCerbo, 2012; Sweet & Rupp, 2012

		& Dwyer, 2009; Valsamidis, Kontogiannis, Kazanidis, Theodosiou, & Karakos, 2012)	

Discussion

Because the papers in this review were selected based on their methodology, they were very diverse in their topics, participant characteristics, functionality of the online environments, and educational contexts. The results show that the existing research can be explored in two dimensions of whether they used systematic approaches including ECD or whether they used EDM methods, creating four categories altogether (see table 1).

1. The first category had 10 papers that used click-track data without systematic models or data mining.
2. The second category had 6 papers all of which used systematic models (including but not restricted to ECD), but didn't use data mining methods.
3. The third category had 17 papers that used data mining, but no systematic models.
4. The fourth category had 3 papers that used both ECD and EDM in the same study.

In the next section we will discuss the emerging themes, strengths and weaknesses within each of these categories.

First Category: Click-track Data Without Systematic Models Or Data Mining

One of the most important aspects of analyzing data for complicated constructs is the process of operationalization, which consists of converting the conceptual goal into measurable components. The connections between components should be established either by research, where existing studies provide precedence, or through systematic exploratory investigation. In this first category we see a group of studies that drew upon click-track data as data sources, but did not develop complex explanations for how those data sources represent the construct under investigation. For example, in early examples of studies using click-track data researchers considered the number of visits or “hits” alone an indication although not the complete picture of user interaction with the online learning environment (Brook, 2003; Rafaeli & Ravid, n.d.).

Such a simplistic approach has been repeated in many other studies investigating online interaction, especially in talking about participation online. One of the studies in this review looked only at the number of discussion posts as a measure of participation, and this metric was used as the basis of the paper for a relationship between procrastination and participation online (Michinov et al., 2011).

In a literature review on the topic Hrastinski (2008) found that the tendency to associate constructs with minor click-track data observations (e.g., number of visits to measure participation) is rather common in educational research. Among the 36 research papers Hrastinski investigated he found that researchers used widely varying definitions and inconsistent metrics for tackling this issue. Some researchers used the amount of writing students did as a measure of participation where reading and interacting with the website were ignored. Other researchers used frequency counts of how many times students viewed materials or the length of time they stayed logged in as measures of user interaction, which are by themselves not necessarily indicative of engagement.

Following this analysis and examples from other research Hrastinksi provides in a later article (2009) a better definition of online interaction that researchers can use as a guide for what to look for. However, his definition does not provide an operationalized metric that online click-track research can use. Hrastinski points out " a process of taking part" but what that translates into depends on the context of the online environment and therefore needs to be translated into a rubric for each environment being studied.

An advantage of using rubrics is that researchers need to consider the relative weights and relationships of variables to each other in order to describe a larger concept. Three of the papers in the review (Ben-Zadok et al., 2010, 2011; Xie, 2013) showed that click-track data analysis suffers from the same problem when several variables are combined to separately explain a construct. These papers chose the approach of looking at correlations of these variables separately and show significant relationships. However the combined effects of multiple variables are not considered and an overarching construct is not defined because they are investigated separately. These papers also do not offer any statistical consideration for collinearity and interaction effects being taken into account.

It is important to note that not all online click-track data collection be paired with the step of operationalization or even systematic analysis. For instance, when the goal is to present descriptive data such as statistics on visits and general use patterns, simple frequencies and descriptive statistics based on click-track data would work, as in the case of showing the rate of usage of online resources (Warwick et al., 2007), or providing trends over time. Another method of using click-track data for descriptive purposes is checking whether students completed online assignments, or checking whether students were present at a synchronous sessions, neither of which require further steps of analysis (Keramidas, 2012). In these cases the researchers are not testing constructs but simply looking at how the users are interacting with the website.

Aas seen in these examples, it is somewhat common to use click-track data to infer relationships between the observed behavior the click-track data represents and the construct. The main methodological drawback in these cases is that these researchers did not follow a rigorous procedure to tie constructs to the online behavior components in their triangulation or comparison efforts. For instance, number of visits in a website cannot be a sole indicator of the amount of user engagement in the classroom, because it lacks other variables that could play a significant role; for instance total amount of time spent on the website, or the breadth of different pages visited.

In summary, we see that papers that used click-track information for descriptive purposes are served well with basic information that shows whether certain components were completed online, and then use this data to triangulate with other more complex or rigorous methods of measuring the construct. For instance, a study could measure learning outcomes following an online intervention with a final exam and also look for how many online quizzes were completed through click-track data to see if that behavior is meaningfully related to scores. However, when the research goal is to use click-track data as a measure for a complex construct such as engagement, limited efforts to operationalize constructs of interest could lead to potentially inaccurate conclusions and misleading relationships.

Second Category: Systematic Models Without Data Mining

As the use of online tools for learning has increased, the concept of click-track data has become more widespread, resulting in articles that make full use of the click-track data as a measurement approach and not just to provide supporting indicators. Compared to the first category, this next category of six articles more systematically made inferences from click-track data by identifying all the variables that made sense for the project and converting them into a rubric.

One approach is to establish a systematic model is take the initial clues from existing literature that aim to establish relationships between constructs and individual metrics, combine several of these metrics and evaluate their relationships, with the goal of creating an accurate model of the construct as a rubric. One of the studies in this review (Morris et al., 2005) looked at time spent online as a multifaceted construct based on several time-related variables while another study picked a set of variables after a careful investigation of the potential impact of each on an outcome construct such as achievement (Kupczynski et al., 2011). Both articles referred to existing research in making the connections that suggested there could be a link between these variables and actions observable online.

A systematic model does not always need to rely on existing literature but instead may provide a new model and test it using click-track data. An example of this type of use of click-track data is evident in the research conducted by Romero and Barberà (2011) in which the authors delve deeper into on-task behavior by looking at the *quality* of the time spent where quality was defined as what that time meant to the individual. They use a combination of click-track and self-report data to learn more about the context of the time spent to have a better understanding of the quality of time. This approach is a good example of using steps to verify relationships between constructs and data as part of the research design when existing research is not available.

While the studies above express relationships between variables in a rubric format (assigning levels to components of a construct), others use more formulaic approaches by using algorithms that are based on existing studies on relationships (Jafari et al., 2013). An algorithm is a method of representing relative impact of different variables which, when adopted a priori, requires building a model of impacting factors. Both rubrics and algorithms are necessary steps in delving deeper into the components of the construct and fulfill the "systematic" aspect of this category.

This review found, however, that such systematic approaches are not as widely used in relation to click-track data. One of the reasons is that basing the model on existing literature is difficult since exhaustive sources on valid measurements in an online environment do not yet exist. In addition, the differences in implementation environments make it difficult to adapt existing studies. Yet, as described above, the ECD framework is highly suitable for developing measurement models in these cases, because its systematic approach walks researchers step-by-step through identifying what needs to be measured, how it can be measured, and how the measurement looks to the user, in order to develop rubrics and/or algorithms for the constructs.

Two of the online research studies in this category have used ECD in complex assessment environments that require analysis of online click-track data. Conejo and colleagues (2013) used ECD to tackle the complex concept difficulty of task for an adaptive learning environment. The authors were able to improve their analysis by using more objective measurements of difficulty made possible by the use of the click-track data. ECD has also been used in a long-running online simulation project where user behavior data has been analyzed to teach students the knowledge and skills required for network administration tasks (Frezzo et al., 2009).

The studies in this section suggest that providing a systematic component in the research has the potential to improve how well the click-track data represents the construct of interest; however, the complexity of forming systematic models necessitates the use of methodologies for developing such models. The relatively few studies employing the ECD methodology demonstrate its suitability for objectively building complex models of constructs, however its use requires reliance on a priori research and methods and thus doesn't leverage all the possibilities inherent in the data. The next section will discuss data mining as an approach that aims to further explore the click-track data.

Third Category: Data Mining Without Systematic Models

With the increase in online educational learning environments researchers increasingly have at their disposal large sets of click-track data that allow for exploratory analyses to be conducted. Data mining has been the collective methodology known in the past 30 years for these types of analyses and it has been a hot topic in education in the last decade. Researchers have provided papers that advocate for data mining methods in education by outlining their benefits for researchers, as well as teachers, especially in making use of the click-track data in online environments (AlShammari et al., 2013; Merceron & Yacef, 2007; Romero & Ventura, 2007). Here, data mining methods applied in education are referred to as EDM. The analysis of studies in this review found five main areas that EDM can improve research in online environments; each is discussed in turn in the following sub-sections.

Provide in-depth representation of constructs. EDM methods can analyze large amounts of click-track data to provide a more complete definition of constructs. For instance in the first category of paper discussed earlier, understanding *off-task behavior* would be defined simply as idle time where the student is not doing anything in the system. EDM, however, has been shown to draw a much more accurate picture of off-task behavior by exploring models that include information from log files in addition to time. Baker (2007) was able to find a more accurate picture by first building a model in which students from a small segment of the group are on-task, and then mined for which attributes in the click-track data logs correspond to this outcome behavior. Other researchers have also been able to identify variables that reliably predict engagement and motivation in an online environment for future users by doing post-hoc analysis of different attributes of online interaction with a sample group (Cocea & Weibelzahl, 2006, 2007; Hershkovitz & Nachmias, 2009). In these cases researchers used EDM to measure the accuracy of their a priori rubrics to see whether EDM variables similarly categorized the users, which then allowed them to enhance their a priori definitions of constructs.

Investigate processes with order of actions and paths of responses. EDM has been used to investigate constructs that include processes such as problem solving, the operationalization of which includes considering strategies combined into multiple steps taken to reach outcomes. Researchers were able to identify and distinguish specific strategies students used from analysis of their behavior online (Fisch et al., 2011) as well as a quantitative content analysis of text looking for patterns (He, 2012). These methods help researchers evaluate strategies as they appear rather than predefining a group of strategies that may not fit into what the students are actually doing, and thereby missing key user activities.

Cluster and group participants or responses. Another method from data mining that can be helpful is *cluster analysis,* which has been shown to successfully work with online learning environment data to categorize users into groups (Antonenko et al., 2012; Kerr & Chung, 2012; Perera et al., 2009; Valsamidis et al., 2012). Cluster analysis can also help create measures by providing the cut-off points for items. This can be especially useful if no a priori guidance from the research literature is available to make informed decisions about cut-off points for such measures.

Investigate relationships. Another area where data mining methods have been shown to be helpful in click-track analysis is social network analysis studies where researchers are analyzing the nature of interactions between members of groups. One of the papers analyzed for this review by Schlager and colleagues (2009) provides both a convincing theoretical argument as well as an implementation example. The authors argue that click-track data analysis makes this process easier, because a large set of data is collected immediately and does not rely on self-report of participants. In turn, using EDM methods researchers are able to look at this large data set to investigate a wide array of characteristics that reflect the social networking opportunities the website provide. To make this case more concrete: If we compare a priori and EDM methods, we see that a priori methods need to identify where in the software the users are interacting with each other and measure whether the users are doing these interactions. EDM methods, however, would look at every possible interaction between users to see which ones are stronger indicators of user interaction and which users do these most. This is the approach that has been used in the paper by Schlager and colleagues to get in depth understanding of teacher interactions in the online system *Tapped In* (www.tappedin.org).

Evaluations of online environments. Operationalizing full system evaluations in an online environment is a challenging task and researchers often have difficulties finding in-depth information to answer general questions such as "Is my online intervention working?" A recent review of online learning community evaluation studies shows that very few of the evaluations use any click-track data for their analysis, instead relying on qualitative data (Ke & Hoadley, 2009). However, EDM has been shown to work well with online programs, especially when the learning management system in use generates a large number of variables about student interactions that allow for large-scale analysis (Hung et al., 2012). This trend is likely to continue as more researchers use learning management systems such as Blackboard, Moodle and others which include functionalities for generating and exporting click-track data (Romero et al., 2008).

Despite all the advantages listed above, EDM methods by themselves also may not provide the best approach to analyzing click-track data, especially when they lack guiding a priori assertions about the constructs. From a post-hoc, or data mining analysis perspective, Herchkovitz and Nachmias (2009) have done some highly interesting work on analyzing click-track data to better understand the consistency of some of the potential variables in rubrics. They looked into the students' pace as they interacted with computer-based instruction to determine whether averaged values for pace are correlated. The low correlations they found showed that a seemingly uncomplicated concept such as pace of interaction could not be reliably operationalized without analyzing it in conjunction with other variables such as the interaction context, type of action required by user, and so on. Herchkovitz and Nachmias suggest the use of data mining to refine the definition and operationalization of seemingly linear variables but they also provide some a priori suggestions for possible relationships that do not only rely on data mining findings.

Another potential problem is that while EDM methods are universal, they can't be universally applied as a standardized measurement instrument to all online learning environments. There are efforts to use click-track data to develop some frameworks that researchers can use in any online learning environment (Agudo-Peregrina, 2013) so that some common conceptual measures can be used across the board, but all the papers investigated in this review have had several unique components that would either create difficulties in alignment or make unified instruments too general to be useful as a measurement tool. This problem also enhances the view that each study, even when they use EDM methodologies, would need to work towards operationalization of constructs through systematic models that make sense for the specific research questions and environments.

Fourth Category: Data Mining with Systematic Models

In the discussion of the existing methods so far we have seen that both systematic models and the data mining procedures provide insights into online learning, but both have drawbacks that limit their use. This raises the question to what extent are they combined in research studies.

This question was addressed in a special issue on integrating ECD and educational data mining; its purpose was to illustrate to the educational data mining community how the strong design components of ECD, as an a priori approach, are complementary to data mining methods (Sweet & Rupp, 2012). In this issue Mislevy also points out to the seemingly contradictory methodologies while one is "structuring situations to evoke particular kinds of evidence" (i.e. ECD) versus "discovering meaningful patterns in available data" (i.e. EDM) (Mislevy, Behrens, Dicerbo, & Levy, 2012, p. 11). He points out that EDM provides several methods that confirm, adjust and discover the components of the evidence model one would create via ECD so that one could have a better idea of what constitutes evidence for the constructs under consideration.

Examples of the applied combination of ECD and EDM are not widespread in the literature since the approach to the use of either of these methodologies is relatively new. One of the examples of the combined use has been to measure scientific inquiry skills by collecting click-track data as students interact with science simulations online. The authors discuss older methods of analysis that included either (a) a *knowledge engineering* perspective in which user behavior would be compared to a pre-determined set of "correct" actions in the simulation, or (b) a data mining approach in which correct behavior would be extrapolated from post-hoc analysis of student interaction. In this example researchers used both approaches enveloped in an ECD design where correct actions were defined but also modified based on data mining results to build a robust evidence model (Gobert & Pedro, 2012). Similar approaches have been used with epistemic games online in science education (Sweet & Rupp, 2012), as well as simulating environments designed for engineering education (Rupp et al., n.d.).

The combined use of ECD and EDM is very recent and as a result a handful of flagship studies illustrate the potential of this combined approach. However, they all share the same perspective that frames the two methods as complementary processes. Not all click-track data studies need to use both methods, but in cases where a relatively complicated and domain specific construct (e.g., motivation, engagement, achievement) is used, it would be fruitful to scan the literature for existing measurement models and build from the literature an ECD model that represents the unique research goals of the study, while supplementing or revising the model with outcomes from EDM techniques.

Conclusions and Implications

Understanding complex constructs in an online environment should follow the same principles as research in an offline setting–constructs should be operationalized systematically. ECD and EDM both offer advantages for a systematic analysis of click-track data and can be used together, which suggests that this combination would be better than using each one individually, or none at all as is the case in many studies of online environments. Such a combination would allow researchers to develop a clear idea about what they are looking for and then to measure the data in a way that corresponds to the users' interactions.

A more important reason to integrate these methods is to ensure that data analysis does not lead to data mining outcomes that rely heavily on correlation and/or show misleading connections because they are not linked to a system of interconnected variables. There is a growing concern in the field about the misuse of data analytics in education to draw conclusions for student behavior (Dringus, 2012). This concern is valid and can be addressed by the structured (ECD and EDM) operationalization procedure advocated for in this review.

It is also important to note that online learning measurements do not need to only rely on click-track data. Other quantitative and qualitative approaches using text analysis can also provide useful outcomes (e.g., the operationalization of sense of community). However, defining and understanding whether participants in an online environment feel a sense of community remains to be a challenge with only click-track data. Rovai's (2002) work on community in online environments has been widely used and relies on an external survey with a good record of reliability and validity. Other studies have successfully used qualitative methods including user generated text analysis of online discussions (Chen, Wang, & Hung, 2009; Ioannou, 2011), conducting interviews with participants in online settings (Barnett & Morran, 2002), or program evaluation reports at the end of online and hybrid programs (Owston, Wideman, Murphy, & Lupshenyuk, 2008) as well as a combination of these methods (Cheung, Hew, & Ling Ng, 2008). Having complementary measures of online behavior with offline data collection such as interviews, or alternative online data collection such as online surveys might give us clues about how to construct our click-track data collection to include all users, especially those who do not leave traces through clicks as much as other users but may be engaged with the content nonetheless (Beaudoin, 2002).

While these concerns are warranted, a large percentage of the papers about online learning are using very little of the data available through tracking users' clicks and instead rely heavily on qualitative methods. Especially in using click-track data, mixed-method approaches are likely to yield the best results and make the most of the affordances of the medium that inherently provides opportunities for both qualitative and quantitative methods (Bruckman, 2006). When the analysis of click-track data is implemented with careful operationalization of constructs and follows the systematic methodology associated with ECD and EDM, it has the potential to provide new insights into online learning and add to the understandings revealed through qualitative methods.

References

Agosti, M., Crivellari, F., & Di Nunzio, G. M. (2011). Web log analysis: a review of a decade of studies about information acquisition, inspection and interpretation of user interaction. *Data Mining and Knowledge Discovery*, *24*(3), 663–696. doi:10.1007/s10618-011-0228-8

Agudo-Peregrina, Á. (2013). we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online. *Computers in Human …*, 1–22. Retrieved from http://www.sciencedirect.com/science/article/pii/S074756321300188X

Allen, I., & Seaman, J. (2013). *Changing Course: Ten Years of Tracking Online Education in the United States.*

Almond, R., Steinberg, L., & Mislevy, R. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *The Journal of Technology, Learning and Assessment*, *1*(5). Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Enhancing+the+Design+and+Delivery+of+Assessment+Systems+:+A+Four+Process+Architecture#0

AlShammari, I., Aldhafiri, M., & Al-Shammari, Z. (2013). A Meta-Analysis of Educational Data Mining on Improvements in Learning Outcomes. *College Student Journal*, *47*(2), 326–333.

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development*, *60*(3), 383–398. doi:10.1007/s11423-012-9235-8

Baker, R. S. J. d. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07*, 1059. doi:10.1145/1240624.1240785

Baker, R., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*. Retrieved from http://www.educationaldatamining.org/JEDM/images/articles/vol1/issue1/JEDMVol1Issue1_BakerYacef.pdf

Barnett, M., & Morran, J. (2002). Addressing children's alternative frameworks of the Moon's phases and eclipses. *International Journal of Science Education*, *24*(8), 859–879. doi:10.1080/09500690110095276

Beaudoin, M. (2002). Learning or lurking?: Tracking the "invisible" online student. *The Internet and Higher Education*, 1–10. Retrieved from http://www.sciencedirect.com/science/article/pii/S1096751602000866

Behrens, J., Mislevy, R., Bauer, M., Williamson, D., & Levy, R. (2004). Introduction to Evidence Centered Design and Lessons Learned From Its Application in a Global E-Learning Program. *International Journal of Testing*, *4*(4), 295–301. doi:10.1207/s15327574ijt0404_1

Behrens, J. T., Levy, R., & DiCerbo, K. E. (2010). *AN EVIDENCE CENTERED DESIGN FOR LEARNING AND ASSESSMENT IN THE DIGITAL WORLD* (p. 46). Retrieved from http://www.cse.ucla.edu/products/reports/R778.pdf

Ben-Zadok, G., Leiba, M., & Nachmias, R. (2010). Comparison of Online Learning Behaviors in School vs. at Home in Terms of Age and Gender Based on Log File Analysis. *... Journal of E-Learning and Learning ...*, *6*. Retrieved from http://www.ijello.org/Volume6/IJELLOv6p305-322Zadok703.pdf

Ben-Zadok, G., Leiba, M., & Nachmias, R. (2011). Drills, Games or Tests? Evaluating Students' Motivation in Different Online Learning Activities, Using Log File Analysis.

Interdisciplinary Journal of E-Learning and Learning Objects, *7*. Retrieved from http://www.ijello.org/Volume7/IJELLOv7p235-248Zadok761.pdf

Brennan, R. L. (2010). Evidence-Centered Assessment Design and the Advanced Placement Program®: A Psychometrician's Perspective. *Applied Measurement in Education*, *23*(4), 392–400. doi:10.1080/08957347.2010.510973

Brook, J. (2003). An initial investigation into the extraction and validity of the quantitative data from web logs of the INFORMS database of online information skills teaching and learning materials. *New Review of Information Networking*, *9*(1), 54–65. doi:10.1080/1361475042000186985

Bruckman, A. (2006). Analysis of log file data to understand behavior and learning in an online community. *The International Handbook of Virtual Learning …*, 1449–1465. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4020-3803-7_58

Chen, D., Wang, Y., & Hung, D. (2009). A journey on refining rules for online discussion: Implications for the design of learning management systems. *Journal of Interactive Learning Research*, *20*, 157–173. Retrieved from http://www.editlib.org/p/26107?nl

Cheung, W. S., Hew, K. F., & Ling Ng, C. S. (2008). Toward an Understanding of Why Students Contribute in Asynchronous Online Discussions. *Journal of Educational Computing Research*, *38*(1), 29–50. doi:10.2190/EC.38.1.b

Cocea, M., & Weibelzahl, S. (2006). Can Log Files Analysis Estimate Learners ' Level of Motivation ? *LWA*. Retrieved from http://www.easy-hub.org/stephan/cocea-abis06.pdf

Cocea, M., & Weibelzahl, S. (2007). Cross-system validation of engagement prediction from log files. *Creating New Learning Experiences on a Global …*, 14–25. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-75195-3_2

Conejo, R., Guzmán, E., Perez-de-la-Cruz, J.-L., & Barros, B. (2013). An empirical study on the quantitative notion of task difficulty. *Expert Systems with Applications*. doi:10.1016/j.eswa.2013.07.084

Dringus, L. (2012). Learning Analytics Considered Harmful. *Journal of Asynchronous Learning Networks*, *16*(3), 87–101. Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ982677

Fisch, S. M., Lesh, R., Motoki, E., Crespo, S., & Melfi, V. (2011). Children's Mathematical Reasoning in Online Games: Can Data Mining Reveal Strategic Thinking? *Child Development Perspectives*, *5*(2), 88–92. doi:10.1111/j.1750-8606.2011.00161.x

Frezzo, D. C., Behrens, J. T., & Mislevy, R. J. (2009). Design Patterns for Learning and Assessment: Facilitating the Introduction of a Complex Simulation-Based Learning Environment into a Community of Instructors. *Journal of Science Education and Technology*, *19*(2), 105–114. doi:10.1007/s10956-009-9192-0

Gobert, J., & Pedro, M. S. (2012). Leveraging educational data mining for real time performance assessment of scientific inquiry skills within microworlds. *... Educational Data ...*, *4*(1). Retrieved from http://www.educationaldatamining.org/JEDM/images/articles/vol4/issue1/GobertEtAlVol4Issue1P153_185.pdf

Haythornthwaite, C., de Laat, M., & Dawson, S. (2013). Introduction to the Special Issue on Learning Analytics. *American Behavioral Scientist*, *57*(10), 1371–1379. doi:10.1177/0002764213498850

He, W. (2012). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*. doi:10.1016/j.chb.2012.07.020

Hershkovitz, A., & Nachmias, R. (2009). Consistency of students pace in online learning. *Educational Data Mining*, 71–80. Retrieved from http://www.educationaldatamining.org/EDM2009/uploads/proceedings/hershkovitz.pdf

Hershkovitz, A., & Nachmias, R. (2011). Online persistence in higher education web-supported courses. *The Internet and Higher Education*, *14*(2), 98–106. doi:10.1016/j.iheduc.2010.08.001

Hrastinski, S. (2008). What is online learner participation? A literature review. *Computers & Education*, *51*(4), 1755–1765. doi:10.1016/j.compedu.2008.05.005

Hrastinski, S. (2009). A theory of online learning as online participation. *Computers & Education*, *52*(1), 78–82. doi:10.1016/j.compedu.2008.06.009

Hung, J., Hsu, Y., & Rice, K. (2012). Integrating Data Mining in Program Evaluation of K-12 Online Education. *Educational Technology & Society*, *15*, 27–41. Retrieved from http://www.ifets.info/journals/15_3/3.pdf

Ioannou, A. (2011). Online collaborative learning: The promise of wikis. *International Journal of Instructional Media*, *38*(3), 213–224. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Collaborative+Learning+:+The+Promise+of+Wikis#1

Jafari, M., SoleymaniSabzchi, F., & Jamali, S. (2013). Extracting Users'Navigational Behavior from Web Log Data: a Survey. *Journal of Computer Sciences and Applications*, *1*(3), 39–45. doi:10.12691/jcsa-1-3-3

Ke, F., & Hoadley, C. (2009). Evaluating online learning communities. *Educational Technology Research and Development*, *57*(4), 487–510. doi:10.1007/s11423-009-9120-2

Keramidas, C. G. (2012). Are Undergraduate Students Ready for Online Learning ? A Comparison of Online and Face-to-Face Sections of a Course. *Rural Special Education Quarterly*, *31*(4), 25–32.

