View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Boise State University - ScholarWorks

Boise State University
ScholarWorks

Educational Technology Faculty Publications and Presentations

Department of Educational Technology

1-1-2015

Data Mining in Online Professional Development Program Evaluation: An Exploratory Case Study

Kerry Rice
Boise State University

Jui-Long Hung
Boise State University

This document was originally published in the *International Journal of Technology in Teaching & Learning* by the Society of International Chinese in Educational Technology (SICET). Copyright restrictions may apply.

Rice, K., & Hung. J. (2015). Data mining in online professional development program: An exploratory case study. *International Journal of Technology in Teaching and Learning*, *11*(1), 1-20.

Data mining in Online Professional Development Program Evaluation: An Exploratory Case Study

Kerry Rice & Jui-Long Hung
Boise State University

This case study explored the potential applications of data mining in the educational program evaluation of online professional development workshops for pre K-12 teachers. Multiple data mining analyses were implemented in combination with traditional evaluation instruments and student outcomes to determine learner engagement and more clearly understand the relationship between logged activities and learner experiences. Data analysis focused on the following aspects: 1) Shared learning characteristics, 2) frequent learning paths, 3) engagement prediction, 4) expectation prediction, 5) workshop satisfaction prediction, and 6) instructor quality prediction. Results indicated that interaction and engagement were important factors in learning outcomes for this workshop. In addition, participants who had online teaching experience could be expected to have a higher engagement level but prior online learning experience did NOT show a similar relationship.

Keywords: Professional Development, Data Mining, Program Evaluation, Online, Engagement, Prediction

INTRODUCTION

The purpose of this case study was to explore the potential applications of data mining combined with other, more traditional evaluation instruments, in educational program evaluation. In this instance, the setting and context was an online teacher professional development (PD) program. Learning management system (LMS) data mining was implemented in combination with survey and demographic data to determine learner engagement with PD workshops in order to more clearly understand the relationship between logged activities and learner experiences – in other words, between user perceptions and user behaviors. Server logs, stored in the LMS that was used to deliver the online PD, recorded all learning activities as well as how learners interacted with instructors, peers, and course materials. Individual learning patterns were constructed by analyzing these logs with data mining techniques. These patterns contain information about learners' navigational behaviors, learning preferences, and interactions.

Online educational environments present some unique challenges and opportunities when attempting to judge the quality and effectiveness of an educational experience. On

Kerry Rice is a Professor and Jui-Long Hung is a Professor, in the Department of Educational Technology at Boise State University. Kerry Rice can be reached at krice@boisestate.edu

the one hand, it is difficult to observe learner behaviors visually because most, if not all activities, occur using asynchronous methods of instruction. On the other hand, data collected through the primary learning interface, typically an LMS, can provide detailed evidence of actual learner behaviors. The combination of these behavioral patterns combined with more traditional evaluation tools, such as self-report satisfaction and usefulness surveys, have the potential to generate a rich source of information for evaluating the effectiveness of instruction and improving the design of the online learning interface (Mor, Minguillon & Carbo, 2006).

LITERATURE REVIEW

EDUCATIONAL DATA MINING

Data mining (DM) is a series of data analysis techniques applied to extract knowledge from raw log data, and has been widely used in business and e-commerce for a variety of purposes, from improving the online customer experience to designing more personalized Internet purchasing profiles (Ngai, Xiu, & Chau, 2009; Levin & Zahavi, 2010; Zhang, Edwards, & Harding, 2007). Data mining has been defined as "the process of discovering patterns in data" (Witten & Frank, 2005, p. 5) and further as a practical approach for finding and describing structural patterns in data, as a tool for helping explain that data, and make predictions from it. Hastie, Tibshirani, and Friedman (2009) defined data mining as the computational process of discovering patterns in data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. In essence, data mining allows us to discover knowledge, in the form of hidden relationships and patterns, from vast amounts of data stored in database systems (Han, Kamber, & Pei, 2011).

Educational Data Mining (EDM) is an emerging discipline, concerned with developing methods derived from data mining for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings in which they learn (Baker, 2011; Siemens & Baker, 2010). Romero and Ventura (2007) reviewed educational data mining articles from 1995 to 2005 and concluded that the field of EDM was growing rapidly. Two major applications, relationship mining (reveals relationships between learning interactions) and prediction (identifies key predictors of learning behaviors or performances), were the most common approaches of EDM. Techniques such as data visualization, clustering, classification, association rule, and decision tree were the most popular. A meta-analysis by AlShammari, Aldhafiri, & Al-Shammari (2013) revealed that EDM could play a critical role in improving learning outcomes by providing more reliable information about how students learn. These EDM techniques empower educational researchers to present data visually, classify students by certain criteria, identify and monitor patterns of learning behaviors, and to predict learning outcomes and task performance accordingly (Hung & Crooks, 2009; Hung & Zhang, 2008; Lim & Morris, 2009; Romero & Ventura 2008).

EDM can be a useful tool for a multitude of stakeholders and a variety of educational purposes. Users of EDM might include learners, educators, course developers, organizations, educational researchers, and administrators. Uses include personalizing learning, risk analysis, performance prediction, adaptive learning, improved course design, and enhanced the decision-making process (Romero & Ventura, 2010). The literature on EDM provides valuable insights into the factors that influence learning outcomes and those that allow us to make performance outcome predictions. Factors include learner engagement measures such as logins and time spent in the LMS (Hung, Hsu, & Rice, 2012; Hung & Zhang, 2008), gender, age, course load, and number of discussion posts (Lowes, Lin, & Kinghorn, in press; Macfadyen & Dawson, 2010; Yasmin, 2013; Hung, Hsu, & Rice, 2012; Hung & Crooks, Hung & Zhang, 2008), Other studies have combined data mining with other forms of data such as demographics, past course registrations, reasons for enrollment, course structure, and traditional end-of-course evaluation instruments to identify predictors of performance outcomes (Calvert, C., E., 2014; Dejaeger, Goethals, Giangreco, Mola, & Baesens, 2012; Guo, 2010; Hung, Rice, Saba, 2012; Xu & Recker, 2012). While there are limited studies in K-12 education, there have been a growing number of studies set in the context of post-secondary education. No studies could be found examining the potential for data mining in PD program evaluation.

