Improving Student Risk Predictions: Assessing the Impact of Learning Data Sources

Authors:
Diego Forteza, Blackboard | John Whitmer, Ed.D., Blackboard
John Fritz, Ph.D., University of Maryland, Baltimore County (UMBC) | Daniel Green, VitalSource

Key Findings

Early activity is a strong predictor of passing a class

Activity patterns differ between courses

Learning activity is a powerful predictor of achievement

Combining data improves predictions

Students who actively use multiple online resources are more likely to pass a class

Using IMS Caliper Analytics® with Blackboard Learn & VitalSource at the University of Maryland Baltimore County

Motivation

Educators are increasingly using multiple systems to provide technology-enhanced learning experiences for students (e.g., LMS, eTextbooks, clickers, etc.). Each of these systems creates a data stream that describes user activity. When student engagement data from these systems is combined with conventional student demographic data from a Student Information System, it becomes possible to investigate complex trends in the relationship between student background and learning behaviors that can impact their outcomes. Better understanding these relationships enables institutions to take actions to help students, whether those are by changing student behaviors or revising the underlying learning experiences.

We frequently hear from institutions that in order to “do” learning analytics, we need to have data from all of these systems. If not, it is argued, we will only have partial understandings of student activity that could be misleading. For example, if a student appeared to not log into Blackboard Learn, they might be choosing to use their study time working with publisher resources, which could be a better use of their time. Intuitively, this idea is reasonable—but to our knowledge, the added value of data from multiple system sources hasn’t been demonstrated empirically. In this era of “big data,” there’s often an assumption that more is better, but what is the actual difference to our understanding and prediction of student outcomes? Is it worth the extra effort and resources required to integrate and analyze this data?

Blackboard In partnership with VitalSource and UMBC

For institutions working within resource constraints (and who isn't?), this is an important question. Perhaps resources could be better spent on analyzing existing data than on further building out a data warehouse or integration. Or better yet, could they be better spent on intervention services to help students with the issues that have already been identified?

We suggest that rather than focusing exclusively on building big data, it'd be wiser to identify the least amount of data that could offer us the greatest value. Of course, there's a bit of a Catch-22: you don't know what data is valuable until an analysis has been performed with that data. Getting some clarity about that additional value provided by multiple data sources was the underlying motivation for us to work on this project. If they provide dramatic additional value, this research would help us to prioritize additional data sources into our descriptive and predictive analytics solutions. If not, we should stick with a smaller set of data sources.

Methodology

Both Blackboard and VitalSource record student interactions using the Caliper Analytics® standard from IMS Global Learning Consortium. This standard allows the integration of learning events from multiple systems through a defined "vocabulary" that describes these events and a protocol for transmitting these events. This research project sought to understand (a) the impact of using integrated, VitalSource-powered materials on student learning in a Blackboard Learn course and (b) the change in understanding of student behavior provided by combining Caliper data sources, from VitalSource and Blackboard Learn, specifically our ability to create early predictions of student risk of failing a course.

This joint project involved teams from University of Maryland Baltimore County (UMBC), Blackboard, and VitalSource.

Data Sources & Course Context

This project used the following data sources:

Data	Definition	Source
Blackboard Learn Caliper Data	Learn Caliper events for courses under investigation. Primarily course navigation events, the most diverse stream.	Caliper telemetry event stream from UMBC's Blackboard Learn SaaS environment.
VitalSource Caliper Data	VitalSource Caliper events for courses under investigation.	Caliper telemetry event stream from VitalSource.
Student Demographic and Prior Academic Achievement	Student characteristics from UMBC Analytics for Learn instance.	UMBC Student Information System.

Activity data was studied until week four of the course to provide results that could be acted upon with enough time for students to change their behavior and outcomes in the course.

Data Privacy

This research project was conducted under research data sharing agreements between UMBC, VitalSource, and Blackboard. Beyond the organizational agreements for data sharing, an Institutional Review Board review was conducted and approved through the appropriate committee at UMBC to confirm that this project was an ethically appropriate use of student data.