Kerr, D., & Chung, G. (2012). Identifying key features of student performance in educational video games and simulations through cluster analysis. *Journal of Educational Data Mining*, *4*(1), 144–182. Retrieved from http://www.educationaldatamining.org/JEDM/images/articles/vol4/issue1/KerrEtAlVol4Issue1P111_152.pdf

Kupczynski, L., Gibson, A. M., Ice, P., Richardson, J., & Challoo, L. (2011). The impact of frequency on achievement in online courses: a study from south texas university. *Journal of Interactive Online Learning*, *10*(3), 141–149.

Merceron, A., & Yacef, K. (2007). Web-based learning tools: storing usage data makes a difference. *... on IASTED International Conference Web-Based ...*, 104–109. Retrieved from http://learngen.org/~aust/WBE07/557-035.pdf

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 33–45). New York: American Council on Education/Macmillan.

Michinov, N., Brunot, S., Le Bohec, O., Juhel, J., & Delaval, M. (2011). Procrastination, participation, and performance in online learning environments. *Computers & Education*, *56*(1), 243–252. doi:10.1016/j.compedu.2010.07.025

MISLEVY, R., BEHRENS, J., DICERBO, K., & LEVY, R. (n.d.). Design and Discovery in Educational Assessment: Evidence-Centered Design, Psychometrics, and Educational Data

Mining. *Researchnetwork.pearson.com*. Retrieved from http://researchnetwork.pearson.com/wp-content/uploads/MislevyEtAlVol4Issue1P11_48.pdf

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the Structure of Educational Assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*(1), 3–62.

Morris, L. V., Finnegan, C., & Wu, S.-S. (2005). Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education*, *8*(3), 221–231. doi:10.1016/j.iheduc.2005.06.009

Owston, R., Wideman, H., Murphy, J., & Lupshenyuk, D. (2008). Blended teacher professional development: A synthesis of three program evaluations. *The Internet and Higher Education*, *11*(3-4), 201–210. doi:10.1016/j.iheduc.2008.07.003

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaiane, O. R. (2009). Clustering and Sequential Pattern Mining of Online Collaborative Learning Data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(6), 759–772. doi:10.1109/TKDE.2008.138

Rafaeli, S., & Ravid, G. (n.d.). OnLine , Web Based Learning Environment for an Information Systems course : Access logs , Linearity and Performance OnLine , Web Based Learning Environment for an Information Systems course : Access logs , Linearity and Performance.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, *33*(1), 135–146. doi:10.1016/j.eswa.2006.04.005

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368–384. doi:10.1016/j.compedu.2007.05.016

Romero, M., & Barberà, E. (2011). Quality of E-Learners' Time and Learning Performance Beyond Quantitative Time-on-Task. *The International Review of Research in Open and Distance Learning*, *12*(5).

Rovai, A. P. (2002). Development of an instrument to measure classroom community. *The Internet and Higher Education*, *5*(3), 197–211. doi:10.1016/S1096-7516(02)00102-1

Rupp, A., Levy, R., & DiCerbo, K. (n.d.). Putting ECD into practice: The interplay of theory and data in evidence models within a digital learning environment. *... Educational Data ...*. Retrieved from http://researchnetwork.pearson.com/wp-content/uploads/RuppEtAlVol4Issue1P49_110.pdf

Schlager, M. S., Farooq, U., Fusco, J., Schank, P., & Dwyer, N. (2009). Analyzing Online Teacher Networks: Cyber Networks Require Cyber-Research Tools. *Journal of Technology Education*, *60*(1). doi:10.1177/0022487108328487

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline . *American Behavioral Scientist* , *57* (10), 1380–1400. doi:10.1177/0002764213498851

SWEET, S., & RUPP, A. (2012). Using the ECD Framework to Support Evidentiary Reasoning in the Context of a Simulation Study for Detecting Learner Differences in Epistemic Games. *Journal of Educational Data Mining*, *4*(1). Retrieved from http://www.educationaldatamining.org/JEDM/images/articles/vol4/issue1/SweetEtAlVol4Issue1P186_226.pdf

Valsamidis, S., Kontogiannis, S., Kazanidis, I., Theodosiou, T., & Karakos, A. (2012). A Clustering Methodology of Web Log Data for Learning Management Systems. *Journal of Educational ...*, *15*, 154–167. Retrieved from http://www.ifets.info/download_pdf.php?j_id=55&a_id=1236

Warwick, C., Terras, M., Huntington, P., & Pappa, N. (2007). If You Build It Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data. *Literary and Linguistic Computing*, *23*(1), 85–102. doi:10.1093/llc/fqm045

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and TechniquesNo Title*. Burlington: Morgan Kaufmann Publishers.

Xie, K. (2013). What do the numbers say? The influence of motivation and peer feedback on students' behaviour in online discussions. *British Journal of Educational Technology*, *44*(2), 288–301. doi:10.1111/j.1467-8535.2012.01291.x

Adding Value To Online Professional Development Research Through The Use Of Data Mining Techniques With Click-Track Data

Caner Uguz

University of Virginia

Abstract

Online professional development (PD) opportunities and research on them have significantly increased with the development of online resources in education. One of the important attributes of the online environment is the availability of tracking and recording user interaction through click-track data collection. This new type of data has promising implications for online PD environments where our understanding of teacher engagement and the efficiency of the delivery of and interactions with content can be improved by looking at users' online behavior. Many researchers have made use of click-track data in the last two decades but often with rudimentary analysis of the available data. With the advent of Educational Data Mining (EDM) as a methodological framework for using advanced statistical analysis, researchers and PD designers may have at their disposal a better system for using click-track data to elicit meaningful information about teacher interaction in online environments. This paper investigates the added value of click-track data when used with EDM by looking at archival data. The results indicate that valuable insights about researcher assumptions, PD environment design, and efficiency can be achieved with a post-hoc analysis of click track data using EDM, placing it as a viable option for PD research in conjunction with other online and offline methodologies.

Adding Value To Online Professional Development Research Through The Use Of Data Mining Techniques With Click-Track Data

The use of web-based tools for educational purposes has seen a dramatic increase in the last two decades as internet access, web based applications and the app economy has grown (Allen & Seaman, 2013). This growth has also been observed in professional development (PD) training and research as benefits of asynchronous online training has led to many online PD programs (Dede, Ketelhut, Whitehouse, Breit, & McCloskey, 2009).

Research on professional development has contributed to our understanding of how teachers learn and improve their practice through PD (Avalos, 2011). However, not all PD research provides meaningful information because the research designs are not rigorous enough to be able to directly show connections between the PD practice and improvement in student and teacher outcomes. One of the main reasons is that traditionally many PD studies have relied on self-report data from teachers regarding attitudes towards or perceptions of satisfaction of the experience (Lawless & Pellegrino, 2007). Thus, they may lack empirical focus on the process of the PD itself or its outcomes (Desimone, 2009).

Despite its prevalence and continued growth, online PD has not contributed markedly to measures of PD effectiveness (Reeves & Pedulla, 2013). One reason is that studies of online PD utilize the same measures of effectiveness as in-person PD, such as qualitative data to understand interactions online like the sense of community among participants, or evaluate the effectiveness of the online learning environment (Cheung & Hew, 2011; Chitanana, 2012; Hew & Knapczyk, 2007; Parker, Maor, & Herrington, 2013). In these studies authors use mainly textual information and make use of the online environment as a means for intervention and a data collection point.

While using qualitative data is very helpful, an online learning environment allows for extending data collection options to get a better sense of what the users are doing. They do this by recording the user behaviors as they interact with the online learning environment components. While this is referenced in the literature in many ways, such as logs and log data, this paper will use the term click-track data to denote that it goes beyond server logs and looks at details of page views and actions within the page.

Click-track data have been available within web applications since the early 1990s but the trend to use this data for educational research is relatively new and has been developing mainly through data mining, which is the analysis of these data to explore and uncover relationships and patterns that may not have been specified and planned for ahead of time within a data set. While the interest of online learning in professional development has grown simultaneously with the interest in analyzing quantitative data for online learning environments, there are very few instances of studies where data mining is being used to analyze and improve online PD environments.

This paper aims to provide further investigation in this area through post-hoc applications of data mining techniques to data from an already completed online PD program. The goal is to demonstrate the advantages and constraints of using data mining efforts in online PD environments so as to speculate how data mining with click-track data might add value to the design and research stages of an online PD environment.

Click-track Data as a Support To Online Professional Development Implementation

In the last two decades our understanding of best practices in professional development has been informed by numerous studies that show PD is most effective when it is provided as an active and continuous process that ties to the context of the classroom and engages communities of teachers (Webster-Wright, 2009). PD environments that include activities where teachers can engage with the PD content through reading, experimenting, reflecting and collaborating are more likely to produce positive results (Kwakman, 2003). However, organizing PD into highly structured content may not yield the best results and delivery models should diversify to include continuous work that is situated in teachers' daily work (Knight, 2002). When it is possible, teachers benefit more from a PD experience that can be personalized to their needs and reflect their preferences (Gibson & Brooks, 2013). Further indicators show that many countries are moving away from top-down PD practices into more dynamic and collaboration-based mentorship and professional learning communities that require more customized solutions compared to typical PD (Collinson et al., 2009). It is also important for teachers participating in PD environments to collaborate and reflect on the learning material through dialogue with colleagues (Snow-Gerono, 2005). Thus, despite the research on face-to-face PD environments not usually being done in a systematic manner, and lacking ongoing evaluation following the PD experiences (Muijs & Lindsay, 2008), it does suggests some useful actions to take in designing PD. However, even less clear is how to implement these principles in online PD, as research on PD in online learning environments is very limited.

The existing online PD research suggests that making it relevant and providing a variety of options that teachers have some control over is part of what contributes to the success of online PD environments (Storandt, Dossin, & Lacher, 2012). Specifically, through its digital method of delivery online PD can provide personalization and evaluation more easily than face-to-face environments. Online PD allows for user interaction information gained by click-track data as a novel tool to understand potential relationships between configuring the learning space, teachers' interactions in it, and outcomes of interest.

Contributions of Click-track Data with Educational Data Mining to Online PD Research

Several studies using online PD environments or other online educational environments for practitioners have explored the advantages offered by using educational data mining (EDM) methodologies to investigate online PD environments. Next we outline the types of advantages illustrated through recent research activities.

Triangulate user information. Click-track data provides reliable records of some level of user participation by recording the exact moments individual users have visited pages online. In many aspects of online PD environments researchers can check user participation information gained from log records against self-report surveys or qualitative data to triangulate their results. Researchers can also triangulate their assumptions about PD design by looking at whether a planned path of web page visits actually take place (Black, Dawson, & Priem, 2008). For instance, if the information to be received through the online environment is spread across many pages, researchers can see how the participants actually traversed these pages and either check if participants who claimed to make significant use of the online environment actually did so or compare that to their assumptions about what the participants would have done.

Provide accurate, detailed records. While teachers may not always remember how they used the system, web logs provide a highly accurate and in-depth recording of their behaviors. These data can then be used for a multitude of purposes, such as understanding shared learner characteristics, evaluating course-related goals, or improving features based on how learners interact with the website (Bruckman, 2006; Hung, Hsu, & Rice, 2012; Klassen & Smith, 2004). A wide variety of questions can be answered with this approach, such as "Are teachers who use the website within school hours spend more time on it than those accessing outside of work?", "Do teachers seemingly skim the material when staying for short periods?", "Do teachers find the relevant information through direct, circuitous or highly varied routes?" and, depending upon demographic information collected, questions such as "Are younger teachers more inclined to visit online resources?"

Scale data collection and analysis. Using qualitative methods such as interviews or content analysis to analyze participants' input as they interact online can prove to be a costly method. Being able to use log data in place of collecting qualitative data is scalable since the work of writing the code to analyze the data can require less time to complete than textual analysis, and need only be done once. This scaling means it may be possible to have more participants in a study and also that more information can be gathered from a small group (Fisch, Lesh, Motoki, Crespo, & Melfi, 2011).

Get real-time feedback. Logging of user behavior happens instantly, so analysis based on these data can be accumulated and analyzed in real time for constructive use, such as providing feedback on medical students as they work with a surgery simulation (Kennedy, Ioannou, Zhou, Bailey, & O'Leary, 2013), or building recommender engines for e-learning settings that provide suggestions for additional or different material for learning based on learner performance (Zaiane, 2002).

Personalize for user needs. Either in real time or after the fact, you can use the log data to provide personalized paths and information to users. This is similar to "choose your own adventure" books, where researchers can customize the user experience based on user input or provide options to students based on their expressed or latent needs (Chen, 2008; Köck & Paramythis, 2011; Lin, Yeh, Hung, & Chang, 2013; Zorrilla, Menasalvas, Marín, Mora, & Segovia, 2005).

Predict outcomes or behaviors. Since log files often provide a large amount of interaction data, this allows researchers to run predictive algorithms that associate behaviors with outcomes (Cocea & Weibelzahl, 2007; Hung & Zhang, 2008). Predictive algorithms work when there is an available external outcome variable that is not measured with click-track data. For instance, if you know which teachers generally include more technology in their classes, you can identify online behaviors of successful teachers and use this data to predict which teachers are more likely to include technology in their classes based on their online behavior. Such a relationship would be correlational, not causal, but would still provide strong methods to classify new data.

Applying EDM to Online PD Research

Such research illustrates how EDM has been useful in online learning environments and most importantly we see how these advantages can apply to online PD research, specifically to improve the following two main areas that have so far been challenging to resolve.

Increasing Assessment and Instructional Support. Using click-track data and data mining strategies researchers can aim for a much more nuanced picture of the user interaction online and how the user experience is shaped as they interact with different aspects of the online PD component. Researchers can also keep track of how well the participants are doing compared to the research goals. For instance, if there is an expected path or visits to certain aspects of the website, or utilization of a key component in the online pages the researchers can analyze the extent to which these interactions happen. These analyses can then be used to provide additional support to users. Or, if participants are not utilizing an online component the PD designers consider to be important they may receive hints about where to find or use them, depending on upon what their specific use pattern to that moment. Collecting and analyzing click-track information can provide this support instantaneously and in individualized ways (Agudo-Peregrina, 2013; Anaya & Boticario, 2011; Lee, 2012).

Performing Iterative Refinements to Design and Structure. One of the main challenges of online PD research design is to make sure that the designed online environment works in the way intended. Since it is often costly to design and deploy these environments researchers could be well served by testing them in pilot studies with click-track data. The results of these pilot studies would provide important information that would give not only usability feedback but also feedback about whether the intervention is working in the way intended. Even though such analyses with click-track data can be done on the fly, in terms of data analysis it would not be desirable to change the intervention variables while the study is ongoing. Yet, the use of click track data in either pilot studies or design-based research would allow developers to improve an online PD environment so that the intervention is favorably designed to allow promoting the desired outcomes for participants and the research alike (Romero & Ventura, 2011; Wiebe, Branoff, & Shreve, 2011).

Purpose and Research Goals

This paper presents a case study demonstrating how EDM can be applied to archival data from an online PD study. It illustrates how EDM methods can provide insights into how the participants utilized the resources in the online PD system to gain meaningful insights for improving the design and structure of it for future iterations.

Background On The Online Professional Development That Produced The Data Set

The raw click-track data used for this paper was collected as part of an online case-based learning environment called Educational Theory into Practice Software (ETIPS). This web-based application coordinates users' logging in, being directed to the case introduction and questions created by their instructor, and then viewing the website and intranet of a school to find and take notes on the information they deem is relevant to answer the questions and complete the case. There are nine of these hypothetical, yet realistic schools (three each at elementary, middle and high schools levels with a school from each level set in a rural, urban, and suburban setting). The detailed information within the pages of these cases give users insights into the school's social and administrative structure, its curriculum and instruction, professional development, and its uses of technology through "public facing" pages, which mimic a typical website of a school open to any visitor, as well as "intranet" pages, with student testing and teacher performance data like what might only be available to school personnel. For example, pages might detail the computer access options such as labs and their availability, the level of technology spending within the bigger budget of the school, student discipline or achievement (See Figure 1 for an example page).

ETIPS cases were used in a study of educational administration students learning leadership decision-making skills while in a course required for administrative licensure at one of the eight participating higher education institutions. The study was conducted within the 2008-2009 academic year, and the 130 research participants who were the students of nine faculty members recruited for participation from a statewide network of schools of education. The faculty members' higher education institutions varied by location type (e.g., urban, rural), student size, selectivity of admission criteria, and the amount of technology typically used during instruction. Six of the eight institutions were publicly funded, and all were located in a southeastern Atlantic state.

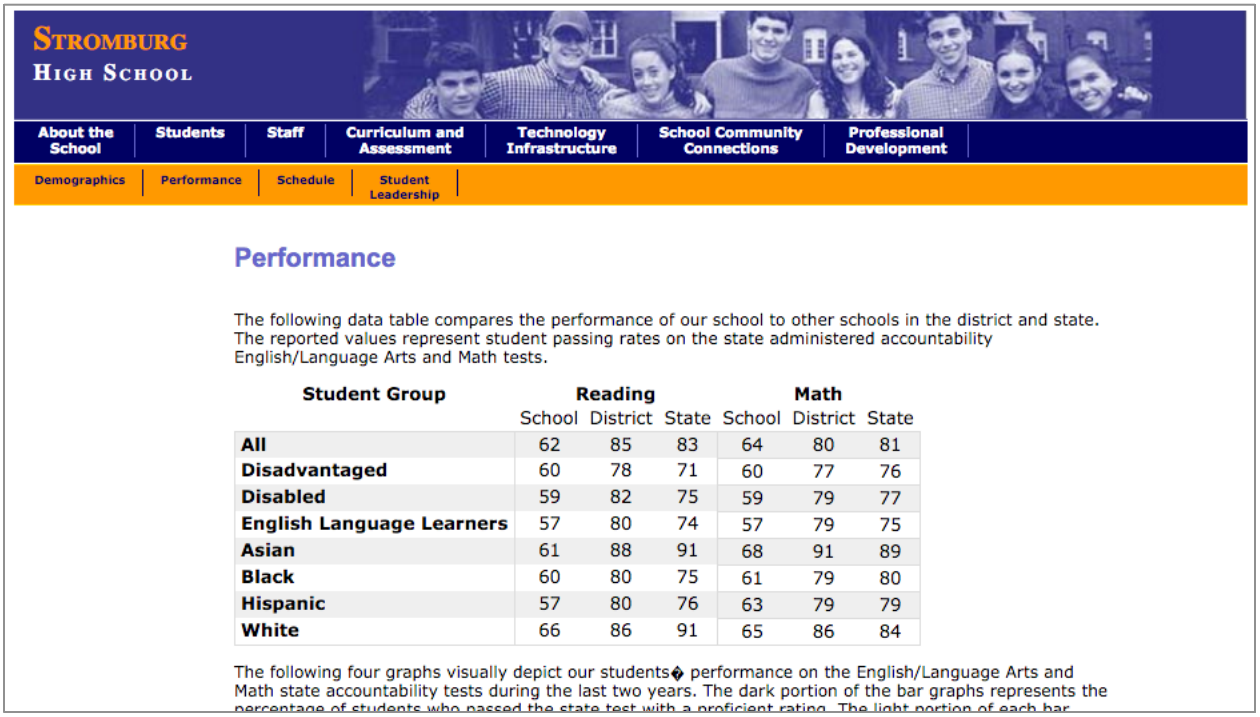
In an administrative licensure course selected by each of the nine faculty members, their students who consented to participate in the study were assigned to complete the same three cases at the ETIPS website. Study participants were first given a pre-intervention survey that gathered information about their gender, academic status, coursework, experience in using cases before, how familiar they were with technology, and perceived readiness to be an administrator, as well as a measure of their self-efficacy about leadership decision making skills. After the three cases a post-intervention survey asked them to evaluate the case experience and their decision-making self-efficacy, which was compared to pre-intervention results.

In each case the participants were presented with a situation and asked to respond to questions designed to elicit their decision-making skills about leadership situations K-12 school leaders might face. Their essay responses to these questions drew on the data of the school in which the case was set and also the ability to apply course content as they were guided through a four-step decision making process. For example, in regards to the overarching situation posed in the case introduction, the question for step one asked: “Consider what is going on at the school. Generate 3-5 explanations that you think could account for this.” As they completed each step’s response the participants were also asked to rate how confident and how certain they were of their answers.

The open-ended essay responses were scored using a 0-3 point scale by multiple researchers who reached a .77 inter-rater reliability score. Student’s high essay scores indicated thorough use of information from the case as well as attention to carrying out well that specific step. The goal of the main analysis of that study was to measure the impact of practicing (i.e., through the three case assignment) on these prospective leaders’ decision-making skills as well as on their confidence about these skills and the certainty of their answers.

Because the online cases were set within a website that required a login, it was possible to collect as click-track data the pages a participant visited within the school website and intranet in which a case was set. The school's website and intranet used navigation and layout components reflecting typical design norms of menus and sub-menus. Participants had to purposefully select what categories of information they deemed relevant and click the links to read content; the path through the material was not prescribed nor was the participants' time to complete a case limited in any way. They also had a notepad available within the website where the notes they took about the specific page they were visiting was saved for them in the ETIPS system. Throughout their work with the cases users' actions were recorded as click-track data, including individual page, user identifier, the start time when they accessed the page and the notes taken within that page.

The goals of the study that utilized these ETIPS cases were related to leadership decision-making skill and self-efficacy. While the click-track data were captured, they were not analyzed at that time, however the ETIPS study's rich source of raw click-track data can now be analyzed post-hoc to ask what value it can add to understanding learners' interactions within the online environment and scaffolding the development of their decision-making skills. In this demonstration case study we will use this set of raw data to allow us to address the research questions we pose about how click-track data can add value to making meaning about learners' work in an online PD experience.


STROMBURG HIGH SCHOOL

About the School | Students | Staff | Curriculum and Assessment | Technology Infrastructure | School Community Connections | Professional Development

Demographics | Performance | Schedule | Student Leadership

Performance

The following data table compares the performance of our school to other schools in the district and state. The reported values represent student passing rates on the state administered accountability English/Language Arts and Math tests.

Student Group	Reading			Math		
	School	District	State	School	District	State
All	62	85	83	64	80	81
Disadvantaged	60	78	71	60	77	76
Disabled	59	82	75	59	79	77
English Language Learners	57	80	74	57	79	75
Asian	61	88	91	68	91	89
Black	60	80	75	61	79	80
Hispanic	57	80	76	63	79	79
White	66	86	91	65	86	84

The following four graphs visually depict our students� performance on the English/Language Arts and Math state accountability tests during the last two years. The dark portion of the bar graphs represents the percentage of students who passed the state test with a proficient rating. The light portion of each bar

Figure 1. An example web page from a simulated school website used in ETIPS.

Research Questions

While educational data mining, or any exploratory statistical analysis, can be used to investigate a large number of avenues, this study applies EDM to ETIPS archival data to illustrate the potential effectiveness of EDM to improve online PD learning environments. Specifically, this paper focuses on three research questions for which we also discuss related EDM techniques.