EDUCATIONAL PROGRAM EVALUATION

Educational program evaluation has been used for a variety of purposes but its primary function is to enable decision-making "about the applicability or worth of something in a situation" (Krathwohl, 2004, p. 588). Borrowing from research methods employed in the social sciences, any number of approaches may be engaged depending on the purpose of the evaluation and its intended audience, but the intent is to inform decision-makers on whether or not the program under review is meeting its intended goals (Nichols & Nichols, 2000). A common form of educational program evaluation is the use of survey analysis methods (Reeves & Pedulla, 2011). Although survey methods alone may be useful in evaluating learners' perceptions of satisfaction and learning, how well these one-time assessments reflect actual learner behaviors, actions, and performance has been questioned (Astin & Lee, 2003; Hung & Crooks, 2009; Hung, Hsu, & Rice, 2012; Picciano, 2002; Xu & Recker, 2012). In the search for increased rigor and practical solutions, systematic frameworks for program evaluation that include more in depth analysis of learner experiences have been proposed (Earley & Porritt, 2014; Yurdakul, Uslu, Çakar, & Yildiz, 2014). However, these can be challenging to implement, especially in programs that use online delivery of instruction. EDM is a technique that offers promise in generating a rich source of information about the hidden relationships between user perceptions and user behaviors and thus improving the accuracy and credibility of traditional program evaluation frameworks.

With respect to this study, it is important to note that evaluation differs from typical research in one very significant aspect; "Because it is decision-driven, the value of an evaluation lies in its usefulness" and "because utilization requires trust in the results, the process of evaluation may be as important as the product" (p. 588). The intent of this study was to examine and judge the value of incorporating data mining into the evaluation process with the use of other more traditional data sources.

SETTING AND CONTEXT: ONLINE PROFESSIONAL DEVELOPMENT

The setting and context for this case study was a series of online PD workshops delivered through the Blackboard LMS. Online PD provides convenient access to training for teachers who might not otherwise be able to participate because of time and/or geographic location factors. The benefits go beyond flexibility and cost savings. Online PD also presents opportunities for teachers to strengthen communication and collaboration activities within their profession by encouraging support from other teachers and the sharing of resources and materials (Fusco & Schlager, 2003). Studies on online PD indicate that teachers gain diverse competencies, new ideas, motivation, self-directedness, and a greater awareness of computer technologies from participation in Web-based learning environments (Bennett, Priest & Mcpherson, 1999; Rodes, Knapczyk, 2000; Spratt, Palmer & Coldwell, 2000).

Although research about online PD has been an emerging field, very little is known about empirically identified practices and methods for effectively training teachers in online environments. However, studies have shown positive teacher outcomes on knowledge and instructional practices (O'Dwyer et al, 2010). Moreover, in a study, examining the relationship between satisfaction and a variety of other variables identified in the literature as important in the design and implementation of online PD, Reeves and Padulla (2011) identified predictors of success through an empirical process of data collection and analysis. However, as online PD programs aim to provide customized learning, it is critical to "observe" and identify learner behaviors in order to generate information for evaluating and improving the design of professional development programs (Mor, Minguillon & Carbo, 2006).

One criticism raised with PD has been with the PD itself. Earley & Poritt (2014) outline several concerns including teacher propensity to participate passively in their own learning. This lack of engagement can be difficult to measure objectively in traditional face-to-face instructional methods. However, with access to data on learner behaviors in online environments, engagement variables can be identified and engagement factors can be coded into the analysis, thus reflecting actual learner engagement with the content. This focus on engagement has been shown to be an important element in predicting learner outcomes in online courses. Learner outcomes and performance are highly related to their engagement level in an online course, and a consistent engagement level is highly correlated with success (Shelton, Hung, & Baughman, 2015).

PROGRAM GOALS AND PURPOSE

The event for this case study was implemented as a statewide professional development workshop portal designed to deliver learning opportunities to teachers. The goals of the program included:

- Provide just-in-time professional learning;
- Create a statewide professional learning community;
- Create job-embedded learning opportunities;
- Ensure content-rich learning opportunities;
- Expand personalized professional development;
- Increase access to professional learning experiences; and
- Leverage economies of scale to reduce the costs of professional development statewide.

Ten workshops were developed or purchased for PD and delivered online in the spring of 2010. Each workshop lasted seven weeks and included a one-week orientation. The following evaluation goals were identified for the workshops:

- Program evaluation of workshop quality for continuous improvement of design and delivery.
- Evaluation of impact on teaching practice as a result of the training.

The program evaluation consisted of participant self-report data from surveys (see Appendix A), analyses of participants' completed work samples and LMS data mining. Data was collected from the Blackboard LMS, which was used to deliver the professional development. Identified variables provided information about the quantity and quality of access to instructional content and participation. These variables were analyzed independently and in conjunction with the data from the surveys (e.g. previous experience) and performance data (e.g. retention data, workshop grades, project grades) to better understand workshop effectiveness, learner outcomes and learner satisfaction. In particular, evaluators were interested in participant satisfaction with the training, the level of engagement participants experienced with the training, and perceived impact on teaching practice as a result of the training. For this study the research questions focused on exploratory data analysis and included: 1) what usage patterns, clusters and predictive variables emerge when mining teacher PD data? and 2) what relationships exist between LMS behaviors, usage patterns, end of course evaluation data, learner outcomes, and demographics?

