Predicting STEM Achievement with Learning Management System Data: Prediction Modeling and a Test of an Early Warning System

Michelle M. Dominguez
Dept. of Educational Psychology & Higher Education, University of Nevada, Las Vegas
4505 S Maryland Parkway #3003
Las Vegas, NV 89154, USA
+001(702)895-3253
doming90@unlv.nevada.edu

Matthew L. Bernacki
Dept. of Educational Psychology & Higher Education, University of Nevada, Las Vegas
4505 S Maryland Parkway #3003
Las Vegas, NV 89154, USA
+001(702)895-4013
matt.bernacki@unlv.edu

P. Merlin Uesbeck
Dept. of Computer Science, University of Nevada, Las Vegas
4505 S Maryland Parkway #3003
Las Vegas, NV 89154, USA
uesbeck@unlv.nevada.edu

ABSTRACT
Learning management systems log users' behaviors, which can be used to predict achievement in a course. This paper examines the implications of data representations (e.g., dichotomous vs. count vs. principled, per learning theory) and applies forward selection algorithms to predict achievement in a biology course. Accuracy is compared across models. The paper closes with a description of an ongoing experiment that employs the prediction model, tests how multiple versions of an early alert message impact students' access of learning resources, and compares the influence of messaging approaches related to personalization and feedback.

Keywords
Learning Management Systems, Prediction Modeling, Early Warning Systems, STEM learning, Learning Theory

1. INTRODUCTION
In response to issues with student performance, retention, progression, and completion [5], universities and educational software providers are developing "early warning systems" to identify students likely to obtain poor outcomes [3]. This paper explores whether logs of students' use of course content can inform models that predict these students' performance. Further, if models can be developed that rely on only behaviors occurring in the earliest weeks of a semester [1], intervention activities can be initiated in time to help students prevent negative outcomes [2].

Undergraduate students utilize a learning management system (LMS) for multiple functions. Based on design features of LMS resources, patterns of student activity may implicate how to represent data in prediction models [4]. For instance, it is more appropriate to model use of a downloadable file as a dichotomous event that should impact learning if it occurs once (indicating that a student has obtained the file) compared to zero times (indicating the student has not). In contrast, resources designed for repeated use online, such as practice quizzes, are best captured as count data. We examine implications of different representations of LMS resource use on the accuracy of prediction models, examine whether the most accuracy model predicts performance in subsequent samples, and whether the model can provide a basis for alerting students about their potential for poor achievement.

2. METHODS
2.1 Participants
For the development of the prediction model, LMS logs capturing behavioral data were gathered for 326 students of an Anatomy and Physiology course at a large, public university in the U.S. Of those sampled, 73% were female and 36% were from underrepresented minority groups. To examine the application of the prediction model on future students, additional samples of 298 and 349 students were drawn from the subsequent Spring and following Fall semesters. All three semesters employed an identical syllabus, an analogous schedule through the observation period, and a cloned set of LMS-hosted materials.

2.2 Materials
Prediction modeling used machine data extracted from server logs of users' behavior-based activity in the LMS from the first four weeks of the course (i.e., prior to any exam). Early warning could then be generated and sent in time for learners to adjust tactics or seek help prior to their first unit exam (i.e., in Week 5). The logs were aggregated and enriched using Splunk [7], a platform for search and modeling of machine data, and tables of metadata about content items. Classification of items into resource types was handled by human research programmers. Models were built and evaluated in RapidMiner [6].

2.3 Procedure
The course that provided data was a traditional large lecture class with a companion site on the LMS, Blackboard Learn. Students could access course materials at any time from the start of the semester, and all use was optional. The frequency and timing of each resource access was recorded and coded by a unique item identifier and time stamp. To represent planful, timely, and recurring use of content items, counts of accesses were captured on a weekly basis. Total use was captured per week and for the four-week period. Behavioral data were merged with performance data. The final grade served as the outcome label. Grades were converted to a binary outcome reflecting students' success (1) or failure (0) to earn a grade of 80%, the minimum "B" score needed to earn credit for STEM majors. Data were parsed into tabular form, enriched, and pivoted into counts per week per student in Splunk. Forward Selection, Weka logistic regression algorithms employing Leave-One-Out cross validation were produced for the models, which were evaluated for accuracy (e.g., κ, recall).

2.4 Model Estimation and Application
Four versions of the data were generated. The first version included the *count* of times a student accessed each content item. The second version treated all data as *dichotomously* used or not used in a period. The third version included *both* count of logs and the dichotomous versions of the data. The final version was a *principled* model guided by learning theory and awareness of instructional design intentions of the instructor; a dichotomous

representation was used for items that could be used only once (i.e. the download of a notes document) and count representations for resources that should provide benefits when used repeatedly (e.g., accessing a quiz to repeatedly self-test).