Research Question 1: Are ETIPS researchers' design assumptions about relevant pages confirmed by actual use?

Instructional designers build online learning experiences that are often used for research purposes, weaving particular constructs of interest into these experiences. For instance, some designers are interested in making the experience engaging, others would like a sense of community to be reflected while others aim for important information being found effectively. Designers need metrics and analysis to help them understand how that construct played out in the user's experience.

For ETIPS researchers one of the important constructs in relation to the online experience was the concept of "relevancy" of pages. The ETIPS simulated school websites were designed to include pages that the ETIPS researchers believed included essential information for participants to be able to answer certain questions, which they deemed as "highly relevant". For instance a case about teachers' technology integration would be best informed if they visited the pages about the school's existing technology capabilities and efforts on technology support. The click-track data recorded all visited pages including the researchers' previously defined highly relevant pages. Being able to recognize pages deemed relevant prior to the study within the click-track data allowed us to differentiate user clicks on the variable of relevancy.

This research question aims to triangulate the research design with click-track data to find out if participants who gave high quality responses to the essay questions in fact did visit the relevant pages or how important the relevant pages were in getting a high score on the essay questions. If the participants who did well on the essay questions did not overwhelmingly visit the pages marked as relevant, this does not mean that the pages may not in fact be relevant but rather that the relevancy matters less than the ETIPS researchers may have assumed. Knowing this information could allow researchers to change the user directions or emphasis in support materials for the cases.

Research Question 2: Which pages are participants making use of and how are they navigating these pages?

A look at the aggregated number of hits does not provide a lot of information about the typical user experience with pages or allow for much insight into that experience. However, if we can see patterns in page views and infer user behaviors from them, it may explain more about participants' use of the online PD environment.

Exploratory analyses to answer this question included the use a Sankey diagram to better understand how participants move between pages. Another method looked at quantitative page characteristics (i.e., derived from information about visits but also content characteristics such as length, readability score, sentiment score, and so on) and see which characteristics of the pages used are associated with participants' higher essay scores. We also determine which specific pages were most related to higher scores, in a sense calculating which pages should have been marked as relevant based not on the content meaning but their relationship with high quality responses to the essay questions.

While all the these analyses support sense-making about the pages learners elected to view and their importance in relation to their essay scores, exploratory analysis can also shed insight on how researchers could nudge people towards seeing the relevant pages. With the help of a time series user experience chart we provide a threshold for researchers to designate a point where participants may need help. This could be implemented as a feature in future redesigns of the ETIPS simulated school websites.

Research Question 3: Based on their online behavior which participant characteristics are indicators of high quality essay scores?

Click-track data can be used to discover different groupings of learners that could potentially be helpful for gain insight in participant characteristics not yet known to impact performance. Identifying cluster behaviors to investigate patterns of groups addresses how participants differ from each other and if there are characteristics that separate them into distinct groups based on their online experiences, which might allow changes to the online PD environment that could eliminate any problematic patterns.

Methods

Participants

As described above, the participants in this study are the 130 teachers preparing for administrative licensure in courses taught by the nine participating professors. The faculty member of each teacher is noted, although they did not complete the online cases themselves and thus are not part of the click-track data to be analyzed here.

Data

Also as described earlier, the archival data for this study was collected as click-track logs while the teachers completed cases within the ETIPS website. The dates when participants interacted with the website depended on when they were given the case assignment by their respective instructors. Participants visited the cases outside of class as homework at a time of their own choosing; they did not have access to their instructor as they worked on the case, except if they emailed him or her or waited to ask questions in class. The participants' exact setting where they accessed the website is unknown and it is not possible to use this as a variable in this analysis.

Data Preparation

While the researchers in the original ETIPS study had already cleaned the essay score and survey data, for a few instances where certain variables in teacher data were missing they received a N/A designation and the omissions were handled in the analysis. Mostly the non-available data were automatically filled in with averages or ignored and missing data or rows were not deleted manually. The click-track data was combed to make sure that only data from participants in the 2008-2009 group were included in the analysis. Click-track data did not include any missing data, and every row of the click-track data was assigned a relevancy score or not after referencing if the actual page visited was one the ETIPS researchers' designated as highly relevant.

Calculating and Formatting Variables

The majority of data preparation work included combining existing data or creating calculated variables from the original data. Three types of data were used in this research. One type was aggregate information from click-track data to build meaningful variables that represented sessions or the user experience, such as total number of visits per user. The second type of data was calculated based on the content of the school or intranet pages used in the cases. These calculations were made from the exact content of the pages as they were used in the study during the original period of research. The third type of data was the participants' surveys and essay scores, which were made available from the researchers of the original ETIPS study. Only one new variable was produced from this data: total self-efficacy, which was the sum of all individual self-efficacy question scores. Next, we provide further detail for each type of data.

1. Click-track data. This data set included all participants' individual page views. An automated programming script recorded each time a user requested a page from the server. The recorded data included the click-track data variables listed in Table 1 below.

Table 1

Click-track Data Variables

Name	**Description**	**Example Value**
Student	Unique ID of the participant	495
Assignment	The assignment number for the online case (the combination of a case and a faculty member got a unique assignment number)	456
First assignment	Whether or not this assignment was the first one of the three for the participant	1 or 0
Case	The number for the online case	345
Relevancy	Whether the page visited was designed as relevant or not for that case by ETIPS researchers	1 or 0
URL	The URL link of the page visited	Home.html
Clicktime	The date and time of the click	10/2/08 10:35
Note	The total notes the user has in the database for this URL (not the notes at the time of the click-track recording).	"There is no gifted program at Stromburg, but the instructional policy indicated that one will be

		provided”

This data set included the click-track information for each participant’s work on only the first and the third cases that they completed, as the second case’s essays were not scored in the original ETIPS research study. Both of these cases were based in the hypothetical Stromburg High School and thus the click track data referenced the same online environment, portraying this hypothetical school’s website and intranet. For both Case 1 and 3, some aggregate variables were calculated from each users’ click-track data to provide sum values representing all of the user’s activity. Details of how these variables were calculated are provided in Table 2.

Table 2

Calculated Click-track Data Variables, By Case and Participant

Name	**Description**	**Example Value**
Totalhits	Total number of times a participant has visited a page within a case	35
TotalLow	Number of hits that are marked as LOW relevancy	19
Totalhigh	Number of hits that are marked as HIGH relevancy	16
TotalUniqueLow	Number of hits that are marked as LOW relevancy without repeat counts	13
TotalUniqueHigh	Number of hits that are marked as HIGH relevancy without repeat counts	8
TotalNoteLength	Number of characters in total notes taken during that case	783

RelevancyRatio	Ratio of TotalHigh to Totalhits	0.45
Unique RelevancyRatio	Ratio of TotalUniqueHigh to total of unique high and low counts	0.38

2. Page content data. Some of the analysis focused on the attributes of the pages themselves and how they potentially affect the essay score outcomes. The content data was calculated for each page instead of per assignment for participants. In other words, the individual pages (e.g., home.html) are the rows and calculations are aggregated across all participants. The calculated page content variables are shown in Table 3.

Table 3

Calculated Page Content Variables

Name	**Description**	**What it is Based on, or Calculated With**
relevancy	Whether the page was designated as relevant or not	ETIPS researchers' data
duration	the average time participants stayed on the page	Click-track data
hits	the total number of views for this page across all participants	Click-track data
length	the length of page in words	Simulated web page content
last page	the number of times this page was the last page visited for the participant's session	Click-track data
readability	readability score of the page content	Readability Test Tool (http://www.webpagefx.com/tools/read-

		able/) ran on each page content
sentiment	sentiment analysis of the page content	Sentiment Analysis Tool (http://werfamous.com/sentimentanalyzer/) ran on each page content
read time difference	The difference between how much time participants spent on the page versus how estimates of length of time it to typically read	Compared to reading time estimate based on readability test tool
weighted scores	The score of the participants weighed by how many hits this page had for that score	Click-track and participant data

3. Participant data. The information about participants included all the survey results and essay scores previously gathered by researchers in the original ETIPS study. Only a small subset of these variables for each participant were utilized for the analysis of the click-track data because the original data set contained a large number of measures not related to the online experience including some demographic information and responses to surveys or not of interest for this analysis.

Table 4

Participant Data Variables

Name	**Description**	**Example value, and range**
degreeLevel	The current degree level of participant from Endorsement to PhD	1-4
requiredCourse	Whether this course was required for the participant or elective	1 or 0
yearsTaught	Number of total years to date worked as a teacher	8

readinessLevel	Self reported readiness to serve as administrator	1-4
casebaseFamiliarity	Whether any of their instructors before used case based teaching	1-4
skillwithTechnology	Self reported skill level with technology	0-4
totalhits	Total number of visits for this participant	62
totalsessions	Total number of sessions, (consecutive hits that are less than 20 minutes apart are one session)	6
totalTime	Total time in minutes spent on the pages	120
averageTime	Total time divided by total hits	1.9
noteLength	Total length of all the notes participant took	1290
noteLengthperPage	Total note length divided by number of hits	20.8
pageComplexity	Average complexity of pages visited by this participant	24.03428571
pageLength	Average length of pages visited by this participant	558.2
pageSentiment	Average sentiment of the pages visited by this participant	9.085714286
beforeSub	Before score subtotal, an external score given to faculty about how well they implemented the topic	0-9
preSETotal	Total score for the self efficacy survey given to this participant before the study	0-72
totalQ1	The essay score for this participant on question one of the qualitative essay	2
scoreClass	A separation of the essay score into three classes, Low: 0-2,Medium: 3-4, High: 5-9	Low, Med, High

Summary. In research using traditional statistical methods, all the variables and data sets described above may not be combined together into one analysis since they are not tied together in terms of a hypothesis. However, in post-hoc exploratory research making use of data mining, all pieces are potentially valuable in that they might provide insight, in terms of click-track data, into different segments of the characteristics of the participants, of the pages that are being viewed, or of the page view experience. The variables described above are believed by the authors to be potentially informative for the original research study and thus were selected to further the understanding of the ETIPS online learning environment, and, might allow us to see some general implications of EDM in online PD, by serving as a demonstrative case study of its uses.

Writing Code for Analysis

Writing code to run the calculations was required to build the new variables described in tables 2 and 3. While such calculations could be done by hand, it is significantly faster and less error prone to use custom computer scripts. Several scripting languages can be utilized towards this purpose. JavaScript was chosen because of the first author's familiarity in using it. The code includes steps to load the data, run calculations on the data to build new rows and provide a comma separated value (CSV) file that can be saved and used with data analysis tools. The entire source code and data used in this step publicly available at the following online repository: https://github.com/caneruguz/thesis.

Software Used For Data Analysis

After the data were cleaned, new variables were calculated and prepared for analysis in a tabular file so they can be analyzed with statistical software. R Studio (https://www.rstudio.com/), a free multi-platform data analysis software run on the open source R programming language, was used for the descriptive statistics and correlations. The data mining analysis, which included cluster, classification and linear regression, were done using the Weka data mining software (http://www.cs.waikato.ac.nz/ml/weka/), developed by University of

New Zealand and available for free use. The Sankey diagram shown in the results was created with the library available from Google (https://developers.google.com/chart/interactive/docs/gallery/sankey) with a local JavaScript implementation available in the script repository indicated above.

Results

For each of the three research questions we first we present the results to answer it, and then discuss some overall results of using EDM for the analysis of such research questions.

RQ1: Constructs of Interest to the Instructional Designer

One of the core constructs of interest to the researchers in the original ETIPS study was the concept of "relevancy". The 73 pages organized as a school website and its intranet had just 8-9 pages designated as relevant, meaning that the original researchers felt they contained information required to be able to respond to the specific case questions. We first created metrics from log files to represent users' search through the 73 pages and then correlated them to essay scores to investigate how learners' use of case material was related to their case performance (see Table 6).

Click-track data show that in the first and third cases users visited an average of 81% and 82% of the relevant pages respectively, meaning almost all participants visited for instance between 7-8 of the 9 possible relevant pages. Data from both cases show that users who visited more of the relevant pages did have slightly higher scores for the case's first essay question. Page visits can be separated into two variables for a more nuanced understanding of user behavior. The first relevancy variable is "total high relevancy page visits" (totalHigh, see Table 2) which is all the visits to any relevant page. The second is the "total unique high relevancy page visits" (totalUniqueHigh, see Table 2), which omits from its calculation return visits to the same page. The reason for this difference is to highlight diversity in page visits and differentiate someone who visited a single page many times with another who visited many different kind of pages. Page diversity is meaningful here since the goal is to search and find relevant information. Both of these relevancy variables correlated with users' essay scores at the same low levels (0.16, $p < 0.05$).

Student essay scores showed a low but statistically significant correlation with total page visits (totalHits, see Table 2) (0.25, $p < 0.05$) suggesting that the more pages participants visited the higher their scores were. Figure 2 shows the relationship of these variables.

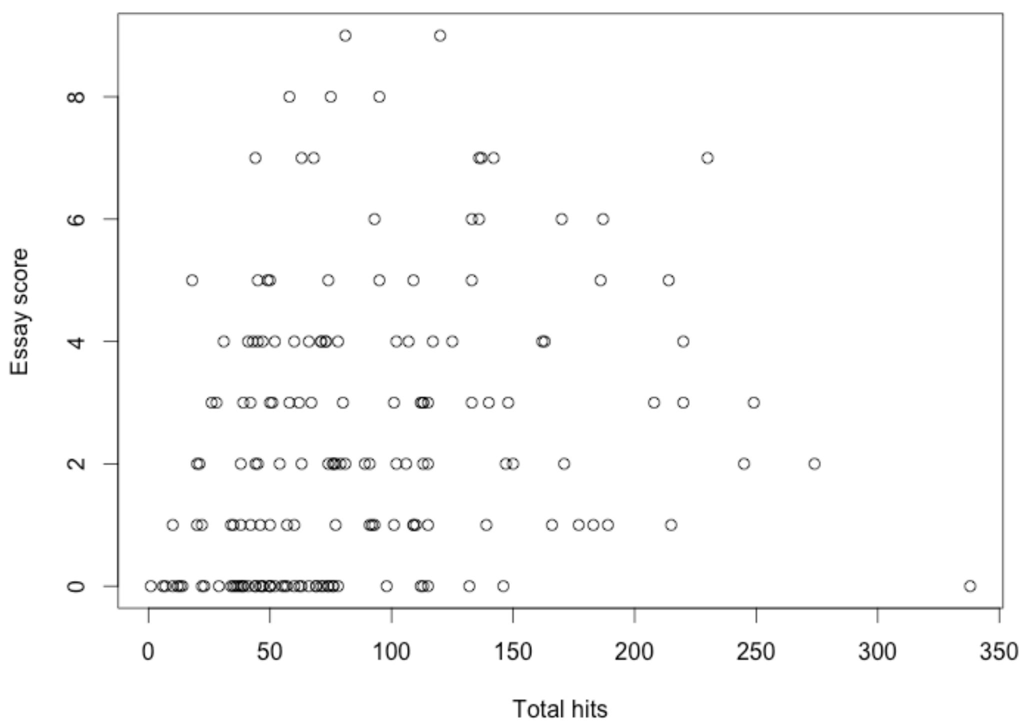


Figure 2. Chart of Total hits by essay scores. Each dot indicates a participant

Similarly, total time spent on the website also had a small, but statistically significant, correlation with the essay score (0.2, $p < 0.05$). On the other hand, the amount of notes participants took during their visits (totalNoteLength, see Table 2) resulted in a very low and not statistically significant correlation with essay scores (0.09, $p > 0.05$).

Table 6

Correlations Between Calculated Click-track Data Variables, By Case and Participant, and Essay Score

	totalHigh	totalUniqueHigh	totalHits	totalTime	totalNoteLength
Step1QTotal	0.16*	0.16*	0.25*	0.2*	0.09

*p < 0.05

The results show that spending more time on the website (totalTime) and visiting more pages (totalHits) did correlate with higher scores in the first essay question. However spending more time on the pages designated by the original researchers as relevant was only weakly related to users doing well on the case's first essay question.

These findings may cast some doubt on the validity of the relevancy construct but the problems may lie not necessarily on the construct, but the way measurements are made as well. While this finding suggests that ETIPS researchers' conception of relevant pages needs further scrutiny —in that relevancy was not confirmed by actual use as indicated by essay scores, it showcases the post-hoc analysis of click-track data as a methodology that allows for such investigations. Where online PD researchers have enough budget and time for a pilot project, using click-track data and basic data analysis methods such as correlation they could find out the strength of such assumptions. The implication is that click-track data is capable of validating assumptions related to the online experience of participants by checking those assumptions with actual usage.

RQ2: Understanding How Users Experience the Website

Building constructs for how learners should utilize the website is a very useful method for designing it but their online experiences might be richer or even unexpected as compared to the instructional designer's or researcher's assumptions about what will happen. In this case study, the exploratory nature of the data mining methods helped us to investigate participant behavior online without preconceived notions based on constructs. This offers an alternate means to make sense of the participant experience and uncover why constructs of interest did or did not work with the users. In the ETIPS example participants' limits of autonomy to travel through the case materials (i.e., school's website and intranet) and make use of the resources provided was limited to the hyperlinking options within the website. Because the navigation of online PD environments are likely built for its custom purposes, looking at how users navigate the space can provide feedback to its designers and researchers.

Navigation menu. Considering the structure of the ETIPS school's website navigation menu and the weak correlations between case performance and the relevancy construct, we may hypothesize that participants navigated through the website in the precise order that the menu structure provided for them. If this has been their path, it would explain why relevancy is not an important aspect since their experience would proceed step-by-step through the entire menu, with all users' searches rather similar. One of the ways to empirically test this is through a Sankey diagram where we look at how participants moved between two pages.

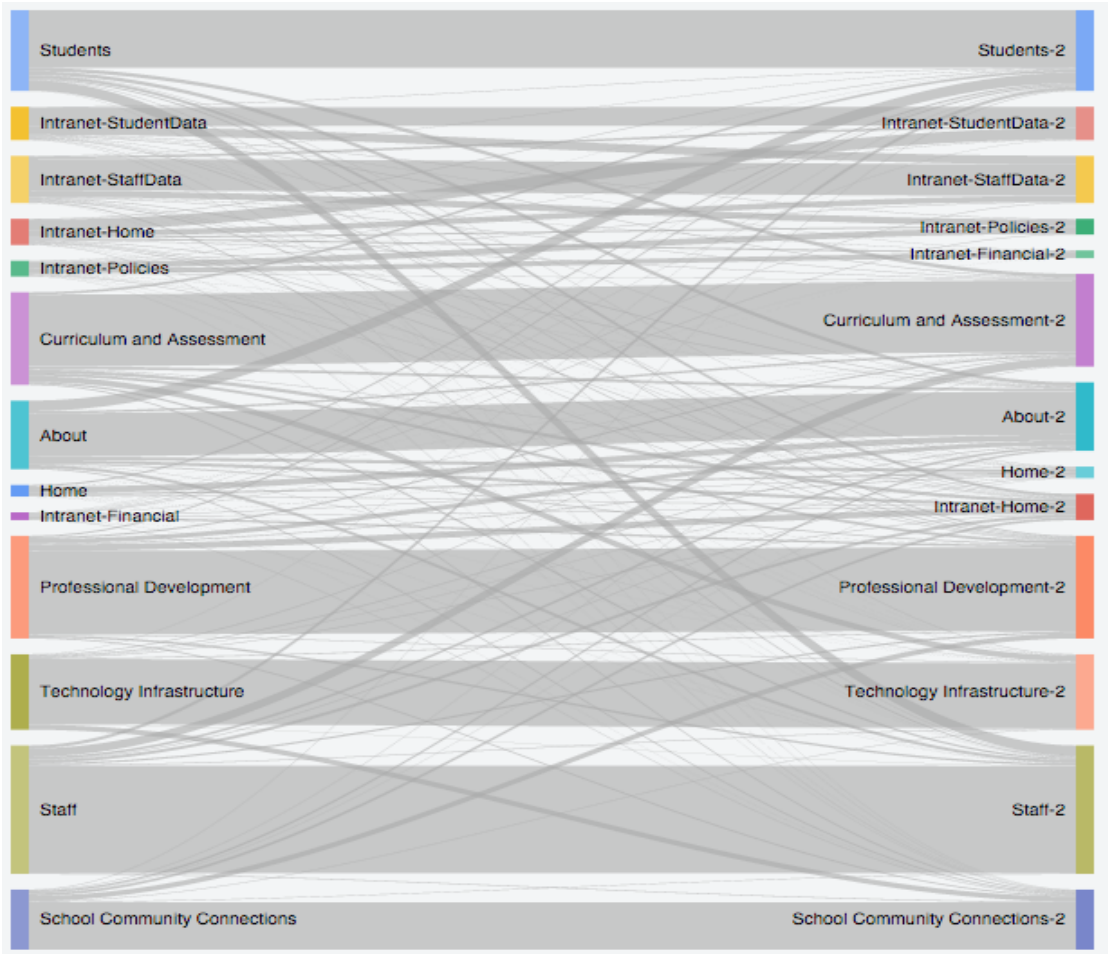


Figure 3. Sankey diagram showing movement between pages. Each line represents a visit between pages on the left and right, the more the numbers of visits, the thicker the connecting lines.

Figure 3 shows, in aggregate, the travel between groups of pages. There is a large amount of traffic between pages within the same menu categories, such as within the Students or Staff menu items. This denotes that users largely navigated using the sub-menu items and not so much in between categories of menu items. When we took out navigation into the same group and only look at the times when they moved between categories of menu items we see how the users moved in between menu categories as seen in Figure 4.

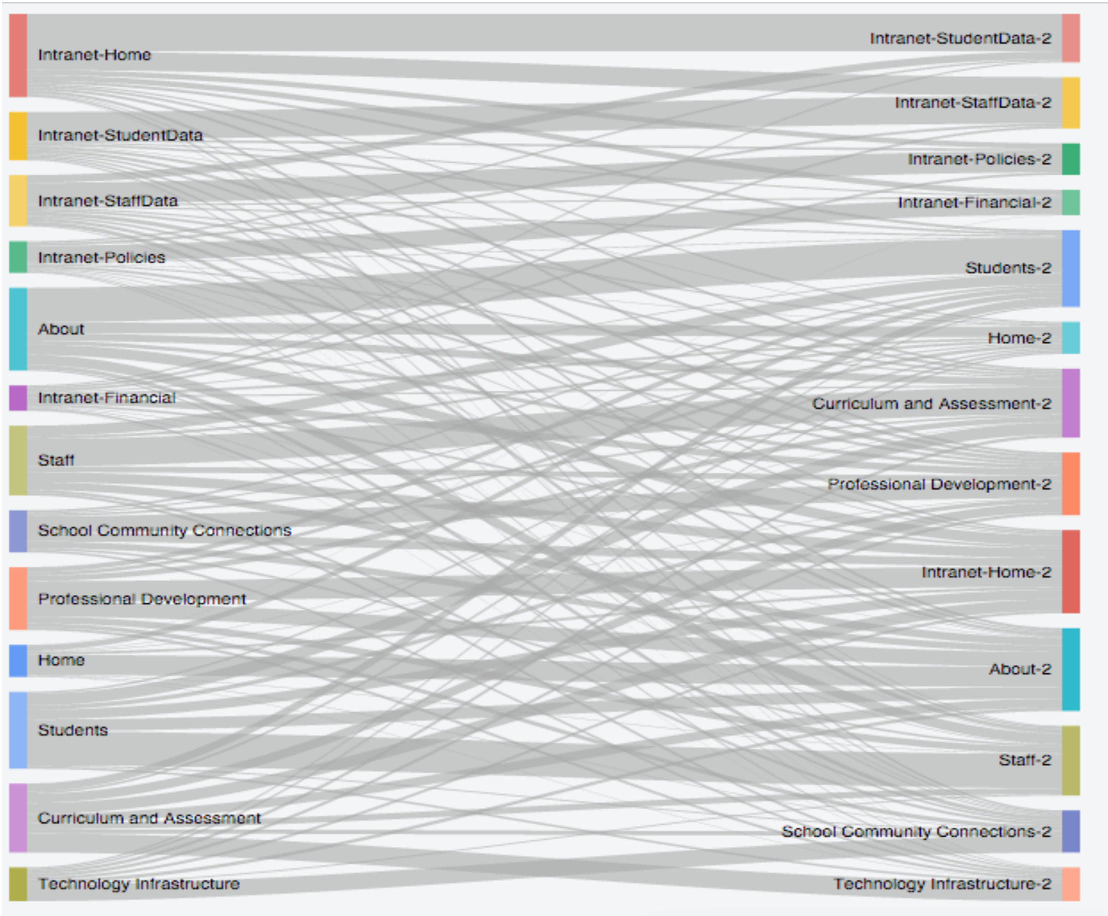


Figure 4. Sankey diagram without navigation into same menu category (i.e., eliminating sub-menu navigation).