METHOD

PARTICIPANTS

There were a total of 142 enrolled participants. Of these, three participants withdrew prior to completion and did not receive a grade for the workshop. One hundred twenty-three (87%) enrolled participants passed the workshop they attended with a score of 80% or better, while 16 (11%) failed. Of the 142 enrollments past the drop deadline, 111 responded to the end-of-workshop survey for a response rate of 78%. Since 8 participants attended more than one workshop only data for the remaining 103 participants was collected for data analysis using data mining. In total, 31,417 learning logs collected from Blackboard.

Respondents were geographically diverse and represented primarily rural school districts. The vast majority of respondents indicated their job role as Teacher (92%). Additional roles included Special Education Teachers (6%), Library Media Specialists (2%), Curriculum Coordinators (2%), Department Heads (4%), and Other (11%). Seven respondents indicated that they were responsible for multiple job roles (Teacher and Library Media Specialist). Years of experience in education was evenly distributed among the following: 0-5 years (26%), 6-10 years (24%), and 11-15 years (20%). However, 30% of respondents indicated 16 or more years of experience. Eighty-six of the respondents (77%) indicated that they worked directly with students across all grade levels with the majority at the high school level.

DATA MINING PROCEDURES

The term "data mining", as often used by statisticians and database researchers, represents the application of specific algorithms to extract patterns or models from the data. Data mining is also considered a step in the Knowledge Discovery in Databases (KDD) process (Fayyad, Pitatesky-Shapiro, Smyth & Uthurasamy, 1996; Han, Kamber, & Pei, 2011) which consists of five major steps—data collection, data preprocessing, data transformation, data mining, and interpretation and evaluation. KDD is a set of specific processes or activities that are necessary for making sense of the data; in other words, for generating knowledge from the data. Data mining, the application of algorithms for extracting patterns, is at the core of this process. Thus, researchers usually follow the KDD process when conducting data mining studies. In addition, when a specific technique is chosen for the analysis of large amounts of data stored in several locations, data extraction and data preparation become the most time-consuming parts of the knowledge discovery process. Table 1 summarizes the major tasks at each step of the KDD process (Fayyad et al.; Roiger & Geatz, 2003).

Table 1. Steps and tasks in the KDD process

Step	Major task	Sub-tasks	End Products or Outcomes
Step 1	Data collection	• Identify the goal • Identify the domain for knowledge discovery • Generate appropriate purpose statement, hypotheses, and desired outcomes • Create a target data set for analysis	Target data set
Step 2	Data preprocessing	• Identify missing data values • Filter noisy data • Determine on time-sequence information	Preprocessed data
Step 3	Data transformation	• Add or delete attributes and instances • Select and apply the transformation methods to normalize, convert, and smooth data as needed	Transformed data
Step 4	Data mining	• Apply one or more DM algorithms • Generate a best-fit model to represent data	Model(s) or Pattern(s)
Step 5	Interpretation and evaluation	• Evaluate output of step 4 • Determine whether or not the discovered knowledge is useful • Determine if it is necessary to repeat previous steps with new attributes and/or instances • Incorporate and apply discovered knowledge to problems	Knowledge application(s) to problems

DATA COLLECTION

A survey for this project was developed based on recommendations from an earlier session, and administered using an in-house evaluation tool, which was integrated into the Blackboard LMS (see Appendix A). This evaluation was undertaken at the conclusion of the ten workshops delivered in spring 2010. It is based on the following sets of data:

- Participant demographics.
- Satisfaction questions related to ease of use, time, and effort in workshop activities, facilitator effectiveness, and workshop expectations.
- Content specific questions related to increased knowledge of the subject matter.
- Ten workshops covering a variety of content areas including: Best Practices for Vocabulary Instruction in the Elementary Classroom; Helping Struggling Readers Improve Comprehension; Promoting Reading Comprehension in the Elementary Classroom; Teaching Writing in the Elementary Classroom; Teaching Writing in the Middle School Classroom; Using Patterns to Develop Algebraic Thinking; Designing a Virtual Field Trip; Learning and Teaching with Web 2.0 Tools; Differentiating Instruction to Accommodate Learning Styles; Transforming the Classroom with Project-Based Learning
- Server Log Data: Identified variables provided information about the quantity and quality of access to instructional content and participation. Data was collected at the course level and at the student level. Collecting data in this way allowed us to develop a learner profile and provided valuable information about individual learning paths for further improving the structure and quality of the workshops.

- Primary variables for data mining included: Logins with timestamp and duration (logout or session ended); Learning activities including content, discussions, and tools with sequential timestamps; Duration of time spent in the LMS.
- Derived Engagement Variables included: Average frequency of logins per course; Average frequency of content accessed; Length of time online (self-report and data mining); Average number of discussion board entries.

RESULTS

Learning log data was combined with survey data and analyzed using clustering analysis, sequential association analysis, and decision tree analysis. SAS Enterprise Miner (SAS Institute Inc., USA) was employed to perform data analysis which focused on the following aspects: 1) shared learning characteristics, 2) frequent learning paths, 3) engagement prediction, 4) expectation prediction, 5) workshop satisfaction prediction, and 6) instructor quality prediction.

SHARED LEARNING CHARACTERISTICS

Learning activities of 103 participants were aggregated and classified into five groups (CL1 to CL5) based on their shared learning characteristics and survey responses. Sixty percent of participants logged in an average of 27.10 times, accessed content 61.03, posted 40.81 discussions, and spent 8 hours and 28 minutes in the LMS over the course of the seven week workshop (Table 2). Compared with the results from the survey, on average participants perceived a longer amount of time spent online than the actual time spent.

