Analyzing Early At-Risk Factors in Higher Education e-Learning Courses

Ryan S. Baker[1], David Lindrum[2], Mary Jane Lindrum[2], David Perkowski[2]
[1]Teachers College Columbia University, 525 W 120th St. New York, NY 10027
[2]Soomo Learning, 9 SW Pack Square, Suite 301, Asheville, NC 28801
baker2@exchange.tc.columbia.edu, david.lindrum@soomolearning.com, maryjane.lindrum@soomolearning.com, david.perkowski@soomolearning.com

ABSTRACT

College students enrolled in online courses lack many of the supports available to students in traditional face-to-face classes on a campus such as meeting the instructor, having a set class time, discussing topics in-person during class, meeting peers and having the option to speak with them outside of class, being able to visit faculty during office hours, and so on. Instructors also lack these interactions, which typically provide meaningful indications of how students are doing individually and as a cohort. Further, online instructors typically carry a heavier teaching load, making it even more important for them to find quick, reliable, and easily understandable indicators of student progress, so that they can prioritize their interventions based on which students are most in need. In this paper, we study very early predictors of student success and failure. Our data is based on student activity, and is drawn from courses offered online by a large private university. Our data source is the Soomo Learning Environment, which hosts the course content as well as extensive formative assessment. We find that students who access the resources early, continue accessing the resources throughout the early weeks of the course, and perform well on formative activities are more likely to succeed. Through use of these indicators in early weeks, it is possible to derive actionable, understandable, and reasonably reliable predictions of student success and failure.

Keywords

At-Risk Prediction, Prediction Modeling, Predictive Analytics, Activity Analytics, Online Course, Webtexts

1. INTRODUCTION

Students enrolled in online courses lack many of the supports available to students in traditional face-to-face classes on campus [13]. Drop rates are typically higher for online courses than traditional courses (see review in [8]), and procrastination is often a major problem in online courses [10]. Part of the reason for the lower success seen in online courses comes from the fact that faculty have less direct contact with students [5, 19] and as a result have fewer indicators of how students are doing, outside of formal assessment. This makes intervention for at-risk students more difficult than in campus-based learning settings.

As a result, many universities and providers of online courseware have moved to models that can automatically identify when students are at risk. These models identify indicators of potential student failure (or lower success). A comprehensive review of work in this area can be found in [10]. In one example of the creation and study of such a model, Barber and Sharkey [4] predicted course failure using a mixture of data from student finances, student performance in previous classes, student forum posting, and assignment performance. In a second example, Whitmer [17] predicted final course grade from student LMS usage activity, including the number of times a student accessed any content, the number of times a student read or posted to the forum, and the number of times a student accessed or submitted an assignment. In a third example, Romero and colleagues [15] predicted final course grade from activity and performance on assignments, including time taken by the student; this work was followed up by additional work, where the same group studied a more extensive set of interaction variables within the Moodle platform [14]. In a fourth example, Andergassen and colleagues [1] predicted final exam score from completion of online learning activities, including when in the semester students engaged those activities, and the total span of time between a student's first and last activities in the online resource.

An area of particular importance is early prediction, as recommended by Dekker and colleagues [7]. Being able to make predictions early in the semester, using the data available from initial student participation in the course, allows for timely intervention. There have been projects that have been successful in identifying at-risk students early in the semester. For example, Ming and Ming [12] developed models that could predict student course success from the first week of course participation, based on the topics students posted on the online discussion forum. In another example, Jiang and colleagues [11] predicted MOOC course completion from grades and discussion forum social network centrality, at the conclusion of the first course week.

Models that can predict student success early in a course, from course participation data, may be more or less useful depending on the features the models are based upon. If models are based on indicators which are interpretable and meaningful to course staff, these models can then provide instructors with data on which students are at-risk along with information on why those specific students are at risk. Systems of this nature have been successfully embedded within intervention practices and had positive impacts on student outcomes. For example, the Course Signals project at Purdue University provides predictions to instructors along with suggested interventions for specific students, in the form of recommended emails to send the students [2]. In one evaluation, Course Signals was associated with better student grades and better retention [3]. Another project, the Open Academic Support Environment, was associated with better student grades [10].