Learning Contexts

Initially seven course sections from the Fall 2017 semester were identified for potential investigation. These courses were selected as the faculty had been using VitalSource for multiple terms and we believed that there was a relatively robust adoption of these course materials in addition to use of the Blackboard Learn gradebook. Although this purposeful selection limits the generalizability of results, for an initial study we decided to start with a "best case" scenario in terms of the potential data context for predictive accuracy.

As illustrated in Tables 1 and 2, the courses investigated here covered a diverse range of subjects and enrollments. The subjects ranged from STEM topics such as Math and Physics to Social Sciences and Languages. In addition to diverse subjects, the course pass rates were also highly variable, with a range from 70%–90%. Further, the distribution of students into quartiles of use was different between courses. Analysis was conducted on the entire population of all courses, and some courses were further identified for in-depth analysis based on distributions of activity and numbers of students that indicated a potential for findings with greater relationship.

While this approach limits the generalizability of the results, we thought that it was appropriate as an initial exploratory study of a "best case" scenario. As a result, we chose the Math and Physics courses for some of the further investigations. However, the statistical tests and calculations were performed on the entire population of courses.

Table 1: Course Enrollments & Pass Rates

Course	Enrollment	Students Passing	Drops	Pass Rate
MATH155 "Applied Calculus," Section 1	173	121	15	0.7
MATH155 "Applied Calculus," Section 2	164	123	6	0.75
PHYS122 "Introductory Physics II,"	284	235	7	0.83
PSYC320 "Psychological Assessment,"	39	33	6	0.85
SPAN201 "Intermediate Spanish I," Section 1	36	31	2	0.86
SPAN202 "Intermediate Spanish II," Section 2	26	21	2	0.81
STAT351 "Applied Statistics for Business Economics," sec 2504	116	104	8	0.9

Table 2: Activity Quartiles by Course Subjects

	Q1 Actions	Q2 Actions	Q3 Actions	Q4 Actions
MATH	23	88	114	112
PHYS	164	83	26	11
PSYC	8	17	10	4
SPAN	4	7	16	35
STAT	9	15	44	48

Results

Finding 1: Early Activity is a strong predictor of Passing a Class

Activity in Blackboard Learn and VitalSource up to week four was combined for each student and separated into quartiles. In Table 3, a cross-tabulation of quartiles and grades is presented. The cell with the largest value is students in the fourth quartile of actions who earned an A; the next-highest value is students in the third quartile with an A. Students with F grades are likewise heavily skewed toward the first quartile; this is the most frequent grade for students in the first quartile.

The results for students in the middle grade levels were not as direct and the overall relationship is not simple or linear. However, we performed a statistical test that indicated this relationship between the two variables is systematic with a very low probability of random chance (approximately 1 in 100 million).

Table 3: Activity Quartiles and Course Grade

	Q1 Actions	Q2 Actions	Q3 Actions	Q4 Actions
A	41	66	81	90
B	61	58	60	60
C	30	37	40	44
D	11	13	10	9
F	65	36	19	7