In Figure 4 we see that the navigation in between menu categories has mainly been between adjacent categories and there is a very strong movement from what users saw as the left to the right side of the menu. That is, participants moved from Home page predominantly to About and then to Students and from Students mainly to Staff, as seen in the actual navigation interface of the school website portrayed in Figure 5, these menu links are side by side, suggesting a systematic and orderly move by most participants.

Figure 5. Screenshot of the navigation for the simulated school website. Each menu item shown has 3 to 6 sub-menu items under it (which are not shown in this image).

The above investigation gives a sense of why site navigation could be giving very little emphasis on finding relevant pages. Participants are given one main navigation scheme as the only way to move between pages. On each page's contents there are few hyperlinks that can send users contextually across the website to other pages. Instead, users are required to make use of the highly hierarchical menu and sub-menu navigation. This resulted in the traversal across the hierarchy that the Sankey diagrams illustrate in Figures 2 and 3. Without a large number of hyperlinks connecting pages in other ways, summary sections with pertinent links, or other diverse contextual navigations solutions, the participants don't have any other means to quickly get to relevant data. They need to instead comb through all the content first to be able to get a sense of where information is provided. Once participants traverse all the pages in this method they apparently had little incentive to visit pages in a different order, since their navigation strategy allowed them to link-by-link acquire the information needed to complete the case. While this structure may be beneficial for long-term use, it failed to help us differentiate users and their use patterns in this homework assignment.

Page attributes. Page attributes were another area of investigation, in addition to navigation, to try to discern which pages might be more effective in contributing to the learners' online PD experience. This required generating the new variables defined in Table 3 to search for more insight about both what each ETIPS case page is like in terms of content. We then used these new variables to explore other ways to determine which page attributes were associated with higher essay scores.

Which variables were important? We are mainly trying to understand which attributes of the pages were most pertinent to participants' success in answering the first essay question and used that essay score as the overall student performance outcome variable. However with individual pages it is not appropriate to use an overall essay score that was created after visits to many different pages. To be able to compare the performance of a page with other pages we calculated a new weighted score for each page. This score is calculated by multiplying each student score with what the page in question represents as the percentage of total pages that student visited. This new score represents the amount that an individual page contributed to a student's overall essay score. For instance if "home.html" was visited by Student A ten times out of her total 60 visits and Student A had a score of 3 on the essay the weighted score for "home.html" for Student A is 3*(10/60) = 0.5. The overall mean of all scores calculated like this for all participants forms the average weighted score of the page.

The initial analysis looked at correlations of the variables described in Table 3 with the outcome variable of weighted scores. Hits and duration had a strong positive correlation with the weighted essay scores for the page, meaning that the longer and more times a page was visited, it was more likely for that page to contribute higher to the participant scores. At first glance this might sound like circular reasoning since the number of visits was initially used to calculate the weighted variables but the number of times a page was visited does not necessarily increase or correlate with the quality of the essay scores. An interesting outcome was that last page count was highly correlated with weighted scores as well, suggesting that participants might have been more likely to end their search on pages that they have found relevant.

How did the page content influence participant experience? While click-track data provided information about how these pages were visited by participants, we can also generate meaningful variables about content of these pages by running measures for readability, page length, and sentiment. Readability shows how difficult an English passage is to understand as calculated by the Flesch Kinkaid score for the text content of a page. Page length is a calculation of total number of words in a page and sentiment is an analysis of the positive or negative associations of words used in the page.

The average readability score was 40 (out of 100), which is considered college level difficulty and is reasonable for pages created for graduate course use in an academic institution. Pages of the school website that portrayed conversations between school personnel were higher in readability (i.e., easier to read), since they were written as dialogues. Length of page was also not an indicator of high essay scores. Interestingly, length of page was not correlated highly with duration, which indicates that participants may not have necessarily read through very long passages but instead skimmed the content. Page sentiment average was 13 and varied very little across pages and overall was slightly more positive than fully neutral (a score of 0), which is expected from a generally factual website.

Overall, the readability, page length, and sentiment of a page did not seem to play a role in the scores. This is important because these attributes should ideally not play a role in helping participants understand school information. These three variables were not correlated with number of hits. The general outcome of these analyses of the page content shows that the website content was well suited for the purpose of the case. The attributes of the page content did not appear to detract participants' focus.

Which individual pages were most important? In previous sections we discussed how the ETIPS researchers' notions about relevancy may not be accurate considering actual online behavior on the ETIPS case websites. Now that we have some data available from actual use, we can use EDM methods to see if a construct such as relevancy can be better formed with data from the actual implementation. In other words, regardless of the ETIPS researchers' conception of relevant information, can we find out which pages had the most apparent relevance to participants? This may be done simply by ordering pages based on weighted scores, where pages with highest essay score contributions would be ranked higher. But we can include the influence of other variables to add depth to the results and improve the accuracy of understanding which pages had more apparent relevancy for the participants.

The variables that seemed to have a meaningful relationship with weighted scores (total hits, duration, readTime, lastPage count) were used in a Simple K means cluster analysis using WEKA. The resulting model had two clusters, one consisting of 54 of the pages with mainly lower scores on these variables and one with 18 consisting of mainly the high scoring pages.

Figure 6 shows a histogram of page visit counts where colors represent the two clusters. We can see that red (or lighter gray if printed) cluster dominates the higher page visits suggesting that the pages included in that cluster were among the most visited. Of the 18 pages that seem to be most relevant 6 of them among those designated as relevant in the original ETIPS research. This analysis suggests a dozen more pages that the ETIPS researchers may want to investigate further to see whether either their allocation of relevant pages might be expanded or a redesign of content may prove to be more useful.

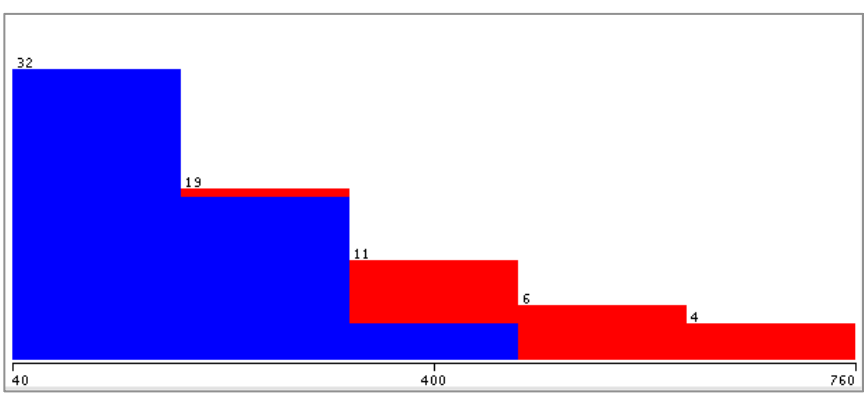


Figure 6. Distribution of clusters over number of hits.

Guiding search. In the investigations above we looked critically at the construct of relevancy and saw that page navigation might have been an influence and some other pages may have been more relevant. While this leads to rethinking the construct of relevant pages, the participant behavior may also be due to the lack of options for proper search. If the researchers in this ETIPS example have clear reasons to think that relevant pages are indeed important for learners to encounter, it might be that the lack of guidance for traveling across the pages might have in turn led to other pages receiving higher weighted scores in the click-track data. Online PD designers and researchers could add hints in their materials to guide participants towards the most relevant pages and to do so could require decision criteria for when it would be worthwhile to show these hints to participants.

To successfully apply hints, designers and researchers would need to have a sense of when participants were "lost", (i.e., straying away from relevant pages). In the case of ETIPS this could be provided in two possible instances, either when participants navigate to pages that are not relevant far more often than to relevant pages, or when participants' sessions begin to get very long. In both cases the software would need to know at what point the participant experience is getting worse. To help with this we can use calculations of average visits of previous participants.

Figure 7 shows the changes in the relevancy ratio for each consecutive page visit. Each line represents a user. On the x-axis are the number of hits, 1 being the first page they clicked, 2 the second page and so on. On the y-axis there is the ratio of relevant page visit, which is total relevant pages divided by total hits.

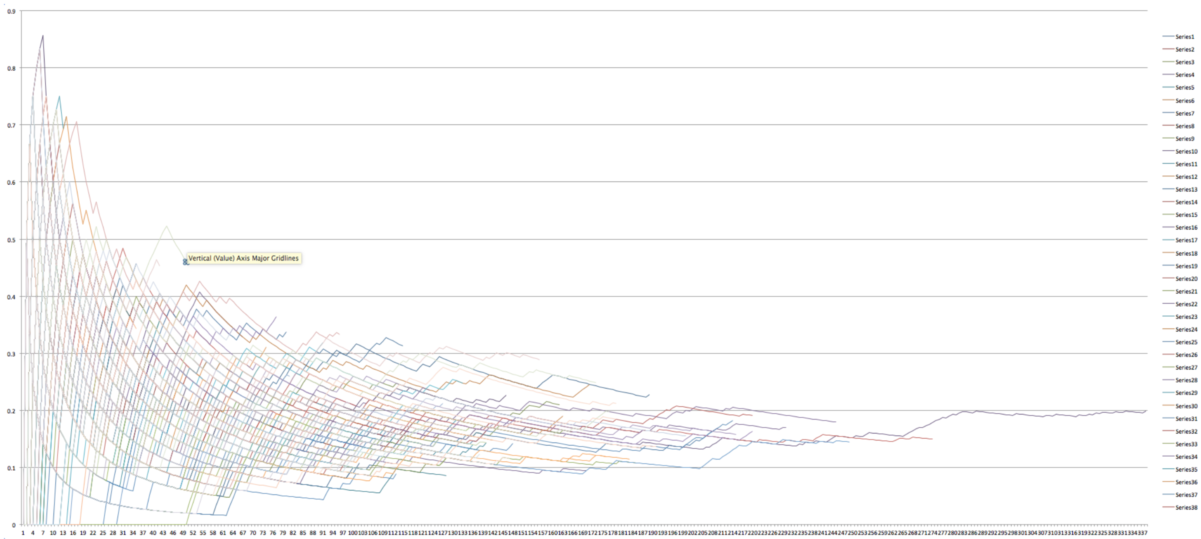


Figure 7. Changes in relevancy ratio on visit path

We see in Figure 7 shows that for most participants the early page hits, or the beginning of their case experience their relevancy ratios are relatively high. Relevancy ratios are the ratio of relevant pages visited to all pages visited. As they continue to browse the pages the relevancy is reduced and since there are many more pages that are not relevant. If relevant pages are clear to the participants and they continue to come back to the relevant pages the ratio would remain high but if they begin to move away from relevant pages the ratio will drop. In the ETIPS case we see that because users navigate to nearly all the pages over their visit the relevant pages they visit become a smaller percentage of their total visits. We also see in Figure 7 that some users go lower on the ratio than others and some users have a significantly longer experience on the website. This chart would allow researchers to find optimum locations where the ratio of relevant pages to total pages visited is too low or the overall number becomes so high as to be an inefficient use of participants' time.

To guide learners' search through online PD environments, researchers could provide hints in a couple of different ways. In the case of ETIPS, for instance, for most users after around 80 page visits the participants have consumed a large amount of the information available. Users going further might require assistance to find relevant information. Another implication of the chart is that in the early sections of their visits, up to around 30th click, most of these users are above 0.3 relevancy and others who have not found many relevant pages are beginning to see them. The ETIPS researchers could provide hints for users who have not reached 0.3 relevancy by their 30th click and guide them towards relevant pages. The ETIPS researchers could also look at the uniqueness of relevant and non-relevant pages to see if after 30 clicks the users have high relevancy but all come from a single page. If that is the case the ETIPS website could provide a hint to diversify their visit. These interventions are possible because click-track data can provide information about how the participant experience is changing from the first to the last click, and metrics for their experience such as relevancy ratio can be calculated at each step and as the participant is using the system.

RQ3: Understanding Participant Characteristics

Data mining can also aid in better understanding what participant characteristics are correlated with successful use of particular online learning materials. Knowing how materials work for whom could help instructional designers or researchers fine tune their online PD, or its implementation, or understand for whom to add hints at relevant points in the user experience to nudge them towards success. User characteristics include demographic information or other relevant variables that influence the way users experience the learning environment. For instance, are participants who self report expertise in technology getting more out of it? Are participants with more years of teaching experience more effective in navigating to the information? However, relationships found between online experience and participant characteristics need to be examined in terms of whether or not it is a problem if the software is not well designed to provide an equal experience for participants with varying levels of expertise and experience.

Twenty three user attributes were available from the original ETIPS research pre- or post-implementation surveys, including characteristics such as years taught, level of technology expertise, pre-implementation self-efficacy score, quality of implementation of the case by the faculty, and so on. Through regression analysis we examined if these variables were predictive of the final outcome of high quality responses in the essay questions. We did not find any statistically significant coefficient, meaning that if we knew all these variables for a participant we could not very accurately predict the quality of their essay score.

Data mining methods such as cluster analysis and classification provide exploratory research methods to see whether participants can be separated into distinct groups based on these variables. Using an expectation-maximization (EM) algorithm with all the variables for an unsupervised clustering resulted in 3 distinct groups (see color coded groups in Figure 8). However, there was no clear variable that distinguished the differences among these groups. We concluded that participants' experiences varied and no particular individual or group characteristics were advantageous or disadvantageous in relation to the online experience of the participants or their essay score.

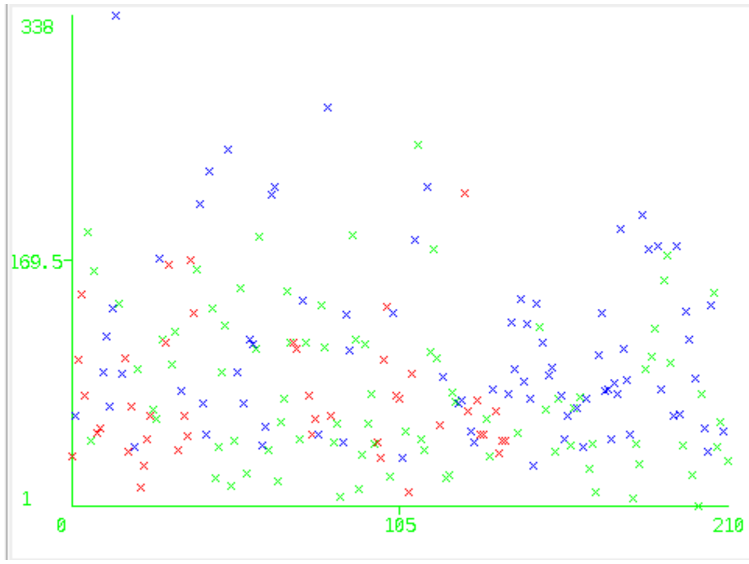


Figure 8. Cluster grouping with all variables showing page visits to demonstrate no visible separation between groups.

The EM clustering was done a second time, with just the analysis variables related to the participants' online experience, such as number of total page visits, total sessions, total durations, note lengths, average times. Again three groups emerged (see Figure 9), where one had lower values for all variables across the board and two high values for all variables and were differentiated by how many notes they took on pages. This indicated that being active in the online environment did correspond to better essay results, especially if those participants also took a lot of notes.

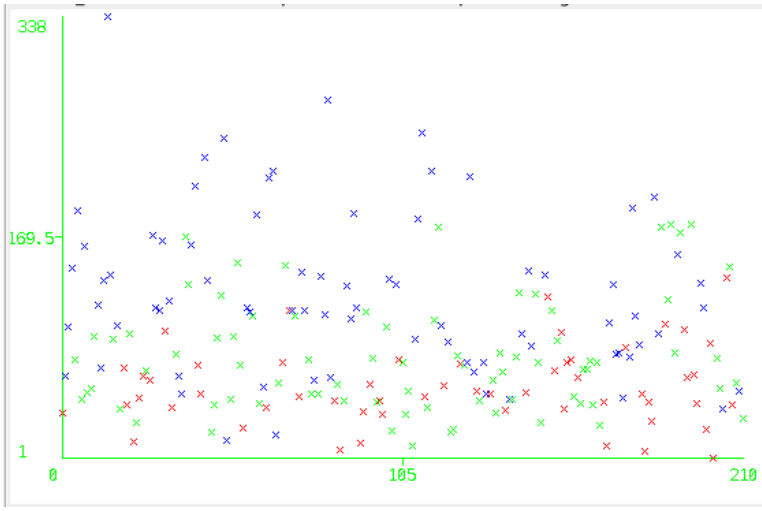


Figure 9. Cluster groups with page visit specific variables.

Discussion

The goal of this study was to demonstrate that using click-track data in conjunction with data mining and advanced statistical analysis methods could provide further insights into research on online PD environments. In this regard the analysis was successful in being able to shed further light into interactions by the participants within the simulated online environments. It provided opportunities to further investigate whether or not a construct of interest to the original ETIPS researchers had a significant effect on participants' quality of responses. While the correlation of page visits may not be the ultimate indicator of whether or not the construct of relevancy works, this demonstration case illustrates methods for investigating areas of interest to designers and researchers of online PD. Here we saw that individual pages were potentially more valuable since their visit had a greater impact on the outcome. In this case, the ETIPS researchers could use these indicators to investigate how they can improve their construct of relevancy. The main contribution of the click-track information in this case was providing to the ETIPS researchers data from actual user behavior by showing which pages participants actually visited.

The results suggested that the ETIPS simulated school navigation menu was a major determinant of the participants' visit patterns, and may have restricted discovery. On the other hand, page content, length, sentiment and complexity were not obstacles to the user experience. The results also added nuance beyond a correlation, that the variable most useful to distinguish user types was how much they made use of the resource. Participants who visited a lot, spent time on the pages and took notes were markedly better in responding to the case's questions.

Based on these results, the ETIPS researchers could revise the content of their relevant pages, reorder navigation, provide more discovery tools such as internal links, suggest hints to participants who seem to be missing valuable information, and find ways to encourage more engagement by participants who are not as active. These efforts would likely lead to more participants making better use of the case information and may then aid this online environment to better meet the goal of being an effective tool for PD.

Overall, the analysis demonstrates how and when an ongoing evaluation of participant search success in real time might add value to improving online experience. Participants' data showed where they were most likely to gain most of their information. Knowing these concrete points could allow researchers to place checks in user interaction, such as if a participant is not hitting relevant items after the average threshold, when and where they should receive just in time hints or navigation help. With click-track data and data mining we can pinpoint an exact ratio, or placement in the user experience where this hint is most valuable.

Conclusions and Implications

Educational data mining of click-track data can help us understand strengths and weaknesses of an online PD environment's design. Through exploratory data mining methods analyses can investigate the characteristics of pages in an online PD environment, such as page relevancy, participants' navigation, and relationships between page details and user characteristics. Perhaps the more significant contribution of a data mining analysis in such environments is the ability to address which aspects are important and valuable. Designing online PD environments are potentially very costly in time and financial resources and therefore any information that can provide more information on how better to structure the intervention can be immensely useful.

Educational data mining of click-track data collected and analyzed as a part of a pilot study can offer important insights into how the learning environment should be designed to best meet the learning interventions, through improvements in navigation, such as distributing content differently across pages, or providing hints to participants when they exhibit a particular pattern. Even without a pilot study, however, the analysis of the click-track data can allow a better sense of what conclusions can be drawn, and whether assumptions about the online learning environment and the users' experience with it are justified. It can also lead to good follow up questions to investigate in further studies.

Building online environments is a significant cost for designers and researchers and thus there is a strong incentive to know in what ways, if at all, the online environment itself supported the change proposed by the hypothesis of the research. Click-track data can show whether or not, and if so how, participants make use of the online environment and if there are differences between those who made more use of it and the others. We can investigate pages' design and any unwanted correlations with their sentiment, length or reading difficulty. Such analysis could show what might do to speed up discovery and improve the information layout and navigation.

While this demonstration case provides explicit direction for individual researchers building high quality online research tools, it shows the potential for improving online PD as well. With the ability to understand user path, impact of page quality, navigation, and depth of interaction, we can have an additional source of information by which we can evaluate the experience of individual teachers in their PD journey. With this information we can use additional tools to help, nudge and generally guide teachers in real time wherever they can access the online PD environments.

While these possibilities are encouraging, it is important to note that using better data could further enhance click-track data and data mining. For instance, the data gathered for this study was limited to page logs. However, current technologies allow for more detailed mouse clicks to be collected to better measure time spent on pages or elements of interest. Embedding meaningful meta-information in the clicks could help to build such constructs. For instance, tagging pages with categories related to the goal they are to serve could provide further avenues for analysis. Researchers could differentiate meta-information about pages such as whether they include tabular data or if they are opinion pieces from the case school's staff members. Also adding independent outcome variables that can help to identify the desired final behavior can greatly increase the data mining analysis options. In this example we used an essay score that was independent from click-track data to determine high quality responses. The more independent outcome variables available, the more options for classification, regression and similar analyses that are possible.

It is also important to remember that click-track data does have limitations in the information that it provides. Small data size is one limitation of click-track data analysis with data mining techniques. Some data mining analyses work best with large amounts of data to be able to build reliable statistical models. Overall we can overcome this by looking at cases, sessions or other ways of organizing and aggregating click-track data but higher number of participants or longer data collection times would help to improve the quality of the analysis. A second limitation is the validity of time-based calculations. It is a generally accepted problem in web analytics that time spent on pages is not always reliable. For instance it is quite hard to understand when participants end their meaningful interaction with the website while the page may still be open in their browsers. Calculating times for last pages viewed therefore remains a challenge. It is also difficult to know if the user did not pay attention in the time that the page was open in their computer. Unique identities in click-track data are also our only way of knowing the identity of the person using the system, so this type of data collection is very vulnerable to participants using another person's login information.

Perhaps the biggest limitation of click-track data analysis is the difficulty of validating variables for complex constructs. For instance, while we can know precisely if a person did make use of a single page or multiple pages, what should define a person as "engaged" with this material? The data itself can't always denote larger constructs without a framework to build a measurement model. In this scenario if we consider online behavior as evidence of a certain construct we can build a model where a number of metrics form the evidence for a competency or behavior that is not directly measured by these metrics. For instance, while engagement is not directly measurable with click-track data, a carefully built combination of metrics may be able to give us a better sense of it. Measurement frameworks such as Evidence Centered Design (ECD) provide methods of building models that can better operationalize difficult constructs and make them more testable. With its roots in testing and assessment systems, ECD is a promising method to enhance the data mining of click-track data demonstrated in this case study and provide researchers a framework to build valid measurements of constructs with click track data. Further research is needed to demonstrate how such relationships contribute to how click-track data can add value to teachers' learning in online PD environments.

References

Agudo-Peregrina, Á. (2013). we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online. *Computers in Human …*, 1–22. Retrieved from http://www.sciencedirect.com/science/article/pii/S074756321300188X

Allen, I. E., & Seaman, J. (2013). *Changing course: ten years of tracking online education in the United States*. Wellesley MA: Babson College/Quahog Research Group. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23494723

Anaya, A. R., & Boticario, J. G. (2011). Content-free collaborative learning modeling using data mining. *User Modeling and User-Adapted Interaction*, *21*(1-2), 181–216. http://doi.org/10.1007/s11257-010-9095-z

Avalos, B. (2011). Teacher professional development in Teaching and Teacher Education over ten years. *Teaching and Teacher Education*. http://doi.org/10.1016/j.tate.2010.08.007

Black, E. W., Dawson, K., & Priem, J. (2008). Data for free: Using LMS activity logs to measure community in online courses. *Internet and Higher Education*, *11*(2), 65–70. http://doi.org/10.1016/j.iheduc.2008.03.002

Bruckman, A. (2006). Analysis of log file data to understand behavior and learning in an online community. *The International Handbook of Virtual Learning …*, 1449–1465. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4020-3803-7_58

Chen, C.-M. (2008). Intelligent web-based learning system with personalized learning path guidance. *Computers & Education*, *51*(2), 787–814. http://doi.org/10.1016/j.compedu.2007.08.004

Cheung, W. S., & Hew, K. F. (2011). Design and evaluation of two blended learning approaches : Lessons learned. *Australasian Journal of Educational Technology*, *27*(8), 1319–1337.