The attributes of a desirable predictive model are tightly connected to the potential uses of that model. For example, highly complex "black box" indicators are hard for instructors to use in interventions, even if they might be perfectly suitable for automated interventions. Beyond this, demographic variables (such as race and financial need) can be predictive [17, 18], but are less immediately useful for instructors wishing to intervene.

In this paper, we study early predictors of student success based on student activity, with the goal of giving faculty immediately useful, easy-to-interpret data.

We analyze these predictors within the context of the Soomo Learning Environment, a system used by over 100 universities to deliver course content and extensive formative assessment to over 70,000 undergraduates a year. Specifically, in this paper we study the learning and eventual success of over four thousand students taking an online course on introductory history at a large 4-year private university.

We find that students who access the resources early, continue accessing the resources throughout the early weeks of the course, and perform well on formative activities are more likely to succeed in the course overall. Through use of these indicators in early weeks, it is possible to derive actionable, understandable, and reasonably reliable, predictions of student success, enabling faculty to identify those students most in need of intervention, and suggesting the kind of guidance each student needs.

2. DATA

We investigate these issues within the context of data from an introductory history course, offered as an online course by a large 4-year private university, using an interactive web-based learning resource from Soomo. The Soomo Learning Environment (SLE) is a web-based content management system built for hosting instructional content and formative assessment. Typically students click a link in their learning management system to open their webtext, hosted in the SLE, in a new tab. All course content, customized for the specific instructor and institution, is presented within this environment. Courses are typically built with a mix of original, permissioned, and open content, combining text, images, audio, video, hosted and linked artifacts, and tools for study. Webtexts are developed by instructional designers at Soomo Learning in conversation with faculty advisors and subject matter experts. Webtexts are then peer reviewed and finally tailored to the needs of a specific institution and/or faculty member.

Webtexts are not just digital copies of traditional paper textbooks; they are distinguished by hundreds of opportunities for students to respond to the content through the course Within Soomo's webtexts, "Study Questions" help students assess their own comprehension of what they just read or watched. "Investigations" present opportunities for application, analysis, synthesis, and evaluation, thereby supporting learners in developing richer understanding.

Final student grades in the US History course were based on performance on a range of assignments. The grade weighting was identical across sections in a specific term, but varied term-to-term as the university and Soomo Learning worked together to tune the course. The final course grade was based on a combination of a final paper and milestones to that final paper, work in the Soomo Learning Environment, and participation in class discussion boards. We obtained data on student course performance and webtext activity, for 4,002 students enrolled across 140 sections of this course, taught over six terms in 2013 and 2014. These students performed a total of 2,053,452 actions in the webtext, including opening pages and answering questions.

Student grades below 60% were considered failing grades; however, the target of our at-risk predictions was to predict whether students would fall below 73%, the minimum grade required to get a C. 990 of the 4,002 students (24.7%) obtained a grade below 73%.

3. ANALYZING INDIVIDUAL PREDICTORS

One of the major goals of predictive analytics is making predictions early in the semester, before the student has fallen behind on the course's material to an extent that is difficult to repair. It is at this stage where instructor intervention can have the greatest impact. In this paper, therefore, we focus on student performance and usage in the first 4 weeks of a 10-week term.

The Soomo webtexts include formative assessment throughout the course, starting on the first pages of the resource. This gives faculty measures of student engagement and performance from the very first week of the course. The predictors analyzed in this paper are not inherent to the Soomo Learning Environment – they could be applied to other online courses that have online readings and assignments. They rely primarily on having measures of student engagement and understanding on a regular basis, from the start of the course.

3.1 Did the student access the webtext at all?

The first feature we analyze is whether students accessed the webtext at all in the early stages of the course. This course was organized into a set of one-week units. Therefore, it might be plausible to analyze whether a student accessed the webtext during the first week of the course; by the end of the first week, the students were expected to have completed the first week's materials. However, many students procrastinate [16], and students are not penalized within this course for completing materials late, so it is possible that many students do not access course materials within this window. We analyze variants of this feature, looking at whether students have failed to access the webtext and activities within the first N days of the course. The canonical value of N is 7; other values are also examined. (We omit data from one course term for this analysis in specific, due to a logging error).