Based on the Kappa (κ) statistic and supplemented with recall metric (i.e., critical for identifying those predicted to struggle), the most accurate model produced during the test phase was then applied to the subsequent two semesters of the same biology course. Content names and date ranges of access were aligned and all potential attributes, as both dichotomous and count, were transformed using the prediction model equation to calculate *z*-values for all students, which was then converted to probability. A probability greater than 0.5 corresponded to passing with a B or better and a probability less than 0.5 corresponded to C or worse.

3. RESULTS & DISCUSSION

Differences in prediction accuracy appear in Table 1. Representing the data as only count or dichotomous produced models with accuracy better than chance, (κ = .161 and κ = .165, respectively). The model with data as both count and dichotomous improved the accuracy to κ = .224, however the recall of students to be targeted by the early warning system (i.e., those who fail to obtain a B or Better) fell. Compared to the metrics obtained by the first three models, the model employing principled representation produced the best combined accuracy, κ = .212; recall = 84.24%. It appears that drawing inferences from LMS design features and learning theory to make data representation choices maximizes the predictive accuracy of a model. We next tested its subsequent utility for identifying students at risk of poor outcomes.

3.1 Application of Prediction Models to Subsequent Samples

Using the most accurate model (Principled, Table 1), attributes and weights were applied to the new data sets to generate predictions. Kappa decreased to .071 compared to training and testing phase (κ = .212). Recall achieved with spring data was 85.14%, on par with recall obtained with the training (84.24%). This model accurately identified more than 4 of 5 future biology students who would eventually fail to earn a B. Of those labeled, half did obtain a B or Better (precision = 51.85%, initial principled model precision was 63.01%). This level of accuracy is sufficient to warrant consideration of the model for utilization in an early warning system as it is high enough to provide accurate warnings to students at risk of a poor outcome.

4. ONGOING RESEARCH

4.1 Implementation of Early Warning Systems

A follow-up study is currently underway to examine the application of the prediction model in an early alert system and whether issuing an alert to students could change student behavior or achievement. The principled version of the data model was programmed into Splunk in order to calculate the likelihood the students (N = 430) in the current semester would obtain a B or better. An early warning message was sent from the instructor through the LMS correspondence tool. Each message included a salutation, indication of the upcoming exam, and a redirect of the student to helpful resources available on the LMS for students to use (i.e., advice from A or B-earners from prior semesters, about tactics used; modules training students to apply these tactics). The students were randomly assigned to 8 groups, which included varying combinations of the message to test the importance of personalizing the message and framing with feedback. The message was sent Monday of Week 5, four days before their exam.

4.2 Preliminary Findings

Of the 326 students that were messaged, 26.4% accessed the Advice page within 24 hours after receiving the message. In total, 37.4% of the messaged students accessed the Advice page before the exam later that week. Effects on motivation, behavior, and achievement will be analyzed when available.

Table 1. Prediction models using different versions of data and using best model on subsequent semesters

Data representation	κ	Accuracy (%)	Precision (%)	Recall (%)	True: Predicted			
					1:1	1:0	0:1	0:0
count	.16	61	61	82	48	94	34	150
dichotomous	.17	60	63	72	63	79	52	132
both	.22	63	65	73	69	73	49	135
principled	.21	63	63	84	51	91	29	155
Future Semesters								
Spring	.07	53	52	85	33	117	22	126
Fall	.15	58	57	81	56	112	34	147

Note. The baseline for test data versions (count, dichotomous, both & principled) is 56%. The baseline for the Spring use data is 51% and the baseline for Fall use data is 52%.

5. ACKNOWLEDGMENTS

This project was supported by National Science Foundation Award number DRL-1420491, university sponsorship and UNLV Information Technology.

6. REFERENCES

[1] Baker, R., Lindrum, D., Lindrum, M.J., Perkowski, D. (2015) Analyzing Early At-Risk Factors in Higher Education e-Learning Courses. Proceedings of the 8th International Conference on Educational Data Mining, 150-155

[2] Hernandez, P. R., Schultz, P., Estrada, M., Woodcock, A., & Chance, R. C. (2013). Sustaining optimal motivation: A longitudinal analysis of interventions to broaden participation of underrepresented students in STEM. Journal of educational psychology, 105(1), 89.

[3] Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. Journal of Learning Analytics, 1(1), 6-47.

[4] Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. Computers & Education, 54(2), 588-599.

[5] Nandeshwar, A., Menzies, T., & Nelson, A. (2011). Learning patterns of university student retention. Expert Systems with Applications, 38(12), 14984-14996.

[6] Rapidminer [Computer software]. (2015). Retrieved from http://www.rapidminer.com

[7] Splunk [Computer software]. (2015). Retrieved from http://www.splunk.com.