Chitanana, L. (2012). A Constructivist Approach To The Design And Delivery Of An Online Professional Development Course: A Case Of The Iearn Online Course. *International Journal of Instrcution*, *5*(1), 23–48.

Cocea, M., & Weibelzahl, S. (2007). Cross-system validation of engagement prediction from log files. *Creating New Learning Experiences on a Global …*, 14–25. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-75195-3_2

Collinson, V., Kozina, E., Kate Lin, Y., Ling, L., Matheson, I., Newcombe, L., & Zogla, I. (2009). Professional development for teachers: a world of change. *European Journal of Teacher Education*, *32*(1), 3–19. http://doi.org/10.1080/02619760802553022

Dede, C., Ketelhut, D. J., Whitehouse, P., Breit, L., & McCloskey, E. M. (2009). A research agenda for online teacher professional development. *Journal of Teacher Education*, *60*(1), 8–19. http://doi.org/10.1177/0022487108327554

Desimone, L. M. (2009). Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures. *Educational Researcher*, *38*(3), 181–199. http://doi.org/10.3102/0013189X08331140

Fisch, S. M., Lesh, R., Motoki, E., Crespo, S., & Melfi, V. (2011). Children's Mathematical Reasoning in Online Games: Can Data Mining Reveal Strategic Thinking? *Child Development Perspectives*, *5*(2), 88–92. http://doi.org/10.1111/j.1750-8606.2011.00161.x

Gibson, S., & Brooks, C. (2013). Rethinking 21st Century Professional Development University of Alberta , 2 Educational Consultant. *Literacy Information and Computer Education Journal*, *4*(2), 973–982.

Hew, K. F., & Knapczyk, D. (2007). Analysis of ill-structured problem solving, mentoring functions, and perceptions of practicum teachers and mentors toward online mentoring in a field-based practicum. *Instructional Science*, *35*(1), 1–40. http://doi.org/10.1007/s11251-006-9000-7

Hung, J.-L., Hsu, Y., & Rice, K. (2012). Integrating Data Mining in Program Evaluation of K-12 Online Education. *Educational Technology & Society*, *15*, 27–41. Retrieved from http://www.ifets.info/journals/15_3/3.pdf

Hung, J.-L., & Zhang, K. (2008). Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching. *MERLOT Journal of Online Learning and Teaching*, *4*(4), 426–437.

Kennedy, G., Ioannou, I., Zhou, Y., Bailey, J., & O'Leary, S. (2013). Mining interactions in immersive learning environments for real-time student feedback. *Australasian Journal of Educational Technology*, *29*(2), 172–183.

Klassen, K., & Smith, W. (2004). Web log analysis: a study of instructor evaluations done online. *Journal of Information Technology Education*, *3*, 291–312.

Knight, P. (2002). A systemic approach to professional development: Learning as practice. *Teaching and Teacher Education*, *18*(3), 229–241. http://doi.org/10.1016/S0742-051X(01)00066-X

Köck, M., & Paramythis, A. (2011). Activity sequence modelling and dynamic clustering for personalized e-learning. *User Modeling and User-Adapted Interaction*, *21*(1-2), 51–97. http://doi.org/10.1007/s11257-010-9087-z

Kwakman, K. (2003). Factors affecting teachers' participation in professional learning activities. *Teaching and Teacher Education*, *19*(2), 149–170. http://doi.org/10.1016/S0742-

051X(02)00101-4

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. http://doi.org/10.3102/0034654307309921

Lee, Y. J. (2012). Developing an efficient computational method that estimates the ability of students in a Web-based learning environment. *Computers and Education*, *58*(1), 579–589. http://doi.org/10.1016/j.compedu.2011.09.008

Lin, C. F., Yeh, Y. C., Hung, Y. H., & Chang, R. I. (2013). Data mining for providing a personalized learning path in creativity: An application of decision trees. *Computers and Education*, *68*, 199–210. http://doi.org/10.1016/j.compedu.2013.05.009

Muijs, D., & Lindsay, G. (2008). Where are we at? An empirical study of levels and methods of evaluating continuing professional development. *British Educational Research Journal*, *34*(2), 195–211. http://doi.org/10.1080/01411920701532194

Parker, J., Maor, D., & Herrington, J. (2013). Authentic online learning: Aligning learner needs, pedagogy and technology. *Issues in Educational Research*, *23*(2 SPL), 227–241.

Reeves, T. D., & Pedulla, J. J. (2013). Bolstering the impact of online professional development for teacher. *The Journal of Educational Research & Policy Studies*.

Romero, C., & Ventura, S. (2011). Preface to the special issue on data mining for personalised educational systems. *User Modeling and User-Adapted Interaction*, *21*(1-2), 1–3. http://doi.org/10.1007/s11257-011-9096-6

Snow-Gerono, J. L. (2005). Professional development in a culture of inquiry: PDS teachers identify the benefits of professional learning communities. *Teaching and Teacher*

Education, *21*(3), 241–256. http://doi.org/10.1016/j.tate.2004.06.008

Storandt, B. C., Dossin, L. C., & Lacher, A. P. (2012). Toward an understanding of what works in professional development for online instructors: The case of PBS teacherline. *Journal of Asynchronous Learning Network*, *16*(2), 121–162.

Webster-Wright, A. (2009). Reframing Professional Development Through Understanding Authentic Professional Learning. *Review of Educational Research*, *79*(2), 702–739. http://doi.org/10.3102/0034654308330970

Wiebe, E. N., Branoff, T. J., & Shreve, M. A. (2011). Online resource utilization in a hybrid course in engineering graphics. *Advances in Engineering Education*, *2*(3), 1–22.

Zaiane, O. (2002). Building a Recommender Agent for e learning Systems Computers in Education. In *Proceedings of the International Conference on Computers in Education (ICCE'02)*.

Zorrilla, M. E., Menasalvas, E., Marín, D., Mora, E., & Segovia, J. (2005). Web usage mining project for improving Web-based learning sites. In *Lecture Notes in Computer Science* (Vol. 3643 LNCS, pp. 205–210). http://doi.org/10.1007/11556985_26

Using Evidence Centered Design and Educational Data Mining Methodologies With Click-Track Data To Explore Teacher Engagement In An Online PD Environment

Caner Uguz

University of Virginia

Abstract

As the access and availability of online technologies have grown, professional development efforts and research have also begun to make wide use of this medium. Using click-track data, these environments may provide further opportunities to measure constructs that are important to understanding PD effectiveness and make valid implementation choices. This paper demonstrates a combined use of an assessment framework and exploratory data analysis techniques to analyze click-track data towards an understanding of engagement in an online PD environment. Evidence centered design (ECD) is the assessment framework used to link the constructs of interest to individual tasks performed online and several educational data mining (EDM) analysis methods are used to create a final score of engagement. This demonstration case provides a qualitative evaluation of whether ECD and EDM together can provide a useful measure of engagement by drawing upon post-hoc data from an online PD environment.

Using ECD and EDM Methodologies With Click-Track Data To Explore Teacher Engagement In An Online PD Environment

Online education has evolved from a niche experiment conducted by a small group of virtual schools in the early 90s to a fledgling public and private sector with an estimated 2.5 million K-12 students taking online courses in 2015 (Dobrovolny, Edwards, Friend, & Harrington, 2015). The permeation of online environments has made this avenue an increasingly important topic of investigation for education researchers (Tsai, Shen, & Chiang, 2013). One of the exciting aspects of the online environments is the availability of new types of digital data for research that is embedded in the learning process context—meaning that it is collected as the activity itself is going on as opposed to after the event takes place (Cope & Kalantzis, 2015). Visitors to online environments leave traces of their user activity—ranging from the pages they visited to the individual mouse clicks, as well as all the content they create such as comments, likes, or blog posts—and with new programming methods, very detailed aspects of the learning experience can be recorded (Jansen, 2009), transforming the information available to us about all levels of the learning process (DiCerbo & Behrens, 2014). We can now have not only information on different age groups and types of content but also about learners' activity reading material online, participating in community activities such as a group discussion, solving puzzles, answering quiz questions, working with simulations, or watching videos. This all produces abundant data as millions of users interact with educational content at a time when online access is ubiquitous.

These attributes of the new digital data create opportunities for measurement that were not available before computers and the internet began to be used in learning environments. In classical measurement theories the testing environment was not adaptive because pen and paper measurements did not allow for insights from previous questions to be applied to the remaining items or the measurement environment in general. Starting with computer aided testing and culminating in the online learning and testing environments, a continuous and fluid testing environment has began to influence how we measure and evaluate (Mislevy, Behrens, Dicerbo, & Levy, 2012). The embedded measures have been able to keep track of user activity in real time where each measurement is linked to the information of the previous measurement from the same user (West, 2012). Online environments also have the advantage of being connected across users, which means individual user activity was stored and evaluated in an environment that also knows about every other user.

In light of these advancements, many researchers and educators have begun serious conversations about how the affordances of the online environments could be used for online professional development (PD) (DiCerbo & Behrens, 2014). This case study demonstrates how click-track data from an online PD environment for educators can be leveraged to better understand teachers' engagement in that environment, using methodological approaches that can be applied to any online PD environment.

Literature Review

As the number of online learning environments grew they soon were used for teachers' professional development (PD) because of how they can provide content not otherwise available to teachers in their schools, and at times that fit teachers' schedules, including real time support (Dede, Breit, Ketelhut, Mccloskey, & Whitehouse, 2005). But just as in traditional face-to-face PD environments, one of the main questions for researchers of online PD has been how to measure effectiveness of the online PD environment (Clark & Oyer, 2005). Many online PD researchers still measure teacher perceptions or attitudes after its delivery but increasingly the field is looking at building conceptual frameworks to identify and build successful PD by getting more direct measurements of success (Desimone, 2009), which could promote the affordances of online PD without losing any quality of training. Preliminary findings suggest that online PD does not have any disadvantages compared to face-to-face PD but teachers reported to be more satisfied with face-to-face training (Fisher, Schumaker, Culbertson, & Deshler, 2010). These indicators however may not be very reliable since quality of either the online or the face-to-face implementations would change teacher perceptions. When researchers include online data in their analysis of online PD environments they often rely on basic metrics such as frequency of logins and time spent on the website (Hrastinski, 2008). These metrics are useful but are not strong indicators of successful outcomes in online learning environments (Kupczynski, Gibson, Ice, Richardson, & Challoo, 2011). To give an analogy from face to face instruction, we know that students who show up for class have better chances at understanding the content of the instruction but we don't take attendance as the sole measure of whether the teaching was effective. Each classroom environment instead creates a set of expectations for learning outcomes, tasks that will accomplish these, and a rubric based measurement to evaluate whether the outcomes were achieved. Similarly, we need a structured framework in online PD environments to be able to measure effectiveness as well as other constructs of interest. Such a framework would require two important components. One is a method of building valid measures (similar to a rubric in a classroom) and the other is statistical methods to analyze and make sense of the data.

A good statistical analysis methodology that can utilize the digital data available in online environments is data mining, an umbrella term for both the processes and statistical analyses methods used to find meaningful patterns in data (Witten, Frank, & Hall, 2011). It can be applied to most types of data from quantitative tabular data to essay responses in any language. Data mining has been used towards very diverse goals from detecting mortgage loan risk (Gerritsen, 1999) to gaining insights in cancer research (Jiang & Liu, 2015).

As education data has grown, partly due to increased collection of digital data, Educational Data Mining (EDM) has also become a significant area of research in the last decade. EDM is essentially the techniques and methods of data mining as applied to educational settings with an interest in better understanding students (Baker & Yacef, 2009). The emergence of EDM at the time of the increasing digital data in education field is not a coincidence since collecting large amounts of data is costly and time consuming without using digital tools. The papers and content of the 2015 EDM conference show that a large amount of research conducted in this field rely on digital data (Boticario et al., 2015).

EDM is often used with the goal of looking at emerging patterns through a process of logging and pre-processing data for the purpose of analysis with diverse clustering methods (Valsamidis, Kontogiannis, Kazanidis, Theodosiou, & Karakos, 2012). Clustering is a commonly used strategy because it provides characteristics to emerge from the data without prior expectations or hypotheses (Vakali, Pokorný, &

Dalamagas, 2005). There are also other approaches to analyzing the data from online learning environments, such as learning analytics, but for both the goal is to provide methodologies for finding patterns by making use of the large data available in new learning environments (De Freitas et al., 2015).

EDM is not simply an open-ended way to look at data to find any pattern. Most studies that include EDM already have goals or questions in mind that guides their search and defines which methodologies to be used. However, EDM itself is still not equipped to make a validity claim between complex constructs and the observed measurements. As a statistical analysis methodology EDM falls short of providing a framework for measurement.

In education, psychometrics is traditionally used to build reliable and valid measures. As EDM has increased in use, the development of the Evidence Centered Design (ECD) has grown in parallel. ECD is framework for evidentiary reasoning in building assessment and evaluation tools, to support the connection between observable behavior and constructs by rigorous documentation and analysis of valid evidence (Mislevy, Steinberg & Almond, 2003). ECD is "a coherent, articulated system of procedures and materials that when interpreted and used properly allows appropriate inferences about a person's standing on some attribute of interest or makes a probability statement about some future behavior (Bond, 2014, p. 37)."

The main goal behind the framework is to guide building pieces of evidence into the construct that is being tested in order to add a stronger element of validity, so that the items used to evaluate the construct are indeed capturing key representations of it. ECD consists of five stages of development (summarized in Figure 1), which help to define constructs and build into a system the chain of logic and reasoning necessary to evaluate the construct of interest.

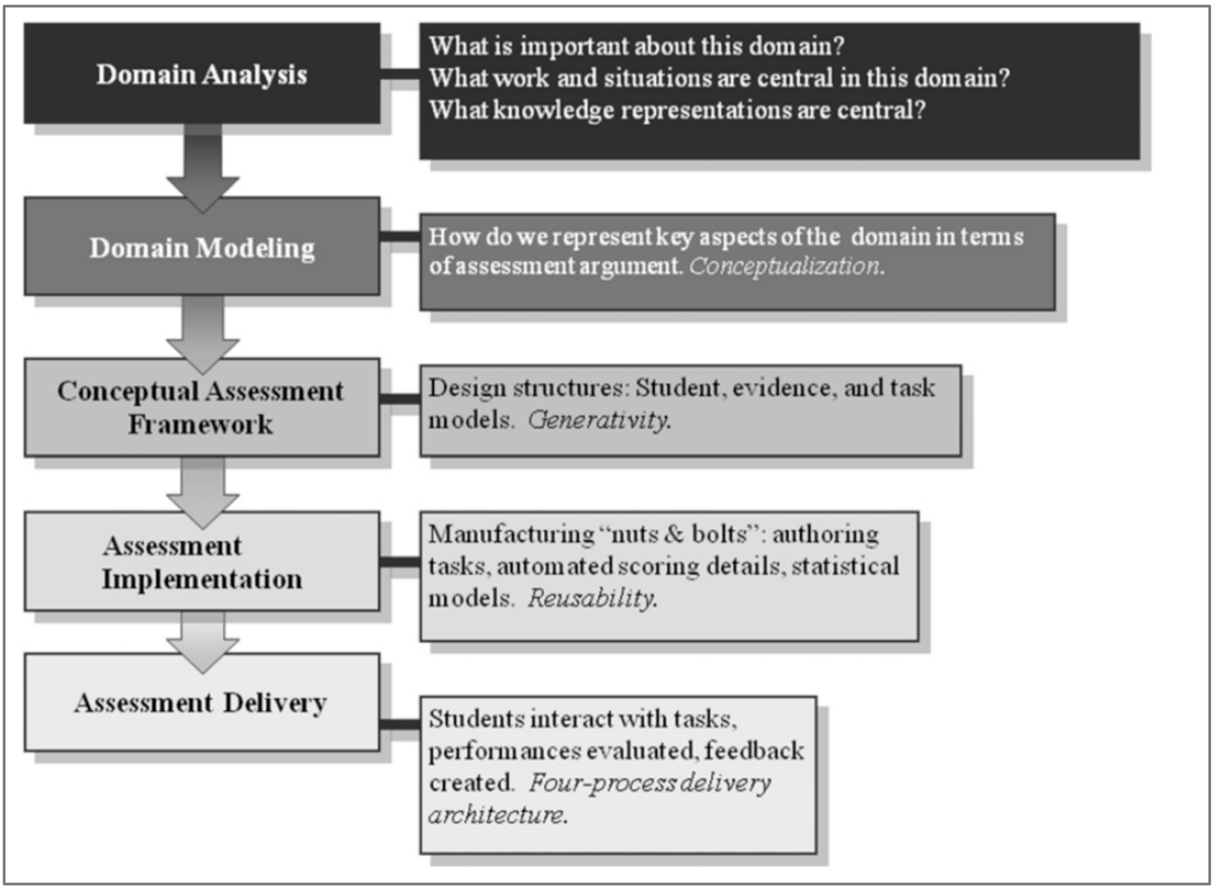


Figure 1. Evidence Centered Design Components from Mislevy et al. (p. 15, 2012)

The two initial stages of domain analysis and domain modeling are very high level conceptualization stages where the researchers and test designers need to think about what domain the measurement is going to address and how it can be defined in observable ways. For instance, when measuring whether an employee of a company should be given a promotion or not, the decision will need to define what aspects of this evaluation are important for the domain, which will include success in work settings and potential to continue to produce good work in the future. In domain modeling we would need to define this further by asking whether we can ask large-scale measurable questions that will get to these points. For instance it would be impossible to measure future work from a person, so a better question would need to be developed that focuses on measuring the potential to continue to create good work. At this stage the focus is not on individual items of measure but the larger questions that help us elicit characteristics that are not directly observable but central to the domain being measured.

In the third stage, the Conceptual Assessment Framework (CAF, shown in Figure 2) guides the ECD process where the actual measurement model is built. The CAF has three important components, the student model, evidence model and task model. The Student Model consists of the variables that represent the knowledge, skills and abilities we are trying to measure. If we are to give an example through an evaluation of tennis playing knowledge and competency we can state that the student model would include variables such as "being able to serve effectively, clear understanding of the game rules" etc.[1] The evidence model is the section where the student model variables would be defined in terms of concrete behaviors that can be observed. The evidence model can be as complex as the variable would require. It can for instance be anything from a single item to a hierarchical and weighted list of a large set of variables. In the tennis example, landing the ball in the correct section of the court during a serve and averaging a certain speed in serves could be two components that can be directly observed and reasonably related to "being able to serve effectively". The task model involves parameterizing the requirements for the environment and individual tasks where the evidence model can be observed. Parameters of the task model are important in making sure a valid

[1] The example of a tennis player is taken from Shute (2010), but some model details are changed for simplicity.

measurement can take place. For instance when measuring whether a tennis player has relatively fast serves, the player should get enough trials to be able to reach a reliable average. Also the conditions of the court, weather elements that would hinder proper testing (such as weather conditions) are important aspects of the task model. Since this is the area where student work products (answered tests, online activity, video recordings etc.) are collected, it's important to define parameters for the environment in which Evidence Model can be reliably measured.

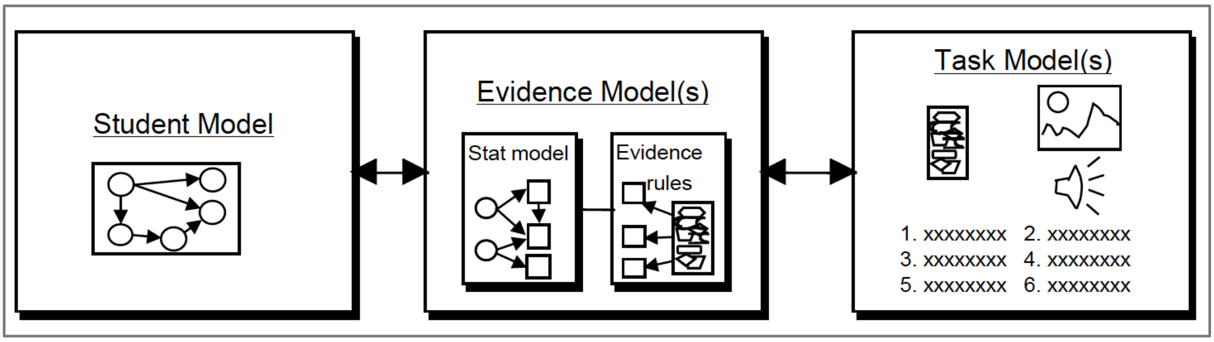


Figure 2. Conceptual Assessment Framework from Mislevy et al., (p. 20, 2012).

After the model for assessment is built with CAF the next stage is the assessment implementation layer, outlining the structure and form of the actual assessment. In online learning environments this would comprise the online content but more specifically the data collection infrastructure where activity logs and click-track data can be collected. This layer is informative for the rest of the model because if a type of information can't be collected due to technical difficulties or other reasons, the CAF would be revisited to align the evidence and task models with the capabilities of the assessment implementation layer.

The last stage of ECD is the assessment delivery layer, which is where the evaluation will actually take place. This can be a physical space, a computer environment, an online website, or any location where data collection is going to occur. In online environments this layer is the online environment itself and not the user's physical environment, over which researchers would not have any control.

These stages of ECD are designed to provide the developers of an assessment a unified vocabulary, and a structure with concrete steps that leads to a better construction of the link between observable behaviors and the evidence for the constructs in question. An important aspect of the model is that the CAF component is dynamic and adaptive, which means the model can guide design changes necessitated by external limitations or in response to evidence model based on data collected results of the process itself (Webb, Gibson, & Forkosh-Baruch, 2013).

Since its introduction two decades ago, the ECD framework has been used widely in diverse assessment areas such as test development in learning contexts (Zieky, 2014), large scale testing environments such as the Advanced Placement program (Huff & Plake, 2010), learning progression student assessment (Zalles, Daniel; Gaertel, Geneva; Mislevy, 2010), assessments embedded in educational games to measure constructs such as physics understanding (Almond, Kim, Velasquez, & Shute, 2014), training in network engineering tasks through a simulation (Benson, Fay, Kunze, Mislevy, & Behrens, 2012) and evaluating performance in epistemic games (Gobert, Pedro, Baker, Toto, & Montalvo, 2012). It is especially useful in online environments because they challenge how we think about individual tasks and require in-depth thinking about what constitutes evidence, particularly when used with the powerful methods of EDM. The continuous and fluid data collection in online environments also works well with this adaptive framework. Overall, ECD has its challenges, in that it’s a multi-step process with considerable preparatory work, but because it guides building evaluations closely tied to evidence it has the potential to provide a better measure for the constructs (Hendrickson, Ewing, & Kaliski, 2013), and when used with EDM can provide the conceptual structure and data analysis tools to respond to the need to analyze education data in the new age of big-data (Rupp, Nugent, & Nelson, 2012).

Traditionally in education, the concept of construct validity is used to guide developing an instrument that connects evidence collected from observable data to the construct of interest (Fraenkel, Wallen, & Hyun, 2011). Also, scoring rubrics are frequently used both in teaching and in research to improve consistency when measuring complex knowledge domains (Jonsson & Svingby, 2007). Practitioners and architects of both the EDM and ECD fields agree that these approaches used together have the potential in online PD environments to bring us closer to better assessments using click-track data.

Engagement is of great interest in the study of online PD effectiveness as active learning is an important component of successful professional development efforts (Birman, Desimone, Porter, Garet, & Yoon, 2000). Online environments often promise an individual experience that can increase user engagement by allowing users to make use of the resources at their own chosen time and pace with interactive components, and it may positively affect subsequent offline behavior. For example, Downer (2009) found that teacher use of online consultancy resources had a higher correlation with teacher responsiveness to intervention. Yet engagement is a very general term that is difficult to operationalize because specific aspects of engagement differ by environment. A measure of engagement in a tennis match will require different metrics compared to engagement in an online educational resource for teachers. The ECD framework's strength is building measures that align highly conceptual domain level definitions to individual tasks germane to the learning environments used for assessment.