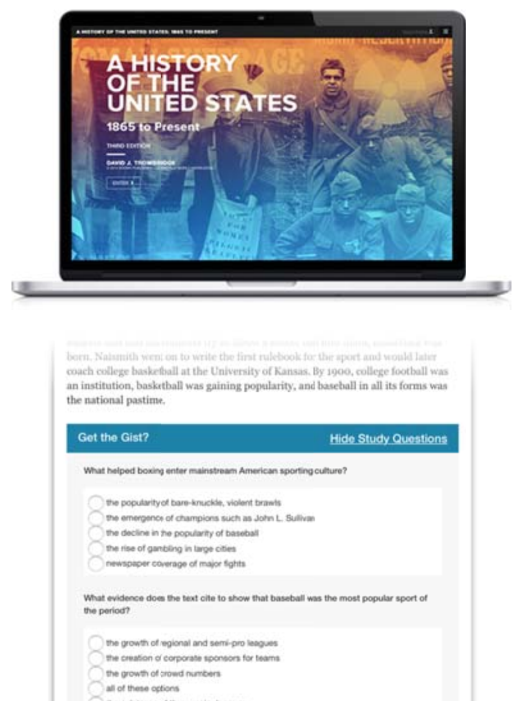

Figure 1. The introductory US History webtext (above) and embedded study questions relevant to that text (below)

As such, we predict whether a student got a course grade under 73% (a.k.a. eventually failed or got a D), from whether the student had accessed the book yet by day N. A precision-recall curve for this relationship is shown in Figure 2. A precision-recall curve [6] shows the tradeoff between precision and recall for different thresholds of a model. Precision represents the proportion of cases identified as at-risk that are genuinely at-risk; recall represents the proportion of genuinely at-risk cases that are identified as at-risk. They are computed:

$$Precision = \frac{true\ positives}{true\ positives + false\ positives}$$

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives}$$

Typically, precision-recall curves are used for different confidence thresholds between a positive and negative prediction; in this case, we display the tradeoff between precision and recall for different thresholds of how many days into a course a student can be before we become concerned that they have not accessed the webtext yet. As will be seen in the paper, studying these curves allows us to study the relative trade-off between precision and recall for different model thresholds and different feature variants. Some instructors may want models with higher recall, so that they can contact a larger proportion of at-risk students; other instructors may want more models with higher precision, to avoid contacting too many total students. While some researchers argue for optimizing a single metric, different instructors (or university administrators) may prefer different models.

As Figure 2 shows, there is a clear trade-off between precision and recall for how many days have passed at the start of the course without the student accessing the webtext. On the far left, almost all students who have not yet accessed the webtext by the 14th day of the class fail. On the far right, almost all students who eventually fail are captured by a model that looks at whether the student has not yet accessed the webtext seven days before class, but precision is only 40%. On the first day of class (day 0), precision is barely higher but recall is much lower. Seven days later (day 7), precision approaches 80% but recall is just below 20%. As such, this indicator changes its meaning considerably with each day that passes during the first 7 days of the class. On day 0, the Cohen's Kappa for this feature (representing the degree to which the model is better than chance) is 0.207. On day 7, Kappa is 0.200. On day 3, it reaches a maximum of 0.277; any value of N higher or lower than 3 has a lower Kappa.

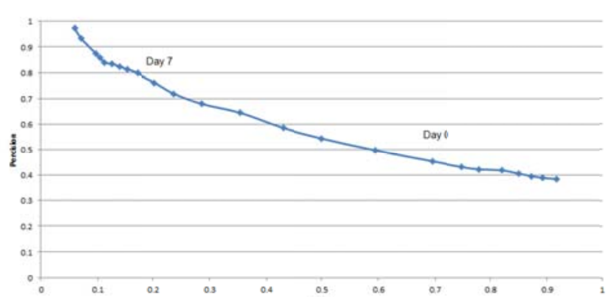

Figure 2. Precision-Recall Curve for how well a final grade below 73% is predicted by whether a student has accessed the webtext by day N.