Table 4: Pearson's Chi-Squared Test of Grades and Action Quartiles

Test statistic	df	P value
83.53	12	8.705e-13 * * *

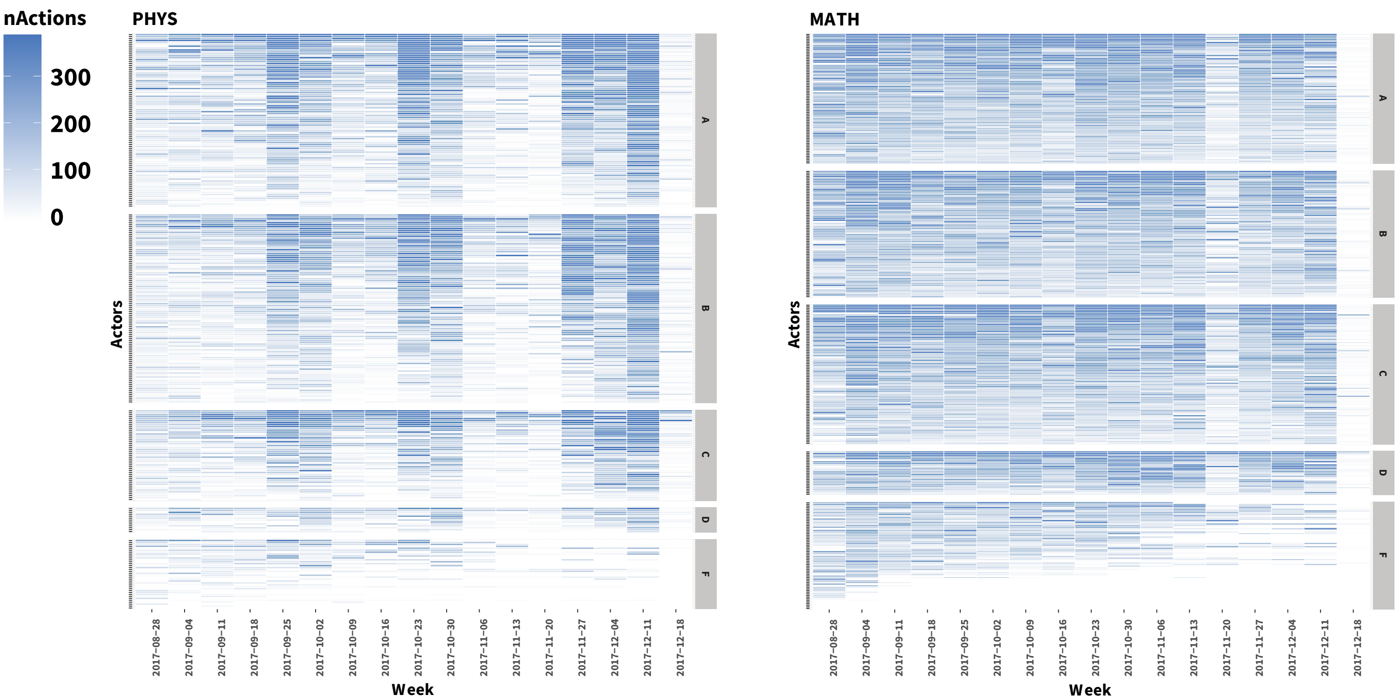
Finding 2: Activity Patterns Differ from Course to Course

In our analysis of the data, we began with overall findings, but also examined course-specific results. In conducting this analysis, we discovered dramatic differences between courses. Chart 1 illustrates the intensity of student activity over the entire duration of the court using a plot that David Wiley termed a "Waterfall" in his **2011 presentation** at the Educause Learning Initiative (ELI) annual meeting.

In this illustration, each row is the activity of a single student, and each column is the week of the course. Higher degrees of activity are illustrated by darker shaded cells. The students are grouped by final grade, and in descending order of activity. The Physics and Math courses were selected as illustrations of the kind of variation that occurs between courses and consistent patterns. In the Math course, the overall amount of activity for all students is greater and is consistent throughout the term (with the exception of Thanksgiving week!). In the Physics course there are three periods of intense activity with lighter activity between them. If you are familiar with teaching strategies, you may have already guessed the cause—the Math course has frequent assessments, while the Physics course has three high-stakes exams. These assessment strategies are clearly related to student activity within the course.

There additional consistent patterns as well. For example, activity starts slowly in each course, and the low grades are clearly associated with a decrease in activity which intensifies later in the course. Students with F grades appear to have given up less than half-way through the course. These patterns have interesting implications for intervention strategies and predictive modeling applications.

Chart 1: Student Activity Waterfalls for Physics and Math Courses



Finding 3: Learning Data is a More Powerful Predictor of Student Achievement than Demography or Educational Background

After conducting exploratory analysis of the demographic data and educational background of students, variables (from Blackboard Learn, VitalSource and the Student Information System) that had a systematic relationship to a student's likelihood to pass a course were included in a logistic regression equation. This regression predicted a student's likelihood to pass a course based on the combination of all values included in the prediction.

This regression equation resulted in not only a statistically significant prediction of a student passing a course but demonstrated interesting differences between the relative importance of the factors in that prediction. Table 5 shows the odds ratios (and confidence intervals) of the variables that were significant in the regression equation. Odds ratios indicate the increased likelihood of a student passing a course with a "one unit" change in the predictor variable. Many of these variables are dichotomous and indicate a change in that category.