This paper investigates potential methodologies of measuring engagement in an online PD environment by leveraging the affordances of click track data guided by data analysis and measurement frameworks. The growing field of EDM, and within it the increased use of ECD methodologies, suggests a potentially useful framework for researchers and designers to investigate online PD environments. We address the question "Can ECD and EDM methodologies used with click-track data provide a viable measure of teacher engagement in an online PD environment?" The goal is to provide a qualitative assessment of the use of this methodology and thereby contribute to the knowledge base of the merits of various methods to measure complex constructs in online PD contexts.

Background On the Learning Environment That Produced The Data

This case study of the use of ECD and EDM with click-track data utilizes data from the online PD environment CANLEAD, a research and development (i.e., Goal Two) grant funded by the USDE's Institute of Education Sciences.

Overview and Purpose

The overall goal of the CANLEAD project is to help increase middle schools' students' understanding of math and science concepts by providing their leadership team, comprised of an administrator, technology integration specialist, and math and science teacher leaders, with a curriculum and yearlong coaching for them to plan and provide professional development for the school's other math and science teachers to integrate technology into their teaching practices. The intervention provides the leadership teams with conceptual guidance, planning tools, and curricular materials and formative assessments for themselves, as well as for their use with their teachers.

Leadership teams attended a face-to-face three-day institute in June 2014 and then throughout the 2014-15 school year received monthly coaching through video conferences. As part of the institute the teams developed plans to onboard their math and science teachers and provide them with support to integrate technology. In addition, within the project-developed online PD environment they and their school's teachers had access to curriculum about integrating technology in math and science teaching. The curriculum materials for teachers included methods and examples of technology integration for topics in the middle school science and math curriculum and protocols for teacher discussion and sharing questions, lessons, and students' work samples. Monthly, leaders and teachers were asked to complete "Recap" surveys outlining their leadership for or integration of technology, respectively, and these formative assessment data were summarized by month in table and graph form and also available in the online PD environment.

Participants

In the 2014-2015 school year, after a three-year period of development and piloting, the CANLEAD researchers implemented the full intervention (i.e., summer institute, video coaching, support materials and formative assessments in the online PD environment) in 5 middle schools. Another 5 middle schools, which served as a comparison group, received only the support materials and formative assessments in the online PD environment. All 10 schools had applied to participate in response to an email invitation sent to all middle school principals of a mid-Atlantic state. The application process screened for access to technology, at least a half-time technology integration specialist for the school, and time built into the schedule on a weekly basis for teachers to meet and collaborate. After selection, the principal was asked to select a leadership team and they committed to attending the summer institute and participating in ongoing meetings and data collection. The five schools in the treatment group were matched with the five in the comparison group on characteristics such as size, geographic location (urban vs. suburban), economically disadvantaged students receiving free and reduced lunch, as well as ethnic and racial composition. Schools from the same district were kept together in the treatment or comparison group. In the following year, 2015-16, the comparison groups received the full intervention.

Click-track data from the classroom teachers from all ten schools are included in this case study with a total of 118 teachers.

Design of Learning Environment

The click-track data for this paper are associated with teachers' uses of the support materials in the online PD environment, which was designed by the project team and built by software developers to provide the teachers and leaders with a collaboration and networking space, a courseware-like learning environment, and a database of resources. Members of the project team prepared all of the website's content and assessments. The specific materials designed for teachers' uses were topic pages, split out by science and math and then by grade level, consisting of detailed guidance on using specific technologies for the curriculum topic, including descriptions of technologies that could add value to creating representations of the concept, rationale for and added value use in teaching specific objectives in the state standards for that topic, the estimated classroom time to implement the suggested lesson, and so on. The goal of each topic page was to ease teachers' preparation for technology integration while also making clear the added value of using this technology for the specified teaching objectives. Figure 3 shows one of the topic pages in mathematics, which includes a URL to a virtual manipulative, one of the types of technologies promoted to the math teachers.

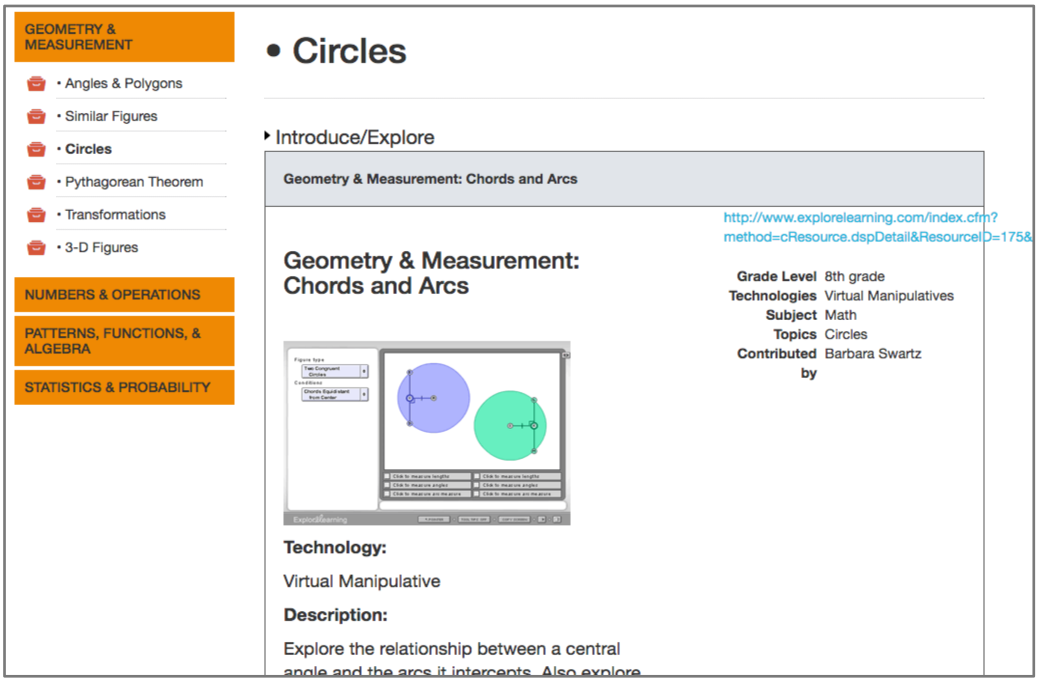


Figure 3. Topic page example in math, highlighting a particular website for using virtual manipulatives to introduce and explore the topic of "circles".

The online PD environment also provides each of the ten schools with a "network" (i.e., like a course in a learning management system) where the school's leadership team members can add tasks or discussion topics for teachers, set implementation times and prioritize them in calendars, and highlight particular school goals or specific curriculum resources to be implemented in their school.

Once teachers are logged in, they can chose to add a picture to their own profile pages, a personal statement and goals, select tags that describe their expertise and interests, and provide contact information such as Twitter handles, personal blog URLs or phone numbers. On the dashboard, where they land upon login, they see recommendations for items perhaps of interest to them (people and topics), based upon matches with any of their own tags they selected. Teachers also complete the formative assessments, a pre-post school-level survey and monthly recap surveys, in the online PD environment by following a direct link in an emailed invitation, or clicking a link to it from the tasks section of their dashboard.

The online PD environment created for this project allows for tracking clicks as the users interact with any of these aspects of the website and it is the participating teachers' click-track data that are used in this case study. The project team designed the parameters of the logging system to record more detailed information than would ordinarily be received about the clicks from a raw server log. These click-track data were collected for pre-defined events, mostly corresponding to page loads and clicks. The data also included meta information on the type of page being viewed and the type of action (e.g., update, view). The software also includes data points on how teachers made use of their profile pages. These include whether they uploaded a photo, added personal statements or added personal goals.

While the overall CANLEAD project had multiple research questions relating to the intervention, this study focuses just on the uses of the materials within the online PD environment and used the ECD model to operationalize teachers' engagement with the online PD materials provided for them. The essential question for CANLEAD in this segment was whether teachers made use of the website as it was intended by the researchers. This is a crucial problem in research project involving online components. The cost of building a custom PD environment is very high and there are many decisions that are made in its development. Having an evaluation of the extent to which teachers made good use of this environment is a valuable outcome both for the overall research on effectiveness of online PD environments but also on practical aspects of delivering the treatment that will provide most value to the teachers as they integrate technology in their teaching.

Methods

We present our methodology in two sections, with the first describing the use of ECD framework for the development of an evidence-based model for analysis, and the second describing the data analysis and EDM process used to provide the scores for the student model.

The ECD Model Applied To the Online PD Environment

The first two sections of the ECD model (see Figure 1) for the CANLEAD project illustrate the general assessment frames, while the detailed structure of the assessment was guided by the CAF (see Figure 2).

Domain Analysis. As first part of the high level conceptualization of what we were trying to measure, the domain analysis focuses on the important components of the domain of interest. In this case we are looking at teacher engagement with the online materials and the step includes separating the aspects that are important to the understanding of this construct versus those that are not, in an effort to begin to build an assessment in terms of tangible measurements. Part of the domain includes knowledge from the literature of what behaviors are important for PD, such as measuring effectiveness and active interaction with the PD content. Another domain area comes from the research of online environments in general, online PD environments in particular, but also in other fields such as marketing, which have pioneered investigations in the measurement of online behaviors.

As shown in the Table 1, the main aspects of the domain of engagement we identified for investigation were frequency, meaningfulness, continuity and variety of use. At this stage of, eliciting the aspects that are not important to investigate is also important to do, so that research resources are used appropriately. For instance, while it may be important in general to know the environment and context of teachers' access to the online PD environment, this information is not unique to the software in question and recording these offline attributes, which change across time and different schools, would require significantly more resources.

Table 1

Domain Analysis Points For The Engagement Measure

Important for this Domain	Not important for the domain
• Interaction with the website • Engage with topical content on technology integration with the website • Engage continuously with the website as opposed to once • Variety of use of more aspects of the website, as opposed to single purpose use	• The physical location where they interacted with the website • The computing environment that they used while interacting with the website • Differences in the prior knowledge of content area • Contextual aspects of ability of access, expertise in technology, scheduling, and so on

Domain Modeling. In this step we think more practically about the measurement methods that can elicit the aspects that are important for the domain without getting into the details of actual assessment model. We need to ask what behaviors teachers are doing that may be indications of the construct we are trying to measure and what the best environment for this would be. Table 2 shows some of the ways the domain analysis items can be arrived at.

Table 2

Domain Modeling For The Main Aspects of the Domain Analysis

Domain Analysis Items	**How teachers may model the behavior**	**The environment in which the behavior occurs**
Interaction with the website	By logging in to the website and remaining active	Through login and on the CANLEAD website
Make use of key resources	By studying topic pages or school pages that include	On school network and topic pages

	tasks related to topics	
Engage continuously	Visits that are meaningfully spread across the school year	On school network and topic pages
Variety of use	Visits that include diverse pages and actions	Anywhere on the CANLEAD website

In this case, the domain modeling decisions reflect that the data is to be collected in specific locations of the online PD environment, and that there is very little that offline information can help. Furthermore, these are events best observed as they happen because they may be too detailed to be able to recall afterwards through an offline method, such as a survey. Making this high level analysis of what measurement we need makes the match between click-track analysis and evaluation of engagement more clear in terms of the data sources best suited for relevant data.

Another clarification that the domain modeling brings is the level of detail in the click track data that is required. Because typical web server logs would not have enough detail for us to be able to discern certain aspects of the page views suggested by the domain analysis, the CANLEAD project created a web log structure that allowed for much more contextual information about each page viewed. For instance, we needed the click track data to differentiate between page types and action types. For a full list of the click track variables please see Appendix A.

Following the general conceptualization of how to measure the construct of interest, teacher engagement, we moved on to the three steps of the Conceptual Assessment Framework, the student, evidence and task models to further build the assessment components.

Student Model. Also known as the "Competency Model", this component supports defining the knowledge, skills and attributes we would like to infer when we make our measurements. In the case of measuring engagement we are looking for specific behaviors as indicators of "engagement as intended" on the CANLEAD website rather than a set of knowledge or skill items. The student model can also be seen as a more fine-grained version of the preceding domain analysis work . The overarching student model for engagement in this case study is shown in Table 3.

Table 3

The Student Model For Evaluating Engagement in CANLEAD

Utilize CANLEAD online PD resources throughout the year with the goal of integrating technology in the classroom through:

1. Engaging significantly by using the website at a frequency that will be sufficient to discover and make use of the material
    a. Spending adequate attention and interest on the website
    b. Taking time to read and properly answer surveys
2. Engaging meaningfully by studying individual topics that are provided
3. Engaging continuously by making use of resources across the school year to make timely use of the pertinent topics for teaching
4. Engaging widely by making wide use of the website as opposed to narrowly using it for a single topic or task by
    a. Making efforts to interact with other users
    b. Building a self profile by adding personal details
    c. Interacting with diverse resources and general sections of the website

The student model serves a basic but crucial purpose of identifying the main elements of the entire evaluation. The working hypothesis is that if the teachers show the behaviors in the student model (Table 3), we can be reasonably certain that they engaged with the website as the CANLEAD project intended it to be used. These ideas are based on the domain knowledge of the developers of the evaluation and should not be seen as the definitive indicators of engagement when we are building the measurement. While student model is quite robust for well-established fields where the aspects of knowledge in a domain are well established, in measurement development of a construct unique to the environment the model itself may need adjustments.

Building the student model focuses measure on knowledge and or skills that are vital to the construct. For instance, this model does not take into account the complexity of the website content being viewed. In other words, it is not an indicator of engagement if the user has visited more pages with highly complex language or structures because the content created in the CANLEAD website follows a uniform complexity and is organized to be familiar in structure and language to teachers. Since this observation is not part of the measurement there won't be any need to collect information on this variable and include it in the evidence and analysis.

A final note of clarification needs to be made about the student model provided above. The items in this model are defined generally as behaviors, which is not typical in ECD models. Usually the student model defines high-level knowledge or skills. Behaviors that help us infer these are defined in the evidence model. However a measurement for online engagement based upon click-track data is by definition built from a set of behavior patterns. The student model therefore defines the large-scale patterns themselves and the evidence model, described next, provides a more detailed and behavior focused elaboration of these patterns.

Evidence Model. One of the most important stages in using the ECD framework is the creation of the evidence model, which is the connection between the high level definitions of the skills and knowledge and the evidence used to make inferences about those constructs of interest. Our evidence model defines the observable behaviors that show whether the patterns defined in our student model (See Table 3) take place or not. Our evidence model consists of two parts. The first part is the evidence rules, which is the list of behaviors, shown in Table 4, we can count or score as these behaviors happen. The second part of the evidence model is its statistical model. This part converts the scores collected across all the components identified in the evidence rules into a single measure for the construct of interest identified in our student model. While in general the amount of statistical modeling required will depend on the task at hand, we use descriptive statistics and EDM methodologies as discussed later in the methods section.

The evidence rules for the CANLEAD data breaks down the patterns defined in the student model into observable and measureable behaviors. Table 4 provides the final variables for our measurement.

Table 4

Evidence Rules for the CANLEAD teacher engagement measure

1. Engage significantly A) Spend adequate attention and interest on the website	
Session number	How many different times do they login
Total pages	Total number of different views
Total Time spent	Total amount of minutes spent
B) Taking time to read and properly answer surveys	

Survey Clicks	Count of clicks that are for survey
Average time between pages	Time spent between view and thanks screen in seconds
Average time on surveys	Time spent on surveys in seconds
2. Engage continuously: Making use of resources continuously across the school year to make timely use of the pertinent topics for teaching	
Session range	Difference between last and first session times
Session gap average	Average time between two sessions
Session gap variance	Statistical variance of the gap
Weekly hours	Average number of hours weekly
3. Engage meaningfully: Using the website by studying individual topics that are provided	
Topic diversity	Number of different topics
Topic engagement	Average time spent on topic
Topic visits	Number of visits to topic
4. Engage widely: A) Making efforts to interact with other users	
Number of searches	How many times has searched anything
Checked profile	How many times have they checked other users profile
Added discussion	How many discussion clicks?
Check bookmark or tasks	How many archive hits
Active actions	How many actions categorizes as active, as opposed to reading?
B) Building a self profile by adding personal details	
Entered personal statement	Length of characters used

Entered goals	Length of characters used
Has photo	0 or 1 if they added photo to profile
C) Interacting with diverse resources and general sections of the website	
Resource use	Number of resource pages are hit
Assignment use	Number of assignment pages are hit
Prioritizing use	Number of actions prioritizing
Working with Google docs	Update titles events count

Task Model. After the evidence rules have been created, our next stage was to make sure that the measurement environment provided the participants opportunities to show these behaviors, so we might collect the evidence from which we wish make inferences about their engagement in the online PD environment. The task model provided the framework for the measurement environment and process, with the actual building of the environment occurring in a later stage. This step is useful in online environments since it helps researchers provide the requirements (or blueprints) for the software to be built so that that observation of the behaviors in the evidence rules is actually possible. In webpages there are many possibilities to consider as tasks, such as interaction behaviors like drag and drop, opening and closing of items, and so on. However, in many online PD environments text-based learning predominates and so the main method of interaction is reading through the available content, and therefore page views remain the main task with which we can measure observable user activity.

The CANLEAD project‘s task model has evolved with the development of the software during the development phase that preceded the data collected and used in this paper. Thus, our task model had to remain dynamic as well as revisions were made to the CANLEAD user interface. For instance, at one point teachers were given curriculum examples on the page as expanding and collapsing text boxes. Teachers would have to click to expand the boxes in order to view the content of the teaching examples, which from a measurement perspective served as a clear indication within that page that the teacher was interested in that specific content. However, because this created an unnecessary hindrance for teachers who wanted to skim through or view everything on the page, as the CANLEAD pages were updated this feature was changed to just keep all content expanded and immediately available on the page and the evidence rules were updated to reflect those changes to the interface. Thus, the important purpose of the task model was to make sure that the behaviors could be measured with the software that is used to present CANLEAD's online PD environment.

An advantage of the CANLEAD project was that the software was custom built for the purpose of the specific PD offerings of the CANLEAD curriculum and the technical development team was available for updates throughout the implementation to make sure that blackouts did not take place and bugs were fixed. It was therefore easy to ascertain that pages required for the measurement existed and worked well. For instance, we were sure to maintain the functionality that the teachers could easily change their profile information and access the topic pages. The leaderships teams for the schools participating in the study were also given demonstrations on how to use the website and they were instructed to conduct similar demonstrations in their schools as well. Making sure teachers knew how to use the website helped to alleviate concerns over whether engagement as we planned to measure it would be confounded by teachers' unfamiliarity with or confusion how to access and navigate the CANLEAD online PD environment.

Summary. As a first part of the methodology, the ECD framework provided the stages and guided the development of building a measure for the construct of engagement in an online PD environment specific to the needs of the CANLEAD project. This included narrowing the behaviors to observe from high-level concepts to individual actions that can be observable online, and building the online environment itself.

The Statistical Model and EDM

The second part of the methodology of using EDM and ECD in combination is centered on the statistical model mentioned above in the description of the evidence model. We describe the statistical model as the analysis methods we used to produce a score for teacher engagement in the online PD environment. The analysis itself generally depends on use of probability, logic, and other means that allow us to make reasonable inferences about the observable behaviors identified in the evidence rules.

Data cleaning and preprocessing. The type of data collected is an important aspect of the statistical analysis as it determines what assumptions can be made and which calculations can be run. In this case study we are using click-track data, which includes logs of page views and interactions. Before the preprocessing the click-track data, we conducted several checks to make sure that the data set was clean and the data collection was valid. Click-track data does not include missing values since the web servers generate the data, but servers indicate null values for categorical variables such as actionType if any specific action is not performed. A subset of the entire click-track data was extracted to include only teacher interactions from the 2014-2015 academic year to make sure only data relevant to this study were being used. Several methods were used to check the data was collected accurately. Spot checking variables and frequency distributions showed whether categorical variables were within the list of pre-determined categories. Timestamps were checked to make sure that times were uniformly recorded and durations were calculated to make sure usage times were reasonable (i.e., not 100 hours per week). Finally, user identification in the click-track data was checked to make sure that all users were from the 10 pre-defined schools.

Log data as it exists wasn't sufficient for our analysis because it was too granular. While each row of the click data is a single interaction a user makes, often a single interaction is not indicative of a behavior. For instance, viewing a topic may require going to the home page, finding the school page, and then choosing the topic page from there. Thus we had to take the click track data in some combination or in aggregate (Losarwar & Joshi, 2012) to produce the observable variables shown in our evidence rules shown in table 4.

The click track data set from the CANLEAD online PD environment included over 60,000 rows of data, each indicating a single activity by one of the CANLEAD teachers on the online PD environment. The authors programmed and used a custom script in JavaScript for data calculation. This tool is open source and available in the Github Repository (https://github.com/caneruguz/canlead-public)[2]. To create the required variables in the evidence model, this data was converted first into sessions, a common step for click track data sets (Munk, Kapusta, & Švec, 2010). Sessions can be seen as all the activities recorded online that constitute a single use instance by the user. For

[2] This repository does not include the dataset for privacy concerns. Please contact the author to request data files.

instance, a teacher using the website in the morning and then again in the late afternoon would have created two separate sessions. The custom script was used to create an intermediate dataset of sessions for teacher aggregate variables. The session variables were not used in the analysis but can be found in Appendix B for reference. The calculation for the sessions resulted in 1,841 sessions for all the participants.

Following the session calculations a new data set was created where each teacher using the CANLEAD website was a single row, each having values for 25 evidence model variables, one school designator (showing which school teachers were associated with) and a unique id for each teacher. The school and unique id columns were not used in the analysis but were useful to check for errors and investigate general patterns.

Data Analysis. One of the main reasons to use a framework for operationalizing a complex construct such as engagement is the difficulty of observing it directly. Our analysis therefore is using a latent-trait approach, meaning that we are using variables we believe may be good observable substitutes for the construct. Latent-trait approach is a good way to find out which traits are valuable in giving us a better approximation of the construct variable (Wang, Dziuban, Cook, & Moskal, 2009), in the literature this process may also be known as using proxy-variables (Kim, Park, Yoon, & Jo, 2016).

Another important aspect of our available data is that we don't have an independent external outcome measure that can give us the engagement level of a teacher without using the click track data. In other words, we don't have another way of knowing if the teacher engaged well in the CANLEAD online environment. Having this variable would allow us to significantly increase the types of analyses we could run, but we still have several methods in our disposal.

This case study uses several statistical methods used commonly for EDM analyses. The first is using a Simple K-Means cluster with 2-6 clusters to group the teachers based on the evidence model variables. The goal is to visually observe patterns within the variables and check whether any of the included variables individually cause a strong separation of the teachers available. The cluster analysis provides background information that will help us interpret the next stage of the data analysis. A second analysis with classification methods is used to determine whether teacher schools can be predicted accurately with the information about teachers' online behavior. The goal of this analysis is to check whether school implementation is a strong indicator of teacher online engagement that should be taken into account.

The third and last method is exploratory factor analysis, utilized here to understand which components (or variable groups) are more important than others in explaining the variance in engagement. This analysis will show whether some variables differentiate among users' engagement levels, which ones we can leave out, or how we can give weights derived from the analysis to individual variables in the evidence rules so that those that better explain the variance have a stronger representation in the final scores. The outcome of the analysis will be the new scores for teachers that represent their engagement level in the CANLEAD online PD software. As a final step in the analysis, we will take look at the structure of the final score.

The open source statistical software package R Studio was used (https://www.rstudio.com/) for the univariate analysis and factor analysis. The open source WEKA data mining tool (http://www.cs.waikato.ac.nz/ml/weka/) was used for the cluster analysis.

Results

Step 1: Checking Data With Univariate Analysis of Variables

The first step of this investigation was a univariate analysis of the 25 variables in the evidence model. Variables were standardized for further analysis, as variable units varied considerably. Histograms for some of the variables showed a number of teachers with very high numbers in some of the variables. For instance, while most teachers had a total number of hits across the entire year below 1000, some teachers had several thousand hits. After checking for calculation errors it was evident that the data did indeed include these uses and the outliers were not consistently the same person, indicating that they are less likely to be outliers. At this step, in order to not make judgments about teachers' use of the website in this exploratory study and to maintain the integrity of the original data set no outliers were removed from the analysis for the next step of analysis.