3.2 Has the student accessed the webtext recently?

Accessing the webtext is an important first step, but it is reasonable to believe that students are most successful if they continue to access the course materials weekly. As such, the second feature we analyze is how long it has been since the student accessed the webtext. This feature has two parameters: the current day N, and the number of days D since the student last accessed the webtext.

As such, we are predicting whether a student got a course grade under 73% (a.k.a. eventually failed or got a D), from whether the student had accessed the book in the last D days, at the time of day N. For tractability, we select four possible values for D: the last 3 days, the last 5 days, the last 7 days, and the last 10 days. We also select values between 1 and 28 for N; the model does not go beyond the fourth week of this course, because after this point, it is relatively late for "early" intervention. Note that students can open the book before the first day of the course (so it is meaningful to compare between values of D, even for N=1).

A set of precision-recall curves is given for these model variants in Figure 3. As Figure 3 shows, the models start out very similar, regardless of value of D, at the beginning of the course, with precisions around 44%-46% and recalls around 65%-70%.

As the value of N goes up, recall drops and precision goes up, until the changes become unstable around the third week of the course. (At that point, however, the changes are relatively minimal). The higher the value of D, the higher the eventual precision and the lower the eventual recall, at the end of the fourth week of the course. For instance, for D = 7, the precision reaches 80.4% by day 14, though the recall is at a relatively low 16.7%. To put this another way, on day 14, a student who has not accessed the textbook in the last 7 days has a 80.4% probability of performing poorly in the course, and 16.7% of students who perform poorly in the course had not accessed the textbook in the last seven days on day 14.

This shift effect is relatively weaker for lower values of D; for instance, for D = 3, the precision goes up relatively little, reaching only 54.2% on day 4, while the recall drops rapidly, reaching 35.8% by day 7. These results, in aggregate, show that this feature manifests different behavior depending on choice of threshold.

Kappa values were relatively unstable across predictors, though the differences in Kappa were generally small, indicating that most of the differences between models reflected a precision-recall tradeoff. The best Kappa, 0.27, was obtained for D=7 and N=28. The second best kappa, 0.247, was obtained for D=7 and N=4. However, the third best kappa, 0.241, was obtained for D=3 and N=4. Kappa values were generally higher for higher values of D, but the differences were extremely small; the average Kappa for each value of D only varied by 0.03.

3.3 Is the student doing poorly on exercises in the webtext?

Another indicator that the student is struggling is if the student is performing poorly on the formative exercises in the webtext. These exercises comprise only a third of the student's eventual grade, but are an indicator that the student does not understand the content. As discussed above, there are two types of assignments within the webtext, Study Questions and Investigations.

