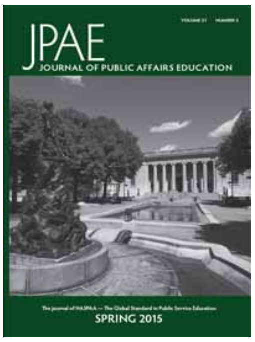


Journal of Public Affairs Education

ISSN: 1523-6803 (Print) 2328-9643 (Online) Journal homepage: www.tandfonline.com/journals/upae20

Using Learning Analytics to Predict At-Risk Students in Online Graduate Public Affairs and Administration Education

Jay Bainbridge, James Melitski, Anne Zahradnik, Eitel J. M. Lauría, Sandeep Jayaprakash & Josh Baron

Using Learning Analytics to Predict At-Risk Students in Online Graduate Public Affairs and Administration Education

Jay Bainbridge, James Melitski, Anne Zahradnik, Eitel J. M. Lauría, Sandeep Jayaprakash, and Josh Baron
Marist College

ABSTRACT
In this global information age, schools that teach public affairs and administration must meet the needs of students. Increasingly, this means providing students information in online classrooms to help them reach their highest potential. The acts of teaching and learning online generate data, but to date, that information has remained largely untapped for assessing student performance.

Using data generated by students in an online Master of Public Administration program, drawn from the Marist College Open Academic Analytics Initiative,[1] we identify and analyze characteristics and behaviors that best provide early indication of a student being academically at risk, paying particular attention to the usage of online tools. We find that fairly simple learning analytics models achieve high levels of sensitivity (over 80% of at-risk students identified) with relatively low false positive rates (13.5%). Results will be used to test interventions for improving student performance in real time.

KEYWORDS
Learning analytics, master of public administration, graduate education, online learning, early alerts

Effectively educating leaders in the public sector is a challenge for both higher education and public service organizations. In a globally interconnected world, leaders in the public sector are given many educational alternatives, increasingly including online training. This article examines critical issues for consideration as educators and practitioners develop and improve online graduate public affairs and administration education.

Using an extract from the Marist College Open Academic Analytics Initiative (OAAI) data systems (Lauría, Baron, Devireddy, Sundararaju, & Jayaprakash, 2012), we constructed a data set of 1,073 students in the Marist College Master of Public Administration (MPA) online program. We analyzed which demographic and educational factors and behaviors provide the most accurate early indications of a student being at risk of poor academic performance. We are particularly interested in the influence of online tool use, such as the number of times students log into the course, the extent to which they access course lessons, and their rate of participation in online discussion forums relative to their peers. We have found these factors are predictive of how well a student will perform in a course, along with more traditional determinants such as overall grade point average, performance on early assignments, class size, and age category.

This work builds on other research in the field of learning analytics, including Course Signals at Purdue University (Arnold, 2010) and Marist College's recent open source evaluation with four other community and undergraduate colleges across the country (Jayaprakash, Moody, Lauría, Regan, & Baron, 2014; Lauría, Moody, Jayaprakash, Jonnalagadda, & Baron, 2013). To the best of our knowledge, this study is among the first efforts to apply these analytic tools to graduate students in general, and to public administration students specifically. For our approach, we tested several common models used in data-mining analytics and found similar results. This article reports on a logistic regression model.

We split our data set into two parts, and used one part to "train" the models and the other part to test them. Because, for analytical purposes, at-risk students are a relatively rare event in our data set, in our training models we oversample the at-risk group and undersample the remainder of the group to produce a more balanced data set. Final results are tested based on models with a group that is not sampled or balanced.

We use results to assess the ability of each model in terms of its sensitivity (percent of at-risk cases identified) and specificity (percent of non-at-risk cases, or false positives, identified). We find that fairly simple models using logistic regression achieve high levels of sensitivity (over 80% of at-risk students identified) with relatively low false positive rates (13.5%). Results of these analyses will be used to design and test interventions for improving student performance.