> **The most powerful predictor of student achievement was Blackboard Learn use in the highest quartile. This resulted in a nearly 8-fold increase in the likelihood that student would pass the class.**

The results demonstrate the power of data provided through IMS Caliper. For example, the most powerful predictor of student achievement was Bb Learn use in the highest quartile; that resulted in a nearly 8-fold increase in the likelihood that student would pass the class. The next-greatest predictor was a student using VitalSource in a Hybrid course. This resulted in greater than 4-times the likehood that student would pass the class. A one-unit change in student college GPA, by contrast, resulted in a slightly lower than 4-times likelihood that a student would pass the course. Factors that have been demonstrated to have a large impact at the individual level such as being a member of a privileged ethnic group (Asian/Anglo) have among the lowest impact of any variable.

Table 5: Interpretation of the Logistic Regression Variable Estimates

	Odds Ratio	2.5%	97.5%
(Intercept)	0.009855	0.002946	0.03081
Bb Learn Fourth Quartile	7.79	3.92	16.30
Use VitalSource (Hybrid Course)	4.72	1.79	15.39
Bb Learn Third Quartile	4.48	2.47	8.34
College GPA	4.26	3.11	5.96
Bb Learn Second Quartile	2.79	1.61	4.95
Use VitalSource (Web Enhanced Course)	2.33	1.46	3.78
Asian or Angle Race/Ethnicity	1.88	1.11	3.17
Bachelor of Science Major	1.75	1.10	2.78
African-American Race/Ethnicity	1.73	0.92	3.29

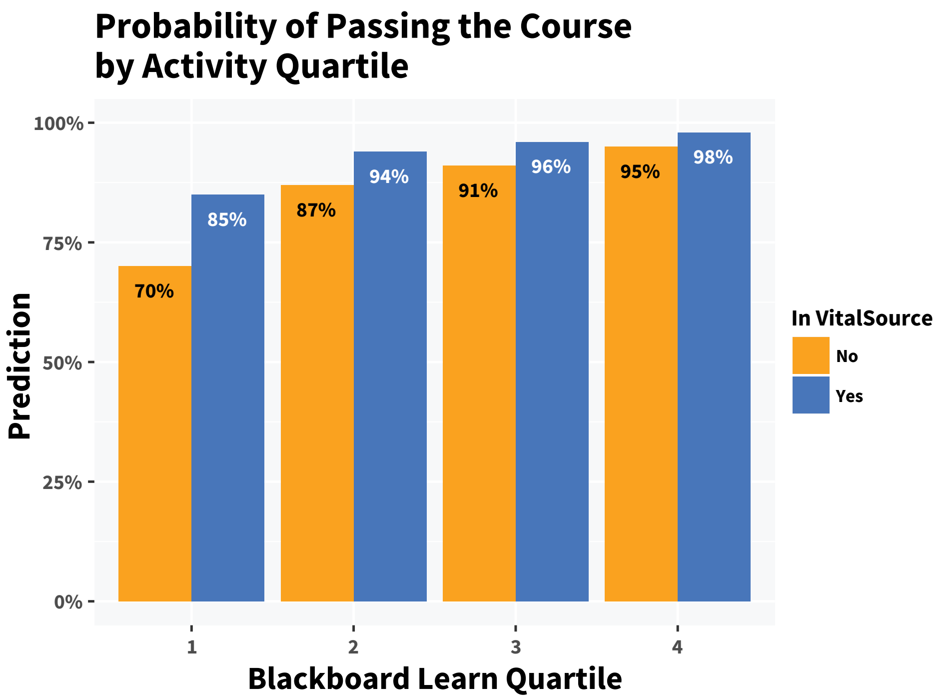
Finding 4: Combining Learning Data from Multiple Sources Improves Prediction Accuracy

Through analysis of Blackboard Learn and VitalSource data, we discovered that students who used VitalSource were more likely to pass a course than those who did not. Unlike the relationship with quartiles of Learn use, the Caliper version of VitalSource data was a binary indicator: Students who accessed the materials did better than students who did not. To isolate the results from Blackboard Learn and VitalSource, we calculated the results from the regression equations for a sample 'high-achieving' student profile. In this case, that was an Asian/Anglo student with a 2.96 GPA, who did not transfer and was enrolled in a Bachelor of Science program.