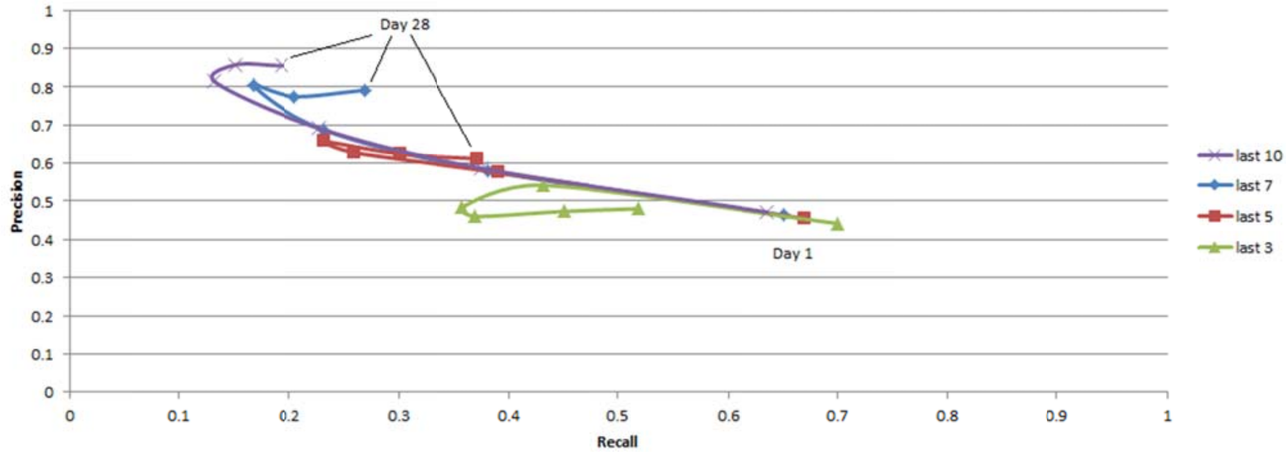

Figure 3: Precision-Recall Curve for how well a final grade below 73% is predicted by whether a student has accessed the webtext in the last D days (indicated by color), by day N.

We can look at student performance on these two types of assignments, first filtering out students who have not completed any assignments, and then looking for students who by the end of the first or second week of content (day N = 7 or 14) have an average below a cut-off S for Study Questions, and a cut-off I for Investigations. As such, we are predicting whether a student got a course grade under 73% (a.k.a. eventually failed or got a D), from whether the student averaged below S on Study Questions and I on investigate assignments, at the time of day N.

Optimizing based on Cohen's Kappa, and setting N = day 7, we find that the value of S has almost no impact (and are therefore not shown on Figure 3). For example, if the I cutoff = 70%, any value of S from 50% to 95% results in a Cohen's Kappa between 0.18 and 0.20. If the I cutoff = 85%, any value of S from 50% to 95% results in a Cohen's Kappa between 0.08 and 0.10.

By contrast, the value of I has substantial impact on model goodness. If the I cutoff = 65% (and S = I), Kappa is 0.20. If the I cutoff = 95% (and S=I), Kappa is -0.05.

The reason for this difference in predictive power between Study Questions and Investigations is likely that Study Questions can be reset. That is, when a student answers a set of Study Questions, the attempt is immediately graded. Students are given feedback and an opportunity to reset the questions and answer them again. Students are encouraged to do this in order to understand the correct answer before they move on. Investigations are more complex, and are also not resettable. In general, then, scores on Study Questions indicate effort and scores on Investigations indicate understanding.

Setting S = I, we can compute the precision-recall curve for different values of I, shown in Figure 4.

As Figure 4 shows, as the required grade to not be considered at-risk goes up, the recall goes up but the precision goes down, leading to very different models for different thresholds. It does not appear to make a big difference, however, whether we look at the first week of content, or the first two weeks of content.

To break this down, students who got below 95% on the first week of Soomo Learning Environment content had a 34.0% probability of performing poorly, and 81.8% of students who performed poorly in the course obtained below 95% on the first week of Soomo Learning Environment content. Students who got below 50% on

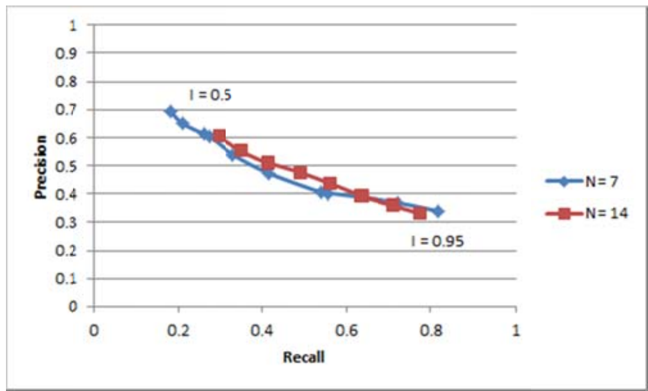

Figure 4. Precision-Recall Curve for how well a final grade below 73% is predicted by average grade on assignments (I), by day (N) 7 and 14.

the first week of Soomo Learning Environment content had a 69.5% probability of performing poorly, and 18.1% of students who performed poorly in the course obtained below 50% on the first week of Soomo Learning Environment content. As Figure 4 shows, the trade-off between precision and recall is roughly even for values of S and I between 50% and 95%.