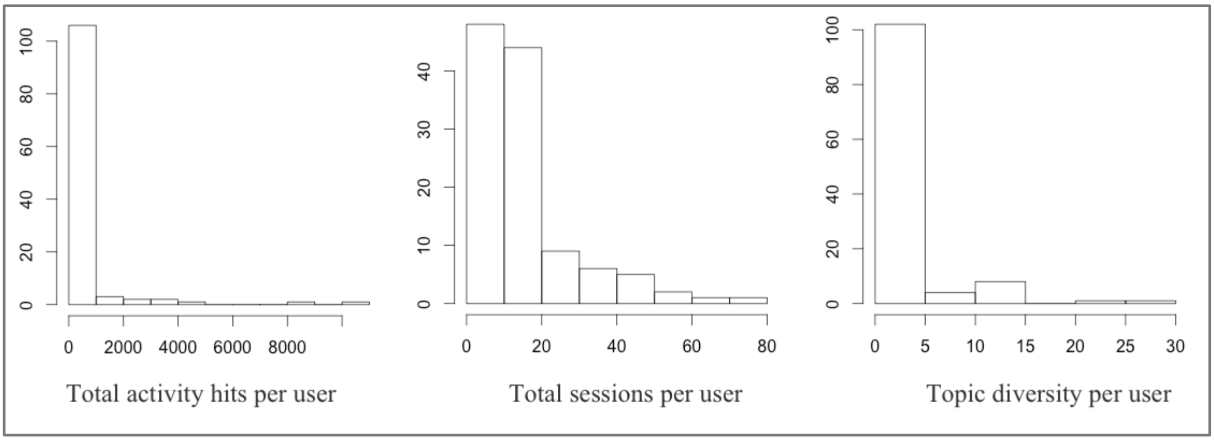


Figure 4. A sample of histograms for three of the evidence model variables.

Step 2: Checking for Patterns Emerging From The Data With Cluster Analysis

The second step of this investigation is an extention of the first step with data mining techniques to further conduct visual inspections about the relationships of the variables in the data. The goal is to see potential errors, noteworthy correlations or patterns emerging from the data. For instance our data mining model assumes that multiple variables play a role in explaining engagement. Checking for patterns before we run this analysis will help us see whether only a single variable is able to provide this insight or whether an exploratory factor analysis is useful for this data set. This step included a cluster analysis of the data with the 25 evidence model variables to check for patterns emerging from the data (see variables in Table 4 for the evidence model variables).

An initial 2-group Simple K Means cluster showed that a prolific group of users caused the very high numbers in many of the fields. This handful of users spent much longer on the website and participated in a larger number of activities than did most other users. They also visited the site much more frequently and made use of many different aspects of the website. However, they did not differentiate markedly in their profile build, survey behavior or general site activity. Figure 5 shows the two-cluster grouping. The large chart on the left represents the number of teachers on the x-axis and the number of sessions for each teacher on the y-axis. The charts on the right shows all the variables condensed to show the cluster groupings where the variable values are on the x-axis. Together these charts illustrate the point that the two clusters are uneven and the blue cluster shows a much higher value for activity related variables.

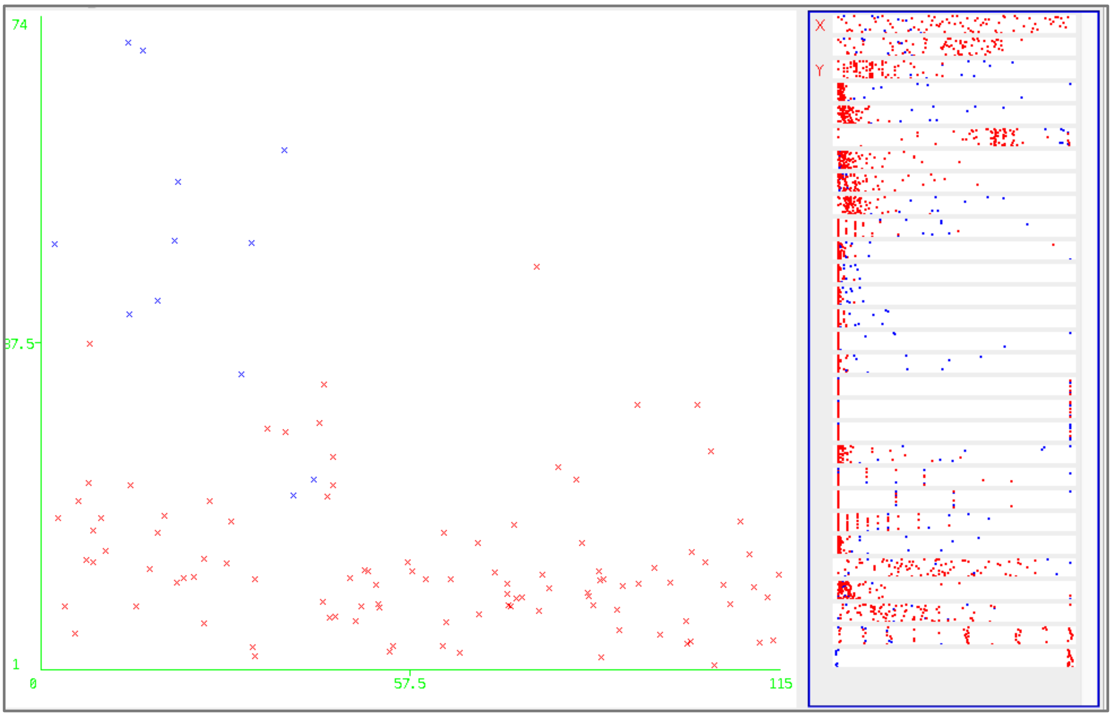


Figure 5. Two cluster grouping chart showing on the left number of sessions for each user and on the right all variables.

Increasing the number of target clusters and checking for group formation with cluster analysis did not yield insights about how the users might be showing certain patterns of behavior, with the possible exception of a group that has a large amount of activity consisting of responding to surveys. In a 4-cluster grouping, one of the clustered groups had higher total time spent on surveys and higher total clicks on survey related pages; suggesting they focused more on responding to surveys. However they also showed some activity in other areas of the website (see Figure 6).

The investigation with cluster analysis showed that there wasn't a single variable that caused a strong separation of users that would allocate them into their own groups. This can be interpreted as user behavior clusters using multiple variables to differentiate themselves and a simple analysis may not be possible. For this purpose a more detailed analysis should be used to explore the latent grouping of these variables.

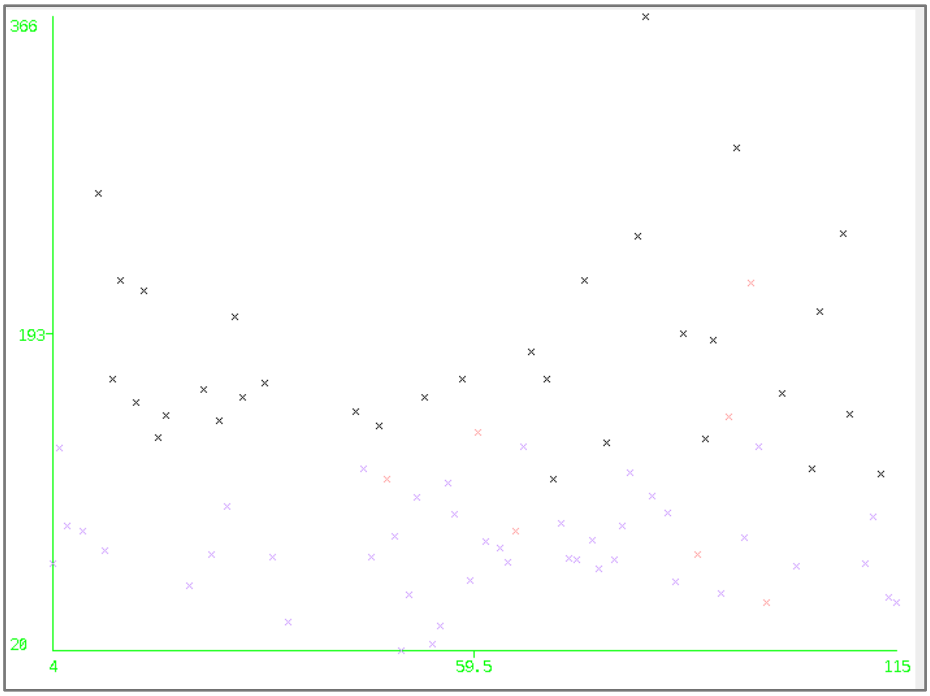


Figure 6. Number of clicks on surveys for each teacher where x-axis are individual teachers and y-axis is the number of visits for survey related pages, showing higher numbers for the black cluster.

Step 3: Checking The Potential Effect of Schools

The third step in our analysis is to see whether large-scale structural differences between teachers may have played a strong role in differentiating their behavior. The theory of action for the CANLEAD intervention recognized how leaders' implementation mattered for the teachers' experiences and that each school had separate leadership teams that implemented the CANLEAD curriculum in different ways. Conducting a classification analysis with the schools as the outcome variable can show us whether the variables available can be used to successfully predict the school location for each teacher. Using the J48 tree classifier with first a cross validation data training method showed that the school location of about 39% of the teachers could be successfully identified from the click track data. After applying the 66% split data training method as an added check, the data showed that 46% of the teachers could be placed into the correct schools. These findings suggest that schools may be an important component of teacher behavior but they are not prevalent enough in our data to an extent that teacher behavior differences can be predicted from school alone.

These exploratory analyses suggest while some patterns of user behavior might be emerging from the data, we can't rule out any variables as the main indicators for a large number of teachers. The distribution of the variables show that many teachers have used their individual approaches to utilizing the website and a scoring method is indeed a better way to make sense of the level of engagement across the variables we established from applying the CAF framework. If this step provided a strong relationship with schools we would need to go back to our CAF framework we built through the ECD methodology and revise our student or evidence models to investigate what aspects of the school environment we should factor in to our model.

Step 4: Uncovering Latent Factors

Our goal in this statistical model was to create scores for each teacher that measures their engagement. Scores will be weighted according to the importance of the latent factors. This step of our analysis aims to find the weights for the factors using an exploratory factor analysis. This is a useful method for finding out the latent components and their relationship to each other (Yong & Pearce, 2013). Similar methodologies are often used in survey development in psychology field to make sense of which items are related in factors and how important these factors are in explaining variance in the results. Our evidence model (see Table 4) similarly suggested 6 factors that may explain engagement

1. Significant engagement as measured by frequency of use
2. Continuous engagement as measured by range and gaps in sessions
3. Meaningful engagement as measured by type of activities
4. Site-wide interaction as measured by non-topic page interactions
5. Completion of profile information as measure by personal statements, goals and avatars
6. Survey interaction as measured by frequency of surveys and time spent completing them

The goal with the factor analysis was to answer whether these latent groupings exist in the data and if so which explain more of the variance between users.

We conducted several checks for assumptions for running factor analysis with the data. The click track data consists of metric values with a linear relationship, which allows for the factor analysis calculations to be made from the data. To satisfy the assumption of no outliers, we removed as outliers the 10 rows higher than 3 standard deviations across the different variables. This resulted in 106 cases with 25 variables and 4.2 cases per variable, which is considered poor in the Comrey and Lee (1992) guidelines of case to variable ratios, but it was the best we could obtain from this data set.

Measure of sampling adequacy test KMO resulted in all but two variables higher than 0.50, and a correlation matrix showed a large number of correlations above 0.3, satisfying the assumption of sampling adequacy for a factor analysis.

Before running the exploratory factor analysis we investigated how many factors should we propose as a target for the model. For this purpose a parallel analysis was run to show how many factors should be used. The analysis suggested using 6 factors (see Figure 7). We can see from this data that while all factors are providing some explanation of the variance there is a significant drop after the first factor, and the first 6 factors are providing the majority of the variance explanation.

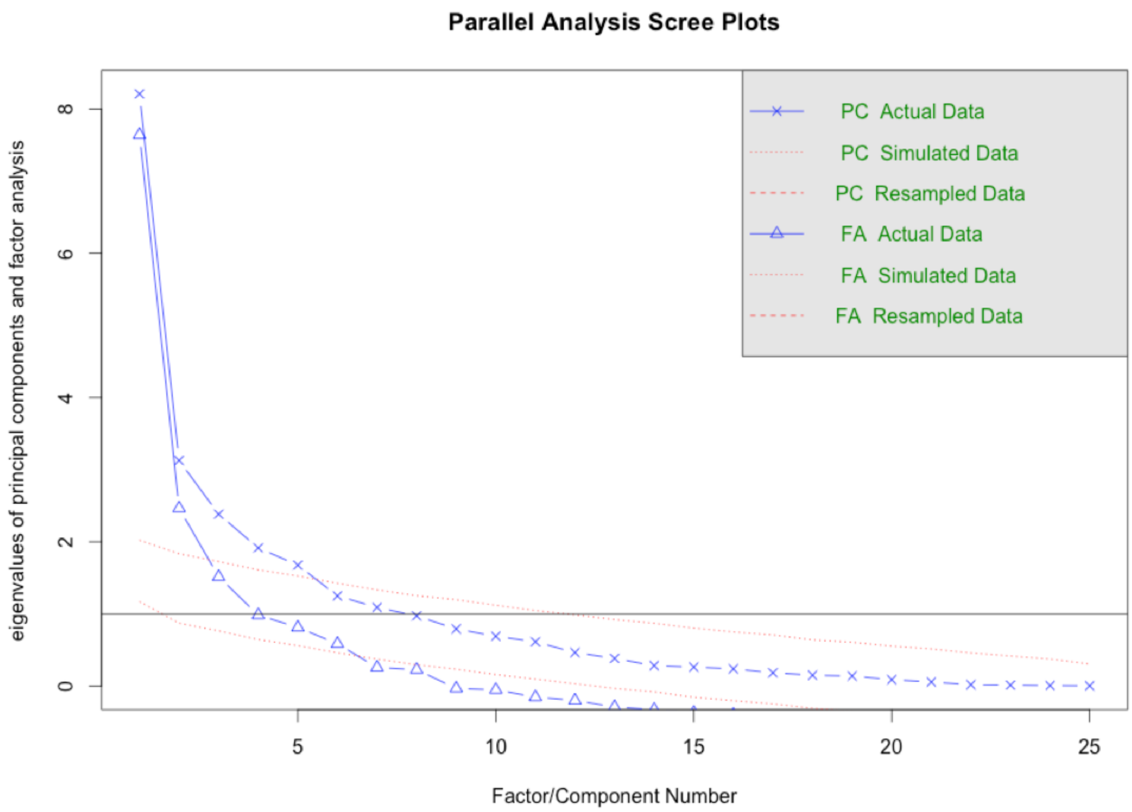


Figure 7. Parallel analysis showing factors explaining most of the variance

The exploratory factor analysis used an oblique rotation to provide loadings for the individual variables. The resulting table interpretation showed that some variables do group together as shown in Figure 8.

	MR4	MR3	MR1	MR2	MR5	MR6
totalSessions	0.40	0.40	-0.02	-0.18	0.39	0.00
totalHits	0.43	0.53	0.38	-0.04	0.09	0.00
totalTime	0.44	0.38	0.31	-0.04	0.17	0.06
sessionRange	0.17	0.34	0.04	0.12	0.26	0.01
sessionGapAverage	0.01	-0.01	-0.02	1.00	0.02	-0.02
sessionGapVariance	0.00	0.02	0.01	0.98	-0.01	-0.01
weeklyHours	0.23	0.29	0.29	-0.01	0.04	0.03
topicDiversity	0.69	-0.08	0.04	-0.05	0.16	0.04
topicAverageTime	-0.10	0.06	0.13	0.00	0.70	-0.04
topicVisitsCount	0.95	-0.10	0.10	0.01	-0.02	0.02
topicResourceUse	0.97	0.01	-0.04	0.02	-0.05	-0.03
topicAssignmentUse	0.77	-0.09	0.07	-0.01	-0.07	0.18
topicPrioritizeUse	0.00	-0.05	0.19	0.03	0.64	0.04
topicUpdateTitlesUse	-0.05	-0.06	1.00	-0.01	0.04	-0.01
personalStatement	-0.07	-0.02	0.01	-0.02	0.00	0.85
personalGoals	0.05	0.02	-0.02	-0.03	-0.06	0.76
personalAvatar	0.15	0.10	-0.01	0.31	0.40	0.27
searchCount	0.76	0.14	-0.11	-0.03	0.03	0.03
searchOtherUsers	0.21	-0.06	-0.14	-0.04	0.10	0.47
searchDiscussionItems	0.43	-0.05	-0.14	-0.06	0.05	-0.25
searchArchiveCount	0.40	0.11	-0.04	-0.10	0.17	-0.10
searchActiveCount	0.03	-0.02	0.97	-0.01	0.02	0.00
surveyClicks	-0.01	0.96	-0.05	-0.01	-0.18	0.00
surveyUntilSubmit	-0.02	-0.33	0.07	-0.08	0.68	-0.05
surveyTotalTime	-0.15	0.85	-0.10	-0.06	0.12	0.02

Figure 8. Evidence model variables with the six factors.

Step 5: Interpreting Latent Factors

Factor loading values themselves and the groupings need to be interpreted with our understanding of the variables and the engagement construct. The main outcome from the distribution in Figure 8 is that there is a very strong component that combines all variables that have some type of numeric counts and aggregates of time and page views. The original group denoting engagement as measured by frequency included only the first three variables but MR4 has many others in the same component. This makes sense since a lot of the engagement does require interaction and our method of measuring engagement in different areas was to count number of interactions in these areas. This was a necessity since we are lacking other types of metrics such as creation of content or mouse interactions that can move away from counts of page views. We can also see that the first 3 variables have high loadings in more than one component because they represent overall use of the site. They are not however as prevalent in measures of completion of profile page or gaps between sessions since these metrics are not related to number of page views.

The model provides evidence for three more of the original groupings as components: continuous engagement, completion of profile, and survey completion behavior. Factor MR6 showed that teachers with personal statements and goals were differentiated from other teachers. These two items are grouped together into a single factor which supports the assumption that filling out personal statement and personal goal are indicators of adding personal information as a sign of further engagement in the ecosystem of the CANLEAD website. Considering that relatively few teachers have these filled out it's good to see that the model confirms the importance of these items as differentiators.

MR2 factor grouping showed that session gaps were a separate latent factor to consider. This is most likely because it provides insights into user behavior in a very different method compared to other indicators by looking at the space in between visits. MR3 factor similarly showed that survey completion was a distinct indicator since it grouped together survey clicks and total time spent on surveys. However "surveyUntilSubmit", which was a variable showing the time in between survey beginning and final submit was not included in that factor. We can interpret this result to mean that the time difference in filling surveys is not a useful indicator when taken together with other survey specific behaviors. On of the variables, the weekly hours measure, did not show high loadings in any of the factors and indicated that as a measure it most likely would not provide additional insights into user engagement.

Two of the latent factors, MR1 and MR5 included variables that did not provide a plausible interpretation with the current data set. MR1 factor is prominently grouping the action of updating titles in Google docs related documents and the number of actions that are considered active. MR5 groups together average time spent on topics, topic prioritizing actions, setting personal avatar and time spent between starting and finishing surveys. It might be possible that these actions make sense together in a way that is not clear from reading the log data but could be further clarified with interviews or qualitative analysis of the visit patterns. It is also possible that these factors, while providing explanation for some variance in the data, are not indicative of latent factors.

This analysis is a very good example of the role the evidence model plays in the ECD framework. Our statistical model is helping us understand how the evidence model items performed to provide a measure for our final construct. We see some of the variables generated having a strong contribution to our understanding of teacher engagement, but some less so. Here we retain all the existing variables and will differentiate their contribution with weights, but if the factors themselves were important to us we would go back to our evidence model and revise the categories and reduce some variables. This integration with ECD would help us create a more fine-tuned assessment model for the next stages of the research or evaluation.

Step 6: Calculating Final Engagement Scores

The last step in the statistical model is to assign scores to teachers with the insights from the factor analysis to finally build the score for the engagement construct developed with the previous analysis steps. One of the affordances of the factor analysis method is that it provides a way to calculate original scores known as "factor scores" with the weights of the components so that each teacher can have a score in the respective component (Distefano, Zhu, & Mîndrilă, 2009). This allows us to see which teachers were stronger in which of the components and also produce an overall score from all the components involved.

For this paper it is useful to briefly look at the final standardized scores produced for the teachers. Figure 9 shows histograms of individual factors showing the distributions of teacher scores for each factor. The scores from the factor analysis are only guidelines that work better at describing variance. CANLEAD Researchers can use them as they are or adjust with weights and other means to give further prevalence to variables that are deemed important.

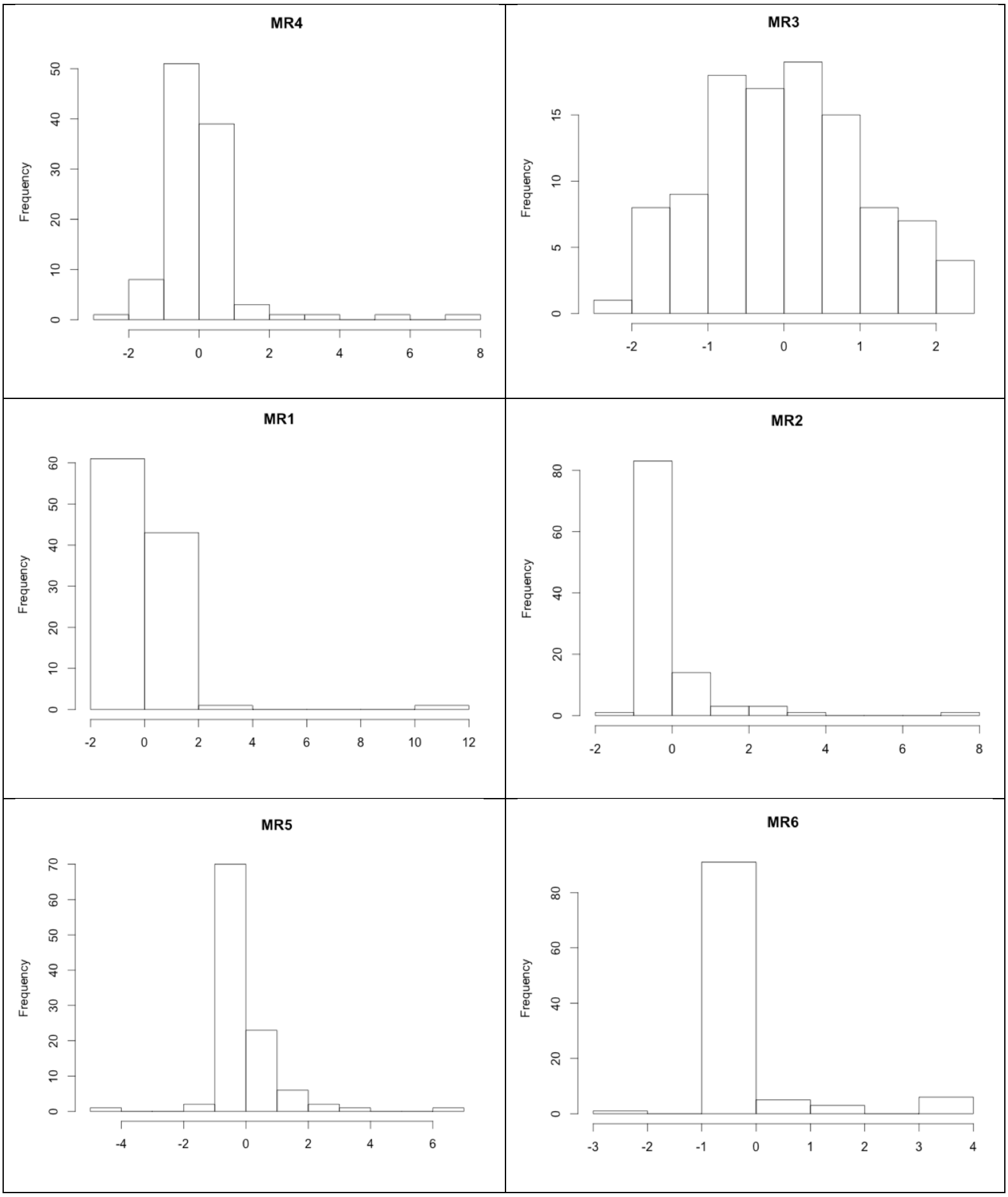


Figure 9. Histograms of the teacher scores on 6 factors of the exploratory factor analysis.