Table 2. Shared learning characteristics for seven-week workshop

CL	FREQ	Logins	Content Access	Discussion Entry	Duration	Q10_4	Q11	Q13_2	Q13_3	Q14_5
1	31	40.90	81.84	75.13	27:32:55	2.42	2.81	1.84	1.90	1.77
2	2	49.00	110.00	161.00	104:40:35	1.50	3.00	1.50	1.50	2.00
3	7	44.29	111.71	72.86	64:22:44	2.29	3.57	2.00	1.86	1.29
4	1	28.00	101.00	46.00	133:37:09	2.00	3.00	1.00	2.00	1.00
5	62	27.10	61.03	40.81	8:28:22	2.27	2.42	1.84	2.15	1.42

FREQUENT LEARNING PATHS

Based on recorded learning behaviors in the LMS, participants' learning paths within one session were generated by sequential association rule analysis. Support and confidence are two terms used to describe the association rules discovered in an association analysis. Support refers to the proportion of all server logs containing learning behaviors defining the rule within one session. Confidence refers to the probability that the entire rule will be observed given the occurrence of the precedent behavior(s) in the sequence. For example, the association rule: "discussion board entry ==> content ==> discussion board entry" has a support score of 82.63% and a confidence score of 96.40% (see Table 4, rule 1). In this example, support means that 82.63% of the server logs within one session contained the following learning behaviors "discussion board entry, content, and discussion board entry" The confidence score means that the probability is .9649 that the learning path "discussion board entry ==> content ==> discussion board entry" will occur if the behaviors "discussion board entry, content, and discussion board entry" has occurred within one session.

A total of 31,417 server logs were extracted for association rule analysis and generated 23 rules. Four association rules (1 to 4) were extracted to depict participants' online learning paths and are listed in Table 4. Instead of focusing on content or discussion only, the results indicate participants tended to switch between content and discussion within one session because the first two rules have higher support and confidence rates than rules three and four. The results also show that different types of interactions (content-participant, participant-instructor, and participant-participant) were well facilitated in the workshops overall.

Table 3. Top daily learning paths

No	Support (%)	Confidence (%)	Rule
1	82.63	96.40	discussion board entry ==> content ==> discussion board entry
2	81.47	98.60	discussion board entry ==> content ==> discussion board entry ==> content
3	44.66	45.17	content ==> content ==> content
4	41.60	49.32	discussion board entry ==> discussion board entry ==> discussion board entry

ENGAGEMENT

Clustering techniques were applied to identify survey questions that were highly related to student online engagement. Online engagement variables included the number of logins, the duration of the login, the number of discussion board entries, and the frequency of content accessed. Figure 1 indicates Q11 (how many hours per week did you spend online, engaged in the workshop activities?), Q12 (how many hours per week did you spend offline, doing homework and completing activities?), and Q4-2 (Online teaching experience) are highly related to online engagement (logins, duration, discussion entry, and content access). The results indicated that the more time respondents spend online, the more time they spent offline as well. Additionally, it was found that respondents who had previous online teaching experience could be expected to spend more hours per week online and offline – indicating a higher engagement level.

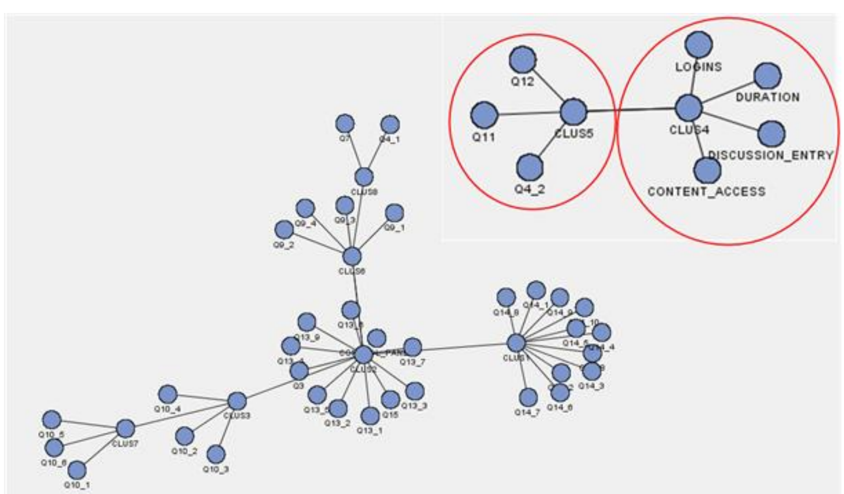

Figure 1. Classification of survey questions and online engagement behaviors based on similarity of participants' responses.

EXPECTATIONS

Decision Tree techniques were applied to survey questions to identify important factors that might influence participants' expectations. The splitting standard was based on Probability F (significance level = 0.2) (Saar-Tsechansky & Provost, 2004). The growing and pruning process for the decision tree analysis followed the steps suggested by Quinlan (1986). Results revealed that Q13-5 (Workshop has produced new knowledge, skills and, awareness.), 13-9 (I would take another workshop.), Q9-2 (The assignments in this workshop helped me produce materials that will be useful in my classroom or school.), and Q10-4 (Information on the portal website was easy to locate.) were important factors in predicting participants' responses to questions about workshop expectations. For example, respondents who "Agreed or "Strongly Agreed" that the workshop helped them produce materials for the classroom, were more likely to respond that the workshop met or exceeded their expectations.