4. INTEGRATED PREDICTIVE MODEL

Having computed these three indicators, it becomes feasible to look at the three in concert, to see how well we can do overall at predicting whether a student is at risk of obtaining a low grade.

The most straightforward way to do so would simply be to combine the single best version of the three operators described above, with an "or" function. Taking the students who obtained below 95% on the first week of Soomo Learning Environment content, the students who had not yet opened the book on day 2, and the students who had not accessed the book in the last 7 days on day 28, and combining them using an "or" function ends up with the prediction that 98.6% of students are at-risk, a model that is not very usable for intervention (the instructor intervenes for all students).

Alternatively, we can use higher-precision, lower-recall versions of these metrics. Taking the students who obtained below 50% on the first week of Soomo Learning Environment content, the students who had not yet opened the book on day 7, and the students who had not accessed the book in the last 3 days on day 7, and combining them using an "or" function ends up with the prediction that 84.7% of students are at-risk, still too many interventions.

If, by contrast, we use "and" across the three operators, trying to find students who are definitely not at-risk (e.g. students who demonstrate none of the three behaviors that are indicative of an at-risk student), the higher-precision, lower-recall version of the metrics identifies exactly four students out of 4002 as being at risk. The lower-precision, higher-recall version of the metrics identifies 14.1% of the students as being at-risk, a more workable number for intervention. However, the model achieves a precision of 25.8% and a recall of 10.2%, much worse numbers than single-feature models.

An alternate approach, which we use in this section, is to use a machine-learned model to combine the features in a more complex way. In these analyses, we conduct cross-validation as a check on over-fitting, to determine how reliable these models will be for new students in future sections of the course. Given the focus on predicting performance for future course sections, we conduct the cross-validation at the grain-size of course sections.

We input to the models the best variants of each feature (in terms of Kappa) seen in the previous sections. We also input extreme threshold variants of the features (high precision-low recall and low precision-high recall) when they achieve comparable Kappa to the best variants. In specific, we include whether the student opened the book on the first N days after the course start (0 days, 2 days, 7 days), whether the student accessed the book recently (D=7, N=28; D=7, N=4; D=3, N=4), and performance on assignments (wk. 1 only, S=I=0.65).

We applied several classification algorithms to these features, and evaluated the resultant models using Kappa, precision, recall, and A', shown in Table 1. A' is the probability that the model can distinguish whether a student is in the at-risk category or not. A model with an A' of 0.5 performs at chance, and a model with an A' of 1.0 performs perfectly [9]. A' is used rather than the theoretically equivalent AUC ROC implementation, due to bugs in existing implementations of AUC ROC.

As is often the case, there is not a single best model across all metrics. The best A' is obtained by W-KStar; but this algorithm's Kappa is much lower than other algorithms with very similar A'. Arguably, Logistic Regression, with A' only 0.015 lower than W-Kstar, but Kappa 0.111 better, should be preferred. Logistic Regression also achieves the best Recall among the algorithms, while obtaining a middling Precision. Of course, it should be remembered that Recall and Precision can always be traded-off by selecting an alternate threshold based on a Receiver-Operating Characteristic curve, or a Precision-Recall curve (as used throughout this paper), shown in Figures 5 and 6. These curves indicate that recall can be increased to 94.3%, while maintaining precision of 35.1%.

Table 1. Performance of Integrated Predictive Models.

Algorithm	Kappa	Precision	Recall	A'
W-J48	0.315	0.636	0.435	0.655
W-JRip	0.265	0.570	0.468	0.578
Naïve Bayes	0.231	0.532	0.483	0.666
W-KStar	0.233	0.670	0.288	0.677
Step Regression	0.305	0.697	0.353	0.658
Logistic Regression	0.344	0.568	0.595	0.662

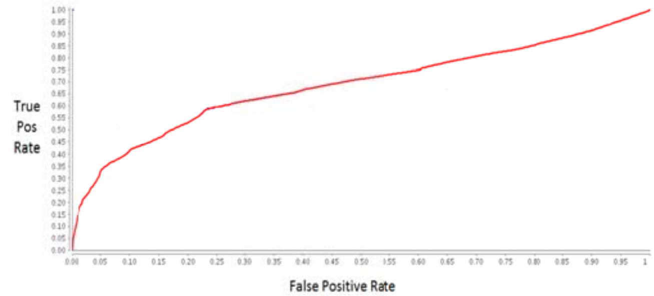

Figure 5. Receiver-Operating Characteristic Curve for (Cross-Validated) Logistic Regression Version of Integrated Predictive Model.