In Chart 2, the increased likelihood of passing a course by Blackboard Learn quartile of use and VitalSource use is illustrated. At the lowest level of online activity in both platforms, this student had a 70% likelihood of passing a course, which increased to 98% at the highest level of activity. At each quartile of use, the additional activity in VitalSource improves the likelihood that this student will succeed.

A noteworthy feature is that the effect of VitalSource is greatest at the lowest level of activity and it decreases at the higher levels of Learn use. In our research team discussions, we discovered that this effect is lower than VitalSource has identified using its native telemetry that includes weighting for various reading activities. For example, using raw Caliper events, page views lasting 10 seconds were weighted the same as those lasting 5 minutes. Furthermore, activities such as notes or highlights were weighted equally with other learning events such as page views or searches. We identified this as an area for follow-on research to improve the fidelity of Caliper events with useful native telemetry.

Chart 2: Prediction of Course Pass by Blackboard Learn Quartile & VitalSource Use



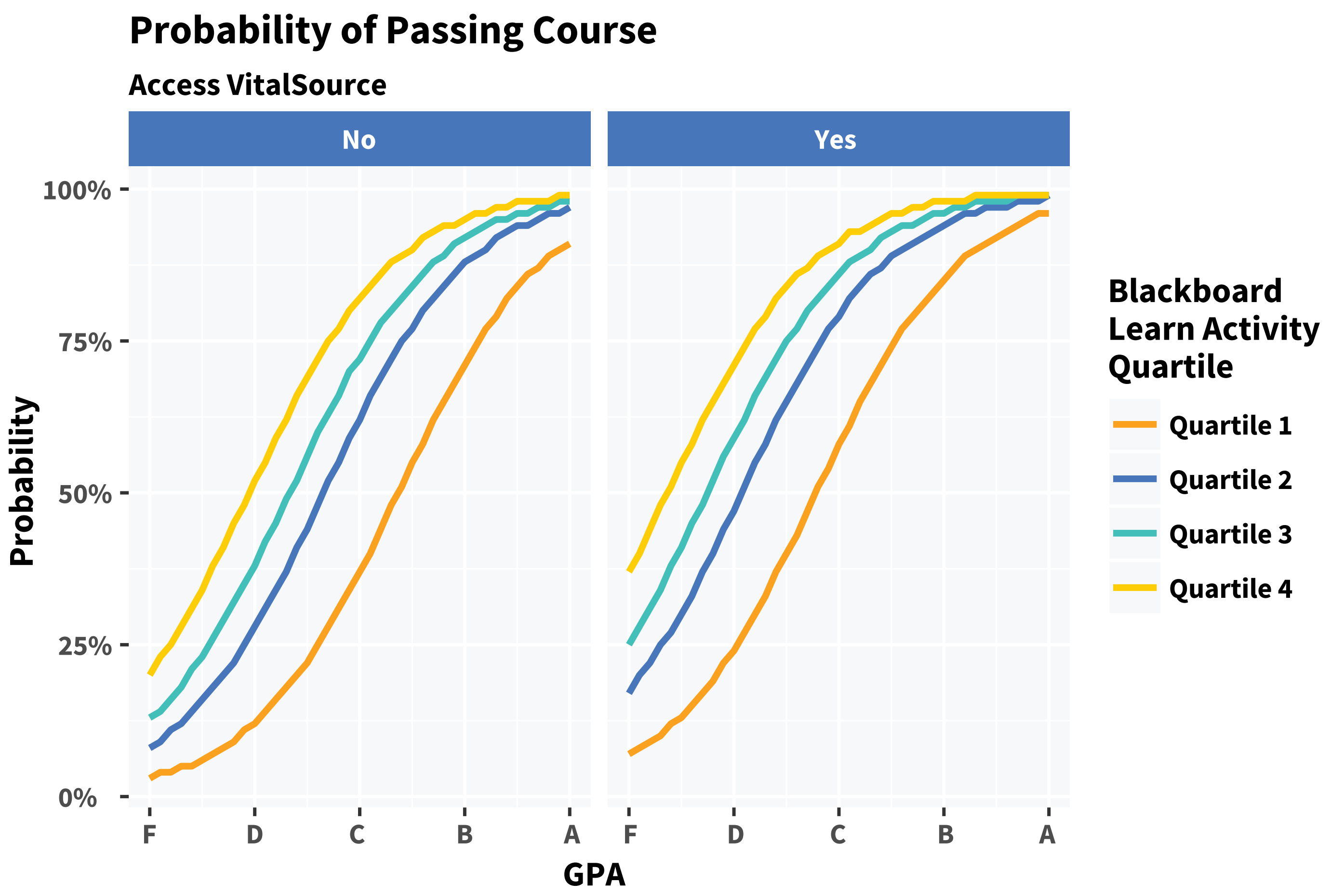
At every quartile of Blackboard Learn activity, additional activity in VitalSource improved the likelihood that a student would succeed.