WORKSHOP SATISFACTION

The same Decision Tree techniques that were used to determine factors influencing expectations were applied to identify important factors that might influence workshop satisfaction. The results revealed that Q10-5 (I was able to easily locate tech support services for my technical issues.), Q10-4 (Information on the portal website was easy to locate.), and 13-5 (Workshop has produced new knowledge, skills and, awareness) were important factors in predicting responses to questions about workshop satisfaction (Q10 responses).

INSTRUCTOR QUALITY

Finally, the same Decision Tree techniques were applied to identify important factors that might influence participant perceptions of instructor quality. The results revealed that Q14_5 (The instructor stimulates interest in subjects.), Q14-2 (The instructor is well prepared for class.), Q14_9 (The instructor treats students with respect.), Q13_8 (Peer collaboration was offered.), Q13-4 (Assessments and/or products reflect course objectives.), and Q10-6 (Support services were helpful in resolving my technical difficulties.) were important factors to predict overall responses to questions about instructor quality (Q14 responses). Three additional questions (Q13-8, Q13-4, & Q10-6) were also important predictors of respondents' perceptions of instructor quality.

PERFORMANCE PREDICTION

A predictive model was constructed by decision tree technique to predict participants' pass/fail rate (Figure 2). The splitting standard was based on Probability F (significance level = 0.2) (Saar-Tsechansky & Provost, 2004). The growing and pruning process for the decision tree analysis followed the steps suggested by Quinlan (1986). Because none of participants who failed a workshop filled out the evaluation survey, only engagement variables such as frequency of logins, frequency of content access, and number of discussion posts were imported for model construction. Frequency of logins was the only significant factor in the model. If participants logged in more than 10 times over six weeks, the pass rate increased from 89% to 94%. When the frequency increased to 17 times, the pass rate improved from 94% to 98%.

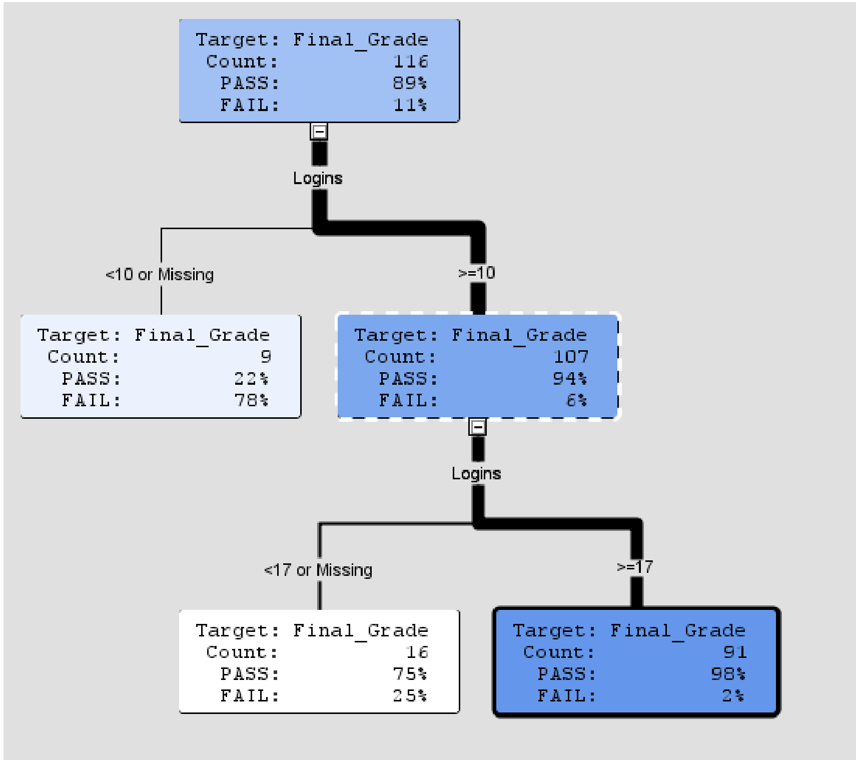

Figure 2. Pass Rate Predictive Model.

Similar results can be found in the predictive model of participants' final grades (Figure 3). The average of participants' final grades improved from 88% to 92% when participants' logged into the LMS more than 10 times over six weeks. The average further improved to 98% when frequency of logins increased to 17 times.

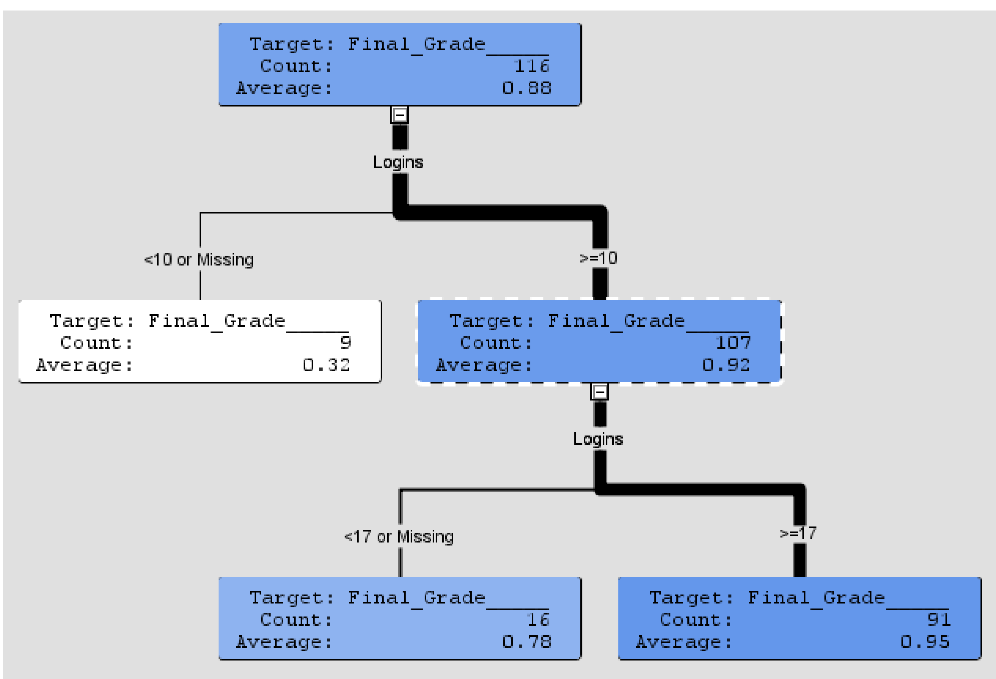

Figure 3. Pass Rate Predictive Model.