The OAAI project, partially funded by the Bill & Melinda Gates Foundation, is an initiative to collect institutional data on students, both undergraduate and graduate, as a means to develop and deploy an open source early alert system; release predictive models under an open license; study the portability of predictive models from one academic context to another; and research the impact of different interventions aimed at improving student performance.[2]

LITERATURE REVIEW

In recent years, online education has inspired wide-ranging debate among higher education professionals. This literature review examines the larger trends in higher education involving online and distance education and examines critical issues facing online graduate programs in public affairs and administration education. The review discusses the relationship between online programs, assessing student learning outcomes, and the issues surrounding student engagement and persistence in online learning. Finally, learning analytics as a field is discussed in the context of online learning.

Trends in Higher Education and Public Affairs and Administration Education

Growth in online programs has increased steadily since the mid 1990s. In 2006, 89% of public higher education institutions reported offering distance education (Allen & Seaman, 2006). In 2011, Allen and Seaman reported 6.1 million students took at least one online course during the Fall 2010 semester. Leaders at more than 84% of institutions surveyed believed that learning outcomes in online education were the same or better than in traditional face-to-face classrooms. Perceptions of student satisfaction by leaders in higher education were also high, as more than 77% of respondents reported that perceived student satisfaction was either the same or higher in online courses as compared to face-to-face classes (Allen & Seaman, 2011).

Online MPA programs are no longer considered a novelty or a fad. Ginn and Hammond (2012) identified 57 schools (11 accredited by the Network of Schools of Public Policy, Affairs, and Administration [NASPAA]) that offered complete MPA degree programs or certificates entirely online and another 41 schools that offered some courses, but not full degrees, online. Allen and Seaman (2011) warn that

"The continued unbridled growth in online enrollments cannot continue forever—at some point higher education institutions will reach a saturation point" (p. 2), but they report that online enrollments have continued to grow at rates far in excess of the total higher education student population, with the most recent data demonstrating continued substantial growth.

These trends in higher education are reflected in public affairs and administration scholarship, where public administration researchers have studied online and distance programs, particularly at the graduate level. A 1997 survey of NASPAA members found that early supporters of online programs appeared to be external and top-level internal stakeholders who were motivated by cost savings, increased enrollment, increased market share, and enhanced ability to generate revenue. At the same time, among the early supporters of online education in public affairs and administration education, the stakeholders closest to the program were more likely to value quality of the program and value added in the learning process (Rahm & Reed, 1997). Interestingly, Stowers (1995) consistently found higher levels of participation among women, although the finding was not statistically significant. Further, several researchers assert that the communication mechanisms used in online education actually engage students in ways that are superior to face-to-face communication (Bonk & Reynolds, 1997; Brower & Klay, 2000; Feenberg, 1999).

Austin (2009) suggests that online education allows MPA programs to reach students who might not otherwise pursue graduate education, but he warns of limits to online education. Specifically, he refers to the lack of authentic relationships among participants in online courses and the ability in face-to-face classrooms to effectively communicate verbally and through the nuanced nonverbal cues students receive from instructors. Beyond concerns about the effectiveness of online education, many faculty in higher education are still resistant to the delivery platform. Allen and Seaman (2011) report that "less than one third of chief academic officers believe that their faculty accept the value and legitimacy of online education. This percent has changed little over the last eight years" (p. 5). Part of the resistance to online education by faculty may be due to concerns about workload. Several studies have shown that both student and faculty workloads are higher in online courses, as are student dropout rates (Barth, 2004; Ebdon, 1999; Leavitt & Richman, 1997; Mingus, 1999; O'Leary & Stowers, 1999; Rahm, Reed, & Rydl, 1999). Palloff and Pratt (2007) suggest that the time needed to develop and deliver an online course is two to three times greater than in a face-to-face classroom. In addition, the academic integrity of online programs and concerns about student honesty are also areas of interest in online programs (H. E. Campbell, 2006).

Growth of Online Programs and Massive Open Online Courses

Despite the concerns about online learning, the number of online programs has grown considerably. The diffusion of innovation theory suggests that any true innovation initially experiences exponential growth, followed by a leveling off (Rogers, 2010). Rahm and Reed (1997) predicted that online education in graduate programs of public affairs and administration would experience such initial growth. More than a decade later, research suggests that growth in online programs is increasing but at a slower rate than in the previous decade, and some programs, particularly in social sciences, psychology, and business are no longer reporting growth, but rather steady enrollments (Allen & Seaman, 2011).