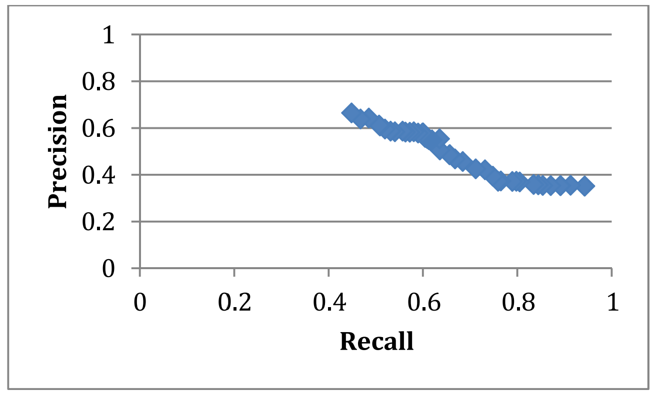

Figure 6. Precision-Recall Curve for (Cross-Validated) Logistic Regression Version of Integrated Predictive Model.

5. DISCUSSION AND CONCLUSIONS

In this paper, we have investigated the degree to which student participation in webtext activities within the Soomo Learning Environment, early in the semester, are predictive of eventual student success in a course. We find that it is indeed possible to achieve a reasonable degree of predictive power, and to identify a substantial proportion of the at-risk students, with reasonable precision. Some of these measures have predictive value from the first day of the course, allowing very early intervention.

In aggregate, we find that a combination of these measures leads to A' values in the 0.65-0.7 range, sufficient for intervention, though not quite up to the level of medical diagnostics. The logistic regression version of the combined model can identify 59.5% of students who will perform poorly, achieving precision of 56.8%, 34.4% better than chance. Of course, with any of the approaches used here, confidence thresholds for intervention can be adjusted, leading to more or fewer interventions. If high recall is the goal – attempting to provide intervention to most at-risk students even if some interventions are mis-applied – then the threshold of the logistic regression model can be adjusted, resulting in a model that can identify 94.3% of the students who will perform poorly, but where only 35.1% of the students it identifies performs poorly. This model does better than a single-feature model; even the high recall model from section 3-3 (performance under 95% on early assignments within the webtext) obtained a recall of 81.8% -- lower than the logistic regression model – while achieving comparable precision (34.0%).

However, if the goal is to provide high-cost interventions to the students who are very likely to perform poorly, the logistic regression model is not an optimal choice. The logistic regression model cannot achieve very high precision, even through adjusting thresholds, as shown in Figure 6. However, an alternate approach can be adopted, through using a different predictor algorithm, step regression. This algorithm obtains more precise prediction than logistic regression, with precision of 69.7% and recall of 35.3% for standard thresholds.

Importantly, these measures are based upon interpretable features. They are based upon features that instructors identified as meaningful and having the potential for intervention. The combination of individual-feature models and a comprehensive model enables us to identify which students are at risk, and then to provide instructors with information about which students are at risk, and why. We can specifically identify that a student is at risk because he/she has failed to access the resources, or because he/she has failed to complete the assignments on time, or because he/she has scored poorly on the assignments. With this information, automatically distilled and placed in a user interface within the Soomo platform, faculty will have a means of finding students who most need support and a basis for encouraging them to access the text, do the assigned work, and take the time to do it well.

The first area of future work planned is to enhance the analytics already offered to instructors by Soomo, based on the findings presented here. The success of these interventions, both in terms of improved student grades and improved student retention, will be evaluated in an experiment or quasi-experiment (the final study design will depend upon negotiation with the university which partnered on the analyses discussed in this paper).