DISCUSSION

For this study, LMS data mining was implemented in combination with traditional evaluation instruments to determine learner engagement with the workshops and more clearly understand the relationship between logged activities and learner experiences. Possibly the most significant conclusion that can be drawn from this analysis is the importance of activities that encourage learners to interact with the course content, with their peers and with the instructor. For example, Association Rule analysis revealed higher support and confidence ratings for participant learning paths that included both discussion forums and content access as opposed to those in which only content was accessed. The results also revealed the inclusion of a variety of interactions (content-participant, participant-instructor, and participant-participant) indicating that the workshops were well facilitated, in terms of interaction, overall. In addition, several findings indicated a direct relationship between the amount of time learners spent online and their average course logins to engagement and performance. Specifically, more time spent online and a higher frequency of logins equated to increased engagement and improved performance. Taken as a whole, we can postulate that interaction and engagement were important factors in learning outcomes for this workshop. On a related note, results indicated that participants who had online teaching experience could be expected to have a higher engagement level but prior online learning experience did NOT show a similar relationship.

Researchers were also interested in predictive variables and indicators in this evaluation. Based on respondents' perceptions as indicated in their survey responses, it was found that if respondents agreed that the workshop helped them produce materials for the classroom, they were more likely to respond that the workshop met or exceeded their expectations. Overall results indicated that two factors influenced respondents' perceptions about workshop expectation ratings: 1) Practical application of new knowledge, and 2) ease of locating information. Three factors were found to influence workshop satisfaction ratings: 1) Usefulness of the subject matter, 2) a well-structured website, and 3) sufficient technical supports. Finally, results indicated that respondents' perceptions of instructor quality were highly related to the following survey items: 1) The instructor stimulates interest in subjects, 2) the instructor is well prepared for class, 3) the instructor treats students with respect, 4) peer collaboration was offered, 5) assessments and/or products reflect course objectives, and 6) support services were helpful in resolving my technical difficulties.

CONCLUSION

This study explored the potential applications of data mining in support of online PD. Two important outcomes were expected as a result. First, it allowed us to begin the process of developing a model for utilizing data as a predictor of learner performance. This outcome is valuable in the creation of warning or recommendation systems that can be used to notify both instructors and students of behaviors that may result in unsatisfactory course outcomes. Second, statistical data demonstrating how learners engage with course materials can lead to improved course design, adjustment of learning strategies, and improvement in learner performance through individualized support mechanisms.

Within this educational program evaluation, we established a secondary target, focused on the need for practical yet rigorous ways in which to evaluate the impact of PD against its intended aims and outcomes. Evaluating the impact of PD is a challenging task. While we acknowledge that teacher PD needs to be evaluated, our lack of tools and skills often results in evaluation for satisfaction with little attention paid to the impact on changes in teaching practice or student outcomes (King, 2014). This exploratory case study sets the stage for future research to improve our analysis of the impact of PD. For example, can data mining help us establish a link between teacher engagement with PD and the outcomes of the learners that they teach? It is expected that a greater understanding of the potential to leverage data collected everyday through LMS's have the potential to inform decisions regarding effectiveness at all levels of education.

REFERENCES

Astin, A. W. & Lee, J. J. (2003). How risky are one-shot cross-sectional assessments of undergraduate students? *Research in Higher Education, 44*(6), 657–672.

AlShammari, I. A., Aldhafiri, M. D., & Al-Shammari, Z. (2013). A meta-analysis of educational data mining on improvements in learning outcomes. *College Student Journal*, *47*(2), 326-333.

Baker, R.S.J.d. (2011) Data mining for Education. To appear in McGaw, B., Peterson, P.,Baker, E. (Eds.) International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier.

Bennett, S., Priest, A. and Macpherson, C. (1999). Learning about online learning: An approach to staff development for university teachers. *Australian Journal of Educational Technology, 15*(3), 207-221.

Calvert, C. E. (2014). Developing a model and applications for probabilities of student success: a case study of predictive analytics. *Open learning, 29* (2), p. 160 – 173.

Chen, L.D., Sakaguchi, T. & Frolick, M. N. (2000) Data mining methods, applications, and tools, Information systems management, 17(1), 65-70.

Dejaeger, K., Goethals, F., Giangreco, A., Mola, L., & Baesens, B. (2012). Gaining insight into student satisfaction using comprehensible data mining techniques. *European Journal of Operational Research, 218* (2), 548-562.

Earley, P, & Porritt, V. (2014). Evaluating the impact of professional development: the need for a student-focused approach. *Professional Development in Education, 40*(1) 112-129, DOI: 10.1080/19415257.2013.798741

Fayyad, U.M., Piatetsky-Shapiro, G., & Smyth, P.. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM, 39*(11), 27-34.

Fusco, J., & Schlager, M.S., (2003, July/August). Teacher professional development, technology, and communities of practice: Are we putting the cart before the horse? *Information Society, 19*(3) 203 – 224.

Guo, W.W. (2010). Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction. Expert Systems with Applications, (37) 4, 3358-3365.