While growth in online programs may be leveling off, enrollment in massive open online courses (MOOCs) appears to be skyrocketing. As the name implies, one of the key differences between MOOCs and other distance and online programs is class size. Stanford University reported in 2012 that enrollment in its three MOOCs had about 100,000 students each (Perez-Pena, 2012). Led by two private organizations, Coursera and Udacity, MOOCs have flourished. In Coursera's first year (2013), it offered about 325 courses and Udacity offered 26 courses. Udacity's largest course enrolled nearly 300,000 students (Waldrop, 2012). The Chronicle of Higher Education surveyed 103 professors who had taught MOOCs and found that the average class size was 33,000 students with an average of only 7.5% students passing their courses (Kolowich, 2013).

Variables for Assessment of Online Student Learning Outcomes

The discussion of online courses and how to effectively adapt assessment to the online format is still in developmental stages, particularly as these methodologies relate to student learning outcomes (Bocchi, Eastman, & Swift, 2004). Vonderwell and Boboc (2013) acknowledge, "Evaluating student learning takes on a new meaning in an online classroom environment where students and instructors do not share physical proximity" (p. 22). They advocate for the development of assessment methods that are "appropriate to online learning and understand the potential of technology tools for monitoring student learning and their own teaching" (p. 22). Vonderwell, Liang, and Alderman (2007) stress that online instructors should understand both the assessment process and particular factors that influence assessment of online learning.

While Mandinach (2005) emphasizes the importance of developing assessment techniques that make the feedback loop between instruction and assessment even more meaningful for students, Elwood and Klenowski (2002) explain assessment for learning is to "enable students through effective feedback, to fully understand their own learning and the goals they are aiming for" (p. 243). Long before online assessment was an issue, Nelson-Le Gall (1981) promoted improving student metacognition skills by guiding them toward more appropriate, better-timed help-seeking skills. Such skills are proving to be essential in online courses. A multiple measures assessment approach is promoted by Gibson and Dunning (2012): "Instructors are encouraged to use multiple assessment and feedback mechanisms tied to the course objectives and providing evidence that a student has acquired the information, understanding, and skills necessary to demonstrate that learning objectives have been met" (p. 217).

The importance of behavioral objectives is emphasized by Grandzol (2004) in a report on matching pedagogy and assessment techniques for an online Master of Business Administration (MBA) statistics course. Van der Merwe (2011) also tends toward behavioral measures, not surprisingly finding that students who spent more time in the virtual classroom (a direct, behavioral assessment measure) generally achieved higher indirect assessment measures, such as grades. Swan, Shen, and Hiltz (2006) reinforce the importance of assessing behavior as well as outcomes for relevant student feedback, reporting a variation of the old management maxim, "What gets measured gets done." They state:

> Value in any instruction system comes from assessment; what is assessed in a course or a program is what is valued; what is valued becomes the focus of activity. The link to learning is direct. Instructors signal what knowledge, skills and behaviors they believe are most important by assessing them. Students quickly respond by focusing their learning accordingly. (p. 45)

While learning outcomes assessment is still an emerging field, the literature for learning outcomes assessment in online courses clearly values *both* direct, behavioral measures and indirect measures. When assessment puts to use all of the data organically generated by the teaching and learning processes, a richer picture, with more depth and insight, is generated. In a review of the literature of online assessment, Gikandi, Morrow, and Davis (2011), found a wide variety of techniques and variables being used, including measurements of self-testing quizzes, discussion forums, and portfolios. They report, however, that the most effective way to address threats to validity and reliability of online formative assessment is the use of direct, ongoing assessment activities. That need for authenticity of engagement, not just participation, should drive the choice of measured indicators.

As Robinson and Hullinger (2008) explain, evaluating online learning needs to go beyond measures of skills and knowledge acquisition to also consider the quality of the learning experience as a whole. They posit that measures of student engagement are one way to determine that. In the mid 1980s, Jacobi, Astin, and Ayala (1987) proposed a clear definition of student engagement. Their definition uses the time and effort students invest in course activities as indicators of how engaged they are. Kuh (2003) also proposed measures of time and effort in studying, practicing skills, obtaining feedback analysis, and problem solving as defining factors of engagement.