However, beyond testing interventions based on the model presented here, there is considerable future work to extend, improve, and study the generalizability of these models. For example, it will be valuable to study what characterizes the students for whom this model functions less effectively. Can additional features, like how much time students spend on assignments, improve overall prediction? And how well will the features identified here apply for different courses, and for different universities, an issue explored by Jayaprakash et al. [10], among others. By answering these questions, we can improve the models, verify their broad applicability, and move to using the models within intervention strategies that can achieve broad positive impact on learners.

6. ACKNOWLEDGMENTS

This research was made possible by the active cooperation of our partner university.

7. REFERENCES

[1] Andergassen, M., Modrtischer, F., Neumann, G. (2014) Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics, 1* (1), 48-74.

[2] Arnold, K. (2010) Signals: Applying Academic Analytics. *Educause Quarterly.* March 2010.

[3] Arnold, K., Pistilli, M. (2012) Course Signals at Purdue: Using Learning Analytics to Increase Student Success. *Proc. of the 2nd International Conference on Learning Analytics.*

[4] Barber, R., Sharkey, M. (2012) Course Correction: Using Analytics to Predict Course Success. *Proceedings of the 2nd International Conference on Learning Analytics*, 259-262.

[5] Beard, L.A., Harper, C. (2002) Student Perceptions of Online versus on Campus Instruction. *Education, 122* (4).

[6] Davis, J., Goadrich, M. (2006) The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine learning.*

[7] Dekker, G., Pechenizkiy, M., Vleeshouwers, J.M. (2009) Predicting Students Drop Out: A Case Study. *Proc. of the 2nd Int'l. Conference on Educational Data Mining,* 41-50.

[8] Diaz, D.P. (2002) Online Drop Rates Revisited. *The Technology Source,* May/June 2002.

[9] Hanley, J.A. and McNeil, B.J. (1982) The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve. *Radiology, 143*, 29-36.

[10] Jayaprakash, S.M., Moody, E.W., Lauria, E.J.M., Regan, J.R., Baron, J.D. Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics, 1* (1), 6-47.

[11] Jiang, S., Williams, A.E., Schenke, K., Warschauer, M., O'Dowd, D. (2014) Predicting MOOC Performance with Week 1 Beahvior. *Proceedings of the 7th International Conference on Educational Data Mining,* 273-275.

[12] Ming, N.C., Ming, V.L. (2012) Automated Predictive Assessment from Unstructured Student Writing. *Proceedings of the 1st international Conference on Data Analytics.*

[13] Muilenburg, L.Y., Berge, .L. (2005) Student Barriers to Online Learning: a factor analytic study. *Distance Education, 26* (1), 29-48.

[14] Romero, C., Olmo, J.L., Ventura, S. (2013) A meta-learning approach for recommending a subset of white-box classification algorithms for Moodle datasets. *Proc. of the 6th Int'l. Conference on Educational Data Mining,* 268-271.

[15] Romero, C., Ventura, S., Garcia, E. (2007) Data mining in course management systems: Moodle case study and tutorial. *Computers and Education, 51* (1), 368-384.

[16] Thille, C., Schneider, E., Kizilcec, R.F., Piech, C., Halawa, S.A., Greene, D.K. (2014) The Future of Data-Enriched Assessment. *Research and Practice in Assessment, 9* (4), 5-16.

[17] Whitmer, J. (2012) *Logging on to improve achievement: Evaluating the relationship between use of the learning management system, student characteristics, and academic achievement in a hybrid large enrollment undergraduate course*. Unpublished Doctoral Dissertation, UC Davis.

[18] Wolff, A., Zdrahal, Z., Nikolov, A., & Pantucek, M. (2013). Improving retention: Predicting at-risk students by analysing clicking behaviour in a virtual learning environment. *Proceedings of the Third International Conference on Learning Analytics and Knowledge,* 145–149.

[19] Zhang, J., & Walls, R. (2006). Instructors' self-perceived pedagogical principle implementation in the online environment. *The Quarterly Review of Distance Education, 7*(4), 413–426.