Han, J., Kamber, M., Pei, J. (2011). Data mining: Concepts and techniques: Concepts and techniques. Elsevier.

Hung, J. L., & Crooks, S. (2009). Examining online learning patterns with Data Mining techniques in peer-moderated and teacher-moderated course. *Journal of Educational Computing Research, 40*(2), 183-210.

Hung, J. L., Hsu, Y.-C., & Rice, K. (2012). Integrating data mining in program evaluation of K-12 online education. *Educational Technology & Society, 15*(3), 27-41.

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.

King, F. (2014). Evaluating the impact of teacher professional development: An evidence-based framework. *Professional Development in Education, 40,* (1), 89-111.

Krathwohl, D. R. (2004). Methods of educational and social science research: An integrated approach (2nd Ed). Long Grove, IL: Waveland Press.

Levin, N., & Zahavi, J. (2010). Data Mining for Target Marketing, In O. Maimon and L. Rokach (eds.), *Data Mining and Knowledge Discovery Handbook* (2nd ed.) (pp. 1189-1220), New York, NY: Springer.

Lim, D. H., & Morris, M. L. (2009). Learner and instructional factors influencing learning outcomes within a blended learning environment. *Educational Technology & Society*, 12 (4), 282–293.

Lowes, S., Lin, P., & Kinghorn, B. (In press). Exploring the link between online behaviours and course performance in asynchronous online high school courses. *Journal of Learning Analytics 2*(2).

Macfadyen, L. P. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers and education, 54* (2), p. 588 – 599.

Marjamaa, J. (2003). Managing your customer resources through data mining. *Community Banker, 12*(4), 28-30

Mor, E., Minguillón, J., & Carbó, J.M. (2006). Analysis of user navigational behavior for e-learning personalization. In C. Romero & S. Ventura, (Eds.), *Data mining in e-learning* (pp. 41-56). Billerica, MA: WitPress.

Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: a literature review and classification. *Expert Systems with Applications*, *36*, 2592-2602.

O'Dwyer, L.M., et al. (2010). Effects of on-line professional development on teachers and their students: findings from four randomized trials. Retrieved from http://www.bc.edu/research/intasc/PDF/EFE_Findings2010_Report.pdf

Picciano, A. G. (2002). Beyond student perceptions: Issues of interaction, presence, and performance in an online course. *Journal of Asynchronous Learning Networks*, *6*(1), pp. 21–40.

Quinlan, J. R. (1986). Induction of decision trees, *Machine Learning, 1*(1), 81–106.

Reeves, T. D. & Pedulla, J. J. (2011). Predictors of teacher satisfaction with online professional development: evidence from the USA's e-Learning for Educators initiative. *Professional Development in Education, 37*(4), 591-611 DOI: 10.1080/19415257.2011.553824

Romero, G., Ventura, S., & Garcia, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education, 51*(1), 368-384.

Romero, C. and Ventura, S. (2007). Educational data mining: a survey from 1995 to 2005. *Expert Systems with Applications, 33,* 125-146.
Rodes, P., Knapczyk, D., Chapman, C., & Chung, H. (2000). Involving teachers in web-based professional development. *T H E Journal, 27*(10), 94-99.
Roiger, R., & Geatz, M. (2003). *Data mining: A tutorial-based primer*. Boston: Addison Wesley.
Saar-Tsechansky, M., & Provost, F. (2004). Active sampling for class probability estimation and ranking. *Machine learning*, *54*(2), 153-178.
Shelton, B., Hung, J. L., & Baughman, S. (2015). Online graduate teacher education: Establishing and EKG for student success intervention. *Technology, Knowledge and Learning*.
Siemens, G., & Baker, R. S. J.d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings from the 2nd International Conference on Learning Analytics and Knowledge*, April 29-May 2, 2012, Vancouver, British Columbia, Canada.
Srivastava, J., Cooley, R., Deshpande, M., & Tan, P.N. (2000). Web usage mining discovery and applications of usage patterns from web data. *ACM SIGKDD Explorations, 1*(2). Retrieved from http://www.sigkdd.org/explorations/issues/1-2-2000-01/srivastava.pdf
Spratt, C., Palmer, S. & Coldwell, J. (2000). Using technologies in teaching: an initiative in academic staff development. *Educational Technology & Society, 3*(3).
Witten, I. H., & Frank, E. (2005). Data mining: Practical machine learning tools and techniques. Burlington, MA: Morgan Kauffman.
Yasmin, D. (2013). Application of the classification tree model in predicting learner dropout behaviour in open and distance learning. *Distance Education, 34*(2), 218-231.
Yurdakul, B., Uslu, Ö., Çakar, E., & Yildiz, D. G. (2014). Evaluation of the Professional Development Program on Web Based Content Development. *Educational Sciences: Theory & Practice*, *14*(5), 1427-1437. doi:10.12738/estp.2014.4.2131
Xu, B., & Recker, M. (2012). Teaching Analytics: A Clustering and Triangulation Study of Digital Library User Data. *Educational Technology & Society, 15* (3), 103–115.
Zhang, X., Edwards, J., & Harding, J. (2007). Personalised online sales using web usage data mining. *Computers in Industry*, *58*, 772-782.