For each chart the x-axis is the standardizes scores and y-axis is the frequencies.

Discussion

The methodology and results illustrate how the ECD framework can be used to operationalize variables and then apply a statistical model based on EDM to build relationships between the data and the measurement. The aim of this study was to add value to online PD research that used click-track data by using a methodology that converted a very high level concept of engagement in an online environment into meaningful and representative data points using EDM methods with an ECD framework. Next we discuss how this methodological approach contributed to gaining a sense of engagement, the benefits and limitations of EDM in this approach, and what aspects could be improved or further investigated for better results in future studies.

Added Value From Using Evidence Centered Design

The nature of the research question required defining a term in the context of the type of data available in this research, since this construct wasn't otherwise already defined with valid available measurement tools—such as with self-efficacy, for example. In this case engagement had to be defined and operationalized in terms of the online PD environment parameters. The task of measuring a construct is a challenging one, and the construct of validity and the statistical processes typically used for devising measurement instruments suggest a part of the process, but as a conceptual framework, ECD provides very concrete guidance, in terms of both steps and focus, to build the measurement system. For the needs of this research ECD provided a logical step-by-step, methodology to build the measurement model that was needed.

One of the difficulties of using ECD with a construct that is based on behavior patterns was the potential confusion in the vocabulary ECD uses. Domain analysis and modeling suggest a knowledge domain and behavior is prescribed specifically to the task and evidence models of the framework. Here, while the construct itself was defined as a desirable pattern of behaviors, not knowledge of the behavior nor a specific skill, ECD was still useful for prompting how to define the behavior patterns. If ECD is to be used more broadly with click track data however, moving away from its roots as a test assessment framework measuring knowledge, further elaboration on how to adapt its use in behavior driven constructs may be useful.

Added Value From Using Educational Data Mining

The availability of the statistical modeling as a process in ECD worked well with integrating some EDM methodologies in this study, showing both exploratory and traditional uses of data mining in educational settings. Cluster analysis of the data was a useful tool to observe groupings among teachers to ask which variables are possibly individual causes for differentiation. This tool showed that no individual variable was a statistically significant differentiator but suggested that frequency related variables may play a major role, a hunch that was confirmed with the final analysis.

EDM also allowed for checking the strength of connections unrelated to the measures that were being tested by using prediction methods such as classification. For instance, in the construct for engagement we did not think the school the teachers were from should be a variable in the evidence model. However, had this variable been a major predictor of teacher behavior it would have signaled we needed to rework our measure of the engagement construct to somehow account for it. EDM allows a study to quickly test such variables and build more robust behavior defined definitions of the construct under question.

Factor analysis, the final analysis method used, is a very common procedure used in educational research to develop instruments and questionnaires but also considered to be especially useful in data mining analysis methods for finding latent factors and reducing dimensions. Factor analysis helped this study confirm its original evidence model and find out possible components that were missed in its development.

What did we learn about teacher engagement?

After applying the research methodologies of ECD and EDM we were able to get a more nuanced picture of teacher engagement. One of the outcomes for instance was that some teachers were very active compared to others and showed significant interaction with different aspects of the website. In such cases, because cluster analysis identifies individual teachers, researchers can talk to these participants to uncover what motivated them or helped them be prolific users of the website. We also found out that engagement styles and magnitude differed from school to school but not in a decisive way; which is an encouraging indication that the website use was not dependent on strict implementation differences within schools.

We have also seen that teachers have used the website continuously across the school year in majority of cases and have checked in frequently. While some of this interaction is due to the periodic filling of the surveys, the interaction for other pages has also been significant across time. This may have been due to the offline elements of the CANLEAD research process and the close guidance and assistance of the researchers that may have kept up the interest and planning to continue to make use of the resources.

Another noteworthy finding is that teachers mostly stayed within a restricted number of topics, meaning that they mainly made use of resources within their school network, which was intended; but did not go beyond requirements to fully utilize the website resources to create their own content or new networks of interaction with other teachers. There is also overall very little creation of content or active involvement through new posts, edits, comments. The website was intended as a medium of consumption primarily with lesson plans and resources for planning. The analysis does show that teachers were very cognizant of this role and utilized the website most for that purpose.

Overall teachers have not spent a very large amount of time on the website, visiting frequently for short periods of time, making use of provided resources and completing surveys as requested. The times of use are not uniform for schools or teachers, showing some concentration at certain times, although we can’t know from the data what these specific periods mean without interviews or other qualitative data.

It can be argued that teachers used the website in the intended way as a resource among the full CANLEAD curriculum and support structures but their use did not utilize the potential for informal learning or sharing of experience that the technical capabilities of the website allowed for.

Limitations Of Using Click track data

While the availability of a large amount of teacher activity information in the form of click track data made the use of EDM possible and allowed for a more robust construction of the traits showing engagement, it also highlighted an important limitation of click track data. Potentially the biggest concern with click track and log data is its reliance on counts of views. Many of the variables in this paper were derived from counts in different areas of the website. The availability of page categories and action types were immensely useful in differentiating the type of use but because they all came down to counting visits they correlated significantly with overall visit statistics. Visits cloud the measurements significantly because they are an overarching requirement for any other interaction. People who visit more have more opportunities to display more of the diverse types of behavior available. This limitation led to the variables in this study relying heavily on a single factor and our analysis merging different aspects of use, which in turn implies that visits were the only important aspect of online engagement. This would be misleading since visits are required for any type of engagement to begin with.

To overcome this methodological problem click track data should rely less on visit counts. For instance, when measuring whether a page was read, if there is an indication of how far the user scrolled down the page that value might be the better basis for inferring a user read a page, instead of just counting visits to a page or time spent on the page. Similarly, opening and closing content within pages can be better indicators of interacting with the content. As user tracking technologies and user interaction design evolves, we can begin to rely less on page loads and more on direct actions to gage user behavior. This change will lead to better differentiation of behavior patterns.

Conclusions

One of the recommendations of Dede et al.'s (2009) call for an online PD research agenda was to make use of the data collection methods available within online environments. A critical question for researchers wanting to follow this advice is where to find demonstrative examples or methodology. This paper provides a potential methodology available to build measures for constructs in an online PD environment comprised of two components that each offer advantages for online PD research. ECD is a well-structured framework that is tested in many assessment environments and well suited to the task of building a measurement for an online PD environment. It's ability to couple with not only EDM but any kind of statistical model makes it an especially ideal framework for quantitative data sets, such as click-track data. As applied researchers in different fields adopt ECD for their use, the vocabulary and descriptions of the framework are evolving (Nelson, Nugent, & Rupp, 2012). Data mining is also of growing interest to the field of online PD and can be applied to any type of data including surveys, demographic information or online content generated by participants (Rice & Hung, 2015). The combination of these two methods has the potential to provide a reliable methodology for analyzing click track data and this paper shows some of the potential to move beyond simple metrics in analyzing online PD data.

However, further research to improve our understanding of the best methodological approaches should be cognizant of the limitations of analyzing click-track related data. Intentions are not clear from user browsing behaviors and the design of pages and website navigation may play a significant role in changing the meaning of a page view. Without additional data sources we also can't be directly measuring whether a person is truly connected with and making use of the content. Our measures of time spent and visits are indirect indicators that allow inferences at best.

It is also important to remember that behavior in online PD environments can be approached from many different theoretical and practical avenues. For example, a qualitative think-aloud method to measure teacher engagement in online PD environments might provide even more detailed and direct measures than those illustrated in this paper. Focus groups, interviews and surveys also have potential for understanding how teachers are interacting online. Understanding the full picture of what works best in online PD environments will undoubtedly require a combination of approaches and methodologies, as well as different types of information.

Further research will be required to better understand the context in which these methodologies work best and what other important caveats to test. Some points emerging from this study show the need to try these methods with click track data that goes beyond page views. Online PD environments should try to include digital interactions (e.g., dragging, reorganizing, collapsing, pressing a key for next item) that may further demonstrate engagement. Other studies might compare, for the same sample of group in an online PD environment, possible measurements from different types of methodologies to understand their relative advantages in terms of insights, cost-effectiveness and effort required. This would be helpful to research teams without the resources to build custom websites or click data tracking tools. Finally, more fine grained studies trying to answer validity of assumptions we make in click track data would be highly useful in the field of education. For instance being able to answer questions such as: "how much has a person read from the page" or "at which point in the reading was the user confused" would be useful. Some exciting new studies using eye tracking have shown the potential of new technologies in answering some of these questions in more detail (Wang, Yang, Liu, Cao, & Ma, 2014).

This study aimed to demonstrate how an existing framework and analytical tools could be used with click-track data to help us gain insight into how teachers use online

PD environments. However, taking on this formidable task of constructing a truly useful, effective and engaging online PD environment will need to also draw on insights from the field, the needs of the target audience, and the ingenuity of researchers and designers.

References

Almond, R. G., Kim, Y. J., Velasquez, G., & Shute, V. J. (2014). How Task Features Impact Evidence From Assessments Embedded in Simulations and Games. *Measurement: Interdisciplinary Research and Perspectives*, *12*(1-2), 1–33. http://doi.org/10.1080/15366367.2014.910060

Baker, R., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*. Retrieved from http://www.educationaldatamining.org/JEDM/images/articles/vol1/issue1/JEDMVol1Issue1_BakerYacef.pdf

Benson, M., Fay, D., Kunze, K. L., Mislevy, R. J., & Behrens, J. (2012). Putting ECD into Practice: The Interplay of Theory and Data in Evidence Models within a Digital Learning Environment. *Journal of Educational Data Ming*, *4*(1), 49–110.

Birman, B. F., Desimone, L., Porter, A. C., Garet, M. S., & Yoon, K. (2000). Designing professional development that works. *Educational Leadership*, *57*(8), 28–33. Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED455227\nhttp://cemse.uchicago.edu/computerscience/OS4CS/landscapestudy/resources/Birman-Desimone-Porter-and-Garet-2000.pdf\nhttp://www.scopus.com/inward/record.url?eid=2-s2.0-0034382095&partnerID=4

Bond, L. (2014). A Brief Note on Evidence-Centered Design as a Mechanism for Assessment Development and Evaluation. *Measurement: Interdisciplinary Research and Perspectives*, *12*(1-2), 37–38. http://doi.org/10.1080/15366367.2014.921486

Boticario, J. G., Santos, O. C., Romero, C., Pechenizkiy, M., Merceron, A., Mitros, P., … Ventura, S. (2015). Educational Data Mining 2015: 8th International Conference on Educational Data Mining. In J. G. Boticario, O. C. Santos, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, … S. Ventura (Eds.), *Educational Data Mining 2015: 8th International Conference on Educational Data Mining*. Madrid.

Clark, T., & Oyer, E. (2005). Studying the Impact of Online Professional Development on Educators' Technology Integration Skills. In *Annual Meeting of the American Educational Research Association*.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

Cope, B., & Kalantzis, M. (2015). Interpreting Evidence-of-Learning: Educational research in the era of big data. *Open Review of Educational Research*, *2*(1), 218–239. http://doi.org/10.1080/23265507.2015.1074870

De Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., … Arnab, S. (2015). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology*, *46*(6), 1175–1188. http://doi.org/10.1111/bjet.12212

Dede, C., Breit, L., Ketelhut, D. J., Mccloskey, E., & Whitehouse, P. (2005). *An Overview of Current Findings From Empirical Research on Online Teacher Professional Development*. *Research Studies*.

Dede, C., Ketelhut, D. J., Whitehouse, P., Breit, L., & McCloskey, E. M. (2009). A research agenda for online teacher professional development. *Journal of Teacher*

Education, *60*(1), 8–19. http://doi.org/10.1177/0022487108327554

Desimone, L. M. (2009). Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures. *Educational Researcher*, *38*(3), 181–199. http://doi.org/10.3102/0013189X08331140

DiCerbo, K., & Behrens, J. (2014). *Impacts of the Digital Ocean on Education*. London: Pearson. Retrieved from http://digresearch.global2.vic.edu.au/files/2014/04/Impacts-of-the-Digital-Ocean-on-Education-1tqx8m2.pdf

Distefano, C., Zhu, M., & Mîndrilă, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20), 1–11. http://doi.org/10.1.1.460.8553

Dobrovolny, J., Edwards, D., Friend, B., & Harrington, C. (2015). Keeping Pace with K-12 Digital Learning: A Review of Policy and Practice. Retrieved from http://www.kpk12.com/reports/

Downer, J. T., Locasale-Crouch, J., Hamre, B., & Pianta, R. (2009). Teacher Characteristics Associated With Responsiveness and Exposure to Consultation and Online Professional Development Resources. *Early Education & Development*, *20*(3), 431–455. http://doi.org/10.1080/10409280802688626

Fisher, J. B., Schumaker, J. B., Culbertson, J., & Deshler, D. D. (2010). Effects of a Computerized Professional Development Program on Teacher and Student Outcomes. *Journal of Teacher Education*, *61*(4), 302–312. http://doi.org/10.1177/0022487110369556

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2011). *How to design and evaluate research in education*. New York: McGraw-Hill Humanities/Social Sciences/Languages.

Gerritsen, R. (1999). Assessing Loan Risks: A Data Mining Case Study. *IT Pro*, (November), 16–21.

Gobert, J. D., Pedro, M. A. S., Baker, R. S. J. d, Toto, E., & Montalvo, O. (2012). Leveraging Educational Data Mining for Real-time Performance Assessment of Scientific Inquiry Skills within Microworlds. *JEDM - Journal of Educational Data Mining*, *4*(1), 111–143. Retrieved from http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/24\nhttp://www.educationaldatamining.org/JEDM/index.php/JEDM/article/download/24/14\nhttp://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/24/14

Hendrickson, A., Ewing, M., & Kaliski, P. (2013). Evidence-Centered Design: Recommendations for Implementation and Practice. *Journal of Applied Testing Technology*, *14*(1), 1–27.

Hrastinski, S. (2008). What is online learner participation? A literature review. *Computers & Education*, *51*(4), 1755–1765. http://doi.org/10.1016/j.compedu.2008.05.005

Huff, K., & Plake, B. (2010). Evidence-Centered Assessment Design in Practice. *Applied Measurement in Education*, *23*(4), 307–309. http://doi.org/10.1080/08957347.2010.510955

Jansen, B. J. (Jim). (2009). Understanding User-Web Interactions via Web Analytics. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, *1*(1), 1–102. http://doi.org/10.2200/S00191ED1V01Y200904ICR006

Jiang, P., & Liu, X. S. (2015). Big data mining yields novel insights on cancer. *Nature Genetics*, *47*(2), 103–104. http://doi.org/10.1038/ng.3205

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, *2*(2), 130–144. http://doi.org/10.1016/j.edurev.2007.05.002

Kerr, D., & Chung, G. K. W. K. (2012). Identifying Key Features of Student Performance in Educational Video Games and Simulations through Cluster Analysis. *JEDM - Journal of Educational Data Mining*, *4*(1), 144–182. Retrieved from http://www.educationaldatamining.org/JEDM/images/articles/vol4/issue1/KerrEtAlVol4Issue1P111_152.pdf

Kim, D., Park, Y., Yoon, M., & Jo, I.-H. (2016). Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments. *The Internet and Higher Education*, *30*, 30–43. http://doi.org/10.1016/j.iheduc.2016.03.002

Kupczynski, L., Gibson, A. M., Ice, P., Richardson, J., & Challoo, L. (2011). The Impact of Frequency on Achievement in Online Courses: A Study from a South Texas University. *Journal of Interactive Online Learning*, *10*(3), 141–149. Retrieved from http://www.ncolr.org/jiol/issues/pdf/10.3.3.pdf

Losarwar, V., & Joshi, D. M. (2012). Data Preprocessing in Web Usage Mining. In *International Conference on Artificial Intelligence and Embedded Systems (ICAIES'2012) July*. Singapore: International Conference on Artificial Intelligence and Embedded Systems.

Miller, V. a, Reynolds, W. W., Ittenbach, R. F., Luce, M. F., Beauchamp, T. L., & Nelson, R. M. (2009). Challenges in measuring a new construct: perception of voluntariness for research and treatment decision making. *Journal of Empirical Research on Human Research Ethics : JERHRE*, *4*(3), 21–31. http://doi.org/10.1525/jer.2009.4.3.21

Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and Educational Data Mining. *Journal of Educational Data Mining*, *4*(1), 11–48. Retrieved from http://www.educationaldatamining.org/JEDM/images/articles/vol4/issue1/MislevyEtAlVol4Issue1P11_48.pdf

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*(1), 3–62. http://doi.org/10.1207/S15366359MEA0101_02

Munk, M., Kapusta, J., & Švec, P. (2010). Data preprocessing evaluation for web log mining: reconstruction of activities of a web visitor. *Procedia Computer Science*, *1*(1), 2273–2280. http://doi.org/10.1016/j.procs.2010.04.255

Nelson, B., Nugent, R., & Rupp, A. A. (2012). On Instructional Utility , Statistical Methodology , and the Added Value of ECD : Lessons Learned from the Special Issue. *Journal of Educational Data Mining*, *4*(1).

Rice, K., & Hung, J. (2015). Data mining in Online Professional Development Program Evaluation: An Exploratory Case Study. *International Journal of Technology in Teaching and Learning*, *11*(1), 1–20.

Rupp, A. A., Nugent, R., & Nelson, B. (2012). Evidence-centered Design for Diagnostic Assessment within Digital Learning Environments: Integrating Modern Psychometrics and Educational Data Mining. *Journal of Educational Data Mining*, *4*(1), 1–10. Retrieved from http://www.educationaldatamining.org/JEDM/images/articles/vol4/issue1/IntroVol4Issue1P1_10.pdf

Shute, V. J. (2010). ECD for Dummies. Retrieved May 20, 2015, from http://myweb.fsu.edu/vshute/ECD for Dummies/ECD for Dummies.swf

Tsai, C. W., Shen, P. Di, & Chiang, Y. C. (2013). Research trends in meaningful learning research on e-learning and online education environments: A review of studies published in SSCI-indexed journals from 2003 to 2012. *British Journal of Educational Technology*, *44*(6), 179–185. http://doi.org/10.1111/bjet.12035

Vakali, A., Pokorný, J., & Dalamagas, T. (2005). An Overview of Web Data Clustering Practices. In W. Lindner, M. Mesiti, C. Türker, Y. Tzitzikas, & A. I. Vakali (Eds.), *Current Trends in Database Technology - EDBT 2004 Workshops* (Vol. 3268, pp. 597–606). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/b101218

Valsamidis, S., Kontogiannis, S., Kazanidis, I., Theodosiou, T., & Karakos, A. (2012). A Clustering Methodology of Web Log Data for Learning Management Systems. *Journal of Educational …*, *15*, 154–167. Retrieved from http://www.ifets.info/download_pdf.php?j_id=55&a_id=1236

Wang, M. C., Dziuban, C. D., Cook, I. J., & Moskal, P. D. (2009). Dr. Fox Rocks: Using data-mining techniques to examine student ratings of instruction. In M. C. S. II, L. D. Yore, & B. Hand (Eds.), *Quality Research in Literacy and Science Education: International Perspectives and Gold Standards* (pp. 383–398). Dordrecht, The Netherlands: Springer Verlag. http://doi.org/10.1007/978-1-4020-8427-0_19

Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems*, *62*, 1–10. http://doi.org/10.1016/j.dss.2014.02.007

Webb, M., Gibson, D., & Forkosh-Baruch, A. (2013). Challenges for information technology supporting educational assessment. *Journal of Computer Assisted Learning*, *29*(5), 451–462. http://doi.org/10.1111/jcal.12033

West, D. (2012). Big Data for Education: Data Mining, Data Analytics, and Web Dashboards. Retrieved April 26, 2015, from http://www.brookings.edu/research/papers/2012/09/04-education-technology-west

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and TechniquesNo Title*. Burlington: Morgan Kaufmann Publishers.

Yong, A. G., & Pearce, S. (2013). A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94.

Zalles, Daniel; Gaertel, Geneva; Mislevy, R. J. (2010). *Using Evidence-Centered Design to Support Assessment , Design , and State Large-Scale Science Assessment*. Menlo Park. Retrieved from http://ecd.sri.com/downloads/ECD_TR10_Learning_Progressions.pdf

Zieky, M. J. (2014). An introduction of the use of evidence-centered design in test development. *Psicologia Educativa*, *20*(2), 79–87.

http://doi.org/http://dx.doi.org/10.1016/j.pse.2014.11.003

Appendix A

Click-Track Raw Data

Column Name	**Data Options**	**Explanation**
id	number	unique id of the row of log items
user_id	number	the id of the user (same as authentication id)
course_id	number	the id of the course (i.e. Middle School Math). Course names may be changed throughout the click track data saves but id numbers always remain the same; so always use id numbers to identify courses (same goes with all other components that use id numbers).
page_id	number	id number of the page within the course. Courses have multiple pages, these numbers uniquely identify individual pages.
page_type	Text	One of the options for the type of page; see rows belos
	Session	topic page inside a course
	Document	shared document in Google
	Resource	A resource page within canlead, shared resource or otherwise
	Course	course home page or topic page, pages related to course that are not resources or assignments
	Assignments	Task pages
	Session	Login or authentication related pages, not all refer to a viewed page, check action_name for better identification of what the page is doing
	Message	Dashboard page where messages are displayed
created_at	time	the time at which the person loads the page
updated_at	time	the time at which the person is still in the page but the page gets updated (checking to see if the user is still there)
last_tracked	time	Last time we know they were still looking at that page
session_start	time	The time when this particular session is started, it helps to group

ed		session activity since every activity in the session will have the same session start time. If more than 20 minutes of inactivity the session restarts.
previous_click_id	number	refers to an id number in one of these rows, refers to a row on this table that preceded this log. It helps with more accurate step by step follow of what user did.
controller_name	text	The name of the group of functions used in programming. We can ignore for data analysis, but useful for debugging.
action_name	text	The more detailed action that is happening since not all logs are page views
	show	View individual topic page
	index	list of sub topics or sub items, refer to controller name (for instance could be search, messages or other page)
	edit	showing the page editing a certain component
	update	an actual update has taken place
	update_title	google docs hook, working with google docs (we are not using this yet)
	new	reaching the page where "create" happens, the page you go to, to create something new but the new item creation has not yet happened.
	create	actually creating or adding an item
	clone	cloning function where course or topic is reproduced.
	discussions	viewing discussions
	topics	top level topic page
	destroy	session information is erased, not indicative of user activity.
	documents	viewing individual document page in a course
	collapse_state	
request_path	url	the url that was the subject of the action in this log, useful for understanding context.

Appendix B

Calculated Variables for Session Data

Column Name	**Data Options**	**Explanation**
session_id	number	unique id of the row
user_id	number	the user this session belongs to
session_start	time	Session start date-time
session_week	number (1-52)	Calendar week of the year for session create, so that we can group sessions into same weeks.
session_day	number (1-31)	Day of the month for session create for analyzing patterns in time of month
session_hour	number (0-24)	Hour of the day for session create for analyzing patterns for hour of the day.
session_duration	number in minutes	Total session duration
Page_type_home	number	Number of "home" page views in session
Page_type_course	number	Number of "course" page views in session
Page_type_assignment	number	Number of "assignment" page views in session
Page_type_resource	number	Number of "resource" page views in session
Page_type_document	number	Number of "document" page views in session
Page_type_message	number	Number of "message" page views in session
Page_type_archive	number	Number of "archive" page views in session
page_type_variety	number	Count unique page types visited in this session
page_variety	number	Count unique pages visited in this session
Action_name: Index	number	Count action items that are "index"
Action_name: new	number	Count action items that are "new"

Action_name: topics	number	Count action items that are "topics"
Action_name: show	number	Count action items that are "show"
Action_name: create	number	Count action items that are "create"
Action_name: documents	number	Count action items that are "documents"
Action_name: edit	number	Count action items that are "edit"
Action_name: update	number	Count action items that are "update"
Action_name: destroy	number	Count action items that are "destroy"
Action_name: update_titles	number	Count action items that are "update_titles"
Variety of action name	number	Count unique actions in this session