Finding 5: Students who actively use multiple online resources are more likely to pass a class.

In addition to the probability held constant for a certain student, we wanted to find results that would apply to students with different incoming characteristics. In Chart 3, a more complex rendering is shown of the relationship between use of Blackboard Learn and VitalSource, separated by incoming GPA. Students' grades in a course are strongly related to their incoming GPA (one could argue that this is true by definition); however, there is a striking difference between their likely grade based on their levels of use. For example, an incoming student with a C (GPA 2.0) in the lowest quartile of use without VitalSource access has approximately a 37% chance of passing the course; where that same student has a greater than 90% likelihood of passing the course if they are in the top Quartile of Blackboard Learn use and access VitalSource.

The gap between lines (quartiles of use) shows that this trend is true at incoming GPA levels, although the strength of the effect is strongest at lower grade levels. This is a promising result, as the students most in need of improvement are those in the lower grade levels. We should note that we did examine the number of observations in each category, and there was a sufficient sample to make a claim that this was a substantial relationship and not due to chance effect.

Chart 3: Probability of Passing a Course by Incoming GPA, Blackboard Learn Activity Quartile, and VitalSource Access



Implications and Next Steps

In this study, we found strong relationships between the use of Blackboard Learn and VitalSource that crossed the boundaries of course subjects, technology adoption, and course format. We also found that IMS Caliper provided useful data that could be used to better understand student activity, the strength of the relationship between student activity using these resources and their course grades, and the ways that these varied by their incoming course characteristics. The study validated the fundamental assumption upon which this data standard, and efforts by campuses and vendors to integrate multiple learning data sources, are based—namely, that we gain improved insights into predictions of student success in passing courses.

The study found strong relationships between the use of Blackboard Learn and VitalSource that crossed the boundaries of course subjects, technology adoption, and course format.

We also found in this study that while quartiles of use were strong predictors of student grade in Blackboard Learn data, for VitalSource data the relationship was more simple: Students who accessed at all had better outcomes than those who did not, but there was not a difference by levels of activity given equal weighting of events. These results may reflect a limitation in the current data profiles or could be a deeper trend in the data.

It is important to recognize that this is not a directly causal relationship—students need to do more than simply access online materials to improve their achievement, and this relationship is likely complex. For example, it could be that students with lower GPAs who access course materials very frequently are motivated and excited about a course, whereas a student with a lower GPA at a lower level of activity may lack motivation and confidence in the materials.

Despite the complexity of these interactions and relationships, we can use this information to better understand our students, motivate them to higher degrees of achievement, and create more accurate interventions to help them succeed. The addition of more data sources appears worth the effort, at least in the sample of courses investigated for this study.

Our next steps for this study are largely related to improving the data sources:

1. Replace data with native application telemetry and compare results to those achieved through Caliper.
2. Increase courses to larger sample and compare results to this initial sample.
3. Identify additional applications that may provide deeper insights and integrate into predictive model.