Appendix A

Spring 2010 Professional Development Satisfaction Survey

1- Your Name (Optional)
2- In which district are you employed?
3- How many years of experience do you have in the field of education?
  1) 0-5 years
  2) 6-10 years
  3) 11-15 years

4) 16 or more years

Please rate your level of experience in online environments.
4-1 Taking online courses.
1) Multiple experiences
2) One or two experiences
3) Limited one-time experience
4) No experiences

4-2 Teaching online courses.
1) Multiple experiences
2) One or two experiences
3) Limited one-time experience
4) No experiences

5- What is your current job role? (Choose all that apply)
1) Teacher
2) Special Education Teacher
3) Student Teacher
4) Library Media Specialist
5) Technology Coordinator
6) Curriculum Coordinator
7) Department Head
8) Principal
9) Vice Principal
10) Superintendent
11) Other

6- If you work directly with students which grade levels apply? (Choose all that apply)
1) Pre-K
2) K
3) 1
4) 2
5) 3
6) 4
7) 5
8) 6
9) 7
10) 8
11) 9
12) 10
13) 11
14) 12
15) Dual Credit
16) Advanced Placement

7- Will you use technology more, less or the same in your teaching or in your work with learners as a consequence of your participation in this online workshop?
1) More
2) Less
3) Same

8-1 If you chose "More" above, which tools will you use more of? (Choose all that apply)
1) Application tools (e.g. MS Word, Excel, PowerPoint)
2) Internet research tools (e.g., search engines, online encyclopedias)
3) Internet communication tools (e.g., discussion forms, email tools)
4) Web 2.0 tools (e.g., blogs, wikis)
5) Learning Management Systems (e.g., Blackboard, Moodle)
6) ePortfolios

8-2 If you chose "More", please list any additional tools you might use and also explain how you will use the tools you have chosen above. If you chose "Less" or "Same" above, please explain why.

Please indicate how strongly you agree or disagree with the following statements about the subject-matter of this workshop.

9-1 I found the subject-matter of this workshop valuable.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

9-2 The assignments in this workshop helped me produce materials that will be useful in my classroom or school.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

9-3 I will use what I have learned in my classroom or school.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

9-4 I expect to see improved student outcomes as a result of my participation in this workshop.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

Please indicate how strongly you agree or disagree with the following statements.

10-1 The workshop registration process was easy and intuitive.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

10-2 I was able to easily locate my workshop on the portal.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

10-3 Information on the portal was helpful.
1) Strongly Agree
2) Agree

3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

10-4 Information on the portal was easy to locate.

1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

10-5 I was able to easily locate tech support services for my technical issues.

1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

10-6 Support services were helpful in resolving my technical difficulties

1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

11- On average, approximately how many hours per week did you spend online, engaged in the workshop activities?

1) 1 or less
2) 2-3
3) 4-5
4) 6-7
5) 8 or more

12- On average, approximately how many hours per week did you spend offline, doing homework and completing activities?

1) 1 or less
2) 2-3
3) 4-5
4) 6-7
5) 8 or more

Please indicate how strongly you agree or disagree with the following statements about the quality of the workshop.

13-1 Workshop objectives are clear.

1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

13-2 Workshop is well organized.

1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

13-3 Student responsibilities are clearly defined.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

13-4 Assessments and/or products reflect course objectives.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

13-5 Workshop has produced new knowledge, skills and, awareness.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

13-6 The amount of work required was in line with expectations.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

13-7 The readings were aligned with course objectives.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

13-8 Peer collaboration was offered.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

13-9 I would take another workshop from Idaho PD.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

Please indicate how strongly you agree or disagree with the following statements about the quality of the instructor.

14-1 The instructor responded to questions in a timely manner.
1) Strongly Agree
2) Agree
3) Neutral

4) Disagree
5) Strongly Disagree
6) N/A

14-2 The instructor is well prepared for class.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

14-3 The instructor has a thorough knowledge of the subject.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

14-4 The instructor provides opportunity for discussion.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

14-5 The instructor stimulates interest in subjects.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

14-6 The instructor communicates ideas and information clearly.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

14-7 The instructor maintains a classroom conducive to learning.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

14-8 The instructor is genuinely interested in helping students.
1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

14-9 The instructor treats students with respect.
1) Strongly Agree
2) Agree

3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

14-10 I would take another class from this instructor.

1) Strongly Agree
2) Agree
3) Neutral
4) Disagree
5) Strongly Disagree
6) N/A

15- Based on your expectations for this workshop, which statement best describes your experience?

1) This workshop did not meet my expectations
2) This workshop met my expectations
3) This workshop exceeded my expectations

16- Based on your experience in this workshop, what uses and opportunities do you see for online professional development in your school, district, or organization?

17- What were the advantages (if any) of participating in an online workshop (versus a face-to-face seminar for example)?

18- What were the disadvantages (if any) of participating in an online workshop (versus a face-to-face seminar for example)?

19- Will the experience of taking an online workshop affect, in any way, your traditional classroom teaching or your work with learners in the schools?

1) Yes
2) No

20- If so, in what ways will your teaching or work with learners be affected?

1) Improved communication with students.
2) Increased empathy for students.
3) Increased awareness of student needs.
4) Increased use of reflection and other metacognitive strategies.
5) Increased use of collaborative activities.
6) Increased awareness of the need to build community.
7) More effective teaching strategies.
8) N/A

21- Please share any other ways in which your teaching or work with learners will be affected as a result of taking an online workshop?

22- Name three things that you learned about the subject-matter as a result of this workshop.

23- How do you plan to use what you have learned in this workshop?

24- Share any additional comments and feedback about the portal. Please be specific.

25- Share any additional comments and feedback about the workshop design. Please be specific.

26- Share any additional comments and feedback about the quality of the workshop instructor. Please be specific.

27- Please explain your answer above and share any additional comments or feedback about your expectations for this workshop.

28- Optional: We are interested in seeing how what you have learned will be applied in your classroom, school, or organization. Please provide contact information if you are willing to be contacted for follow-up questions. Thank you.

Copyright of International Journal of Technology in Teaching & Learning is the property of Society of International Chinese in Educational Technology (SICET) and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.