Karaksha, Grant, Anoopkumar-Dukie, Nirthanan, and Davey (2013) studied graduate students with the similar measures of time and energy as proxies for engagement, but added measurement of "quality of effort and involvement in productive learning activities" (p. 1). Conrad and Donaldson (2004) also measured and reported qualitative factors: "This collaborative acquisition of knowledge is one key to the success of creating an online learning environment. Activities that require student interaction and encourage a sharing of ideas promote a deeper level of thought" (p. 5). While previous researchers provided insights into engagement through the indirect measure of reported student perceptions, recent advances in analytics are allowing us to measure and analyze direct observations of student behavior along with their indirect, self-reported behavior and perceptions.

Data Analytics and Online Learning

The course management systems (CMS) underlying online classes provide colleges and universities with large amounts of data that, until recently, have not been analyzed to improve student learning or to evaluate online programs. CMS integrate message forums along with a wide array of designed applications for managing messages, course content, and syllabi, and for submitting assignments. They also collect information about student performance, participation, and system use.

This research uses data collected through the Sakai learning management system,[3] which came into existence in 2004, when programmers at the University of Michigan and Indiana University began working together to develop a new online CMS. Sakai represents an open source alternative to commercially available products such as Blackboard/WebCT.

The model developed in this article builds on previous academic research in learning analytics, which uses data from similar systems to identify students who are at risk of performing poorly and makes recommendations for improving student performance (Ma & Klinger, 2000). In management, analytics involves standardizing reports, explaining data, formulating higher order questions, analyzing, forecasting, predicting behavior, and ultimately, optimizing organizations (Davenport & Harris, 2007). Learning analytics has been used to recognize patterns in students' online behavior to better evaluate student learning (Zaiane & Luo, 2001). Previous research has assessed discussion forums in an effort to improve instructors' abilities to assess students (Dringus & Ellis, 2005). In addition, demographic data (J. P. Campbell, 2007) and student academic performance (Morris, Wu, & Finnegan, 2005) have been used to predict student success. Most similar to our research is Purdue University's Course Signals, which developed an academic alert system, based on overall grade point average (GPA) and measures of CMS activity, to inform instructors about students at risk of performing poorly in class, and to alert potentially at risk students early as well (Arnold, 2010).

Trends in higher education suggest that online education has experienced tremendous growth in the last decade. Online education at the graduate level in public affairs education is not a passing fad likely to go away. At the same time, recent shifts in online education suggest many programs will experiment with larger classes, which could limit the opportunities for discussion between students and faculty. Experiments with MOOCs have resulted in single-digit retention rates. As online MPA programs wrestle with different program structures and pedagogical approaches, it is critical for researchers to use analytic methods to determine success factors for improving online student learning outcomes and streamline workload for instructors.

METHODS

The Sakai online learning management system used by Marist College integrates multiple communication mechanisms, including announcements, messages, calendars, forums, chat rooms, and videoconferencing. In addition, the system provides a suite of tools to organize course content such as the syllabus, assignments, lessons, and resources. As students and faculty use these systems, data are collected regarding student participation—the amount of course content read, the number of forum discussion threads read, the number of contributions students made to the discussion forums, and the number of exams and assignments submitted, as well as grades given. In addition to the data collected by CMS systems, most universities also maintain enterprise resource planning (ERP) systems that manage enrollment, student records, financial aid, finances, human resources, and advancement. This article uses learning analytics to assess data from enrollment systems, an ERP, and a CMS to predict student success in an online MPA program.

Data

Data were collected through the OAAI project at Marist College. The data came to us already having been cleaned, which means a process is in place for regularly downloading, verifying, and storing institutional data. As part of data collection, all information is de-identified, given a unique random string, and scrambled for storage on the network.

For this article, a subset of 1,073 cases from the OAAI data set was used. It consists of records for graduate students enrolled in the MPA program at Marist College for at least one course from Fall 2010 through Fall 2011 semesters. The Marist MPA program enrolls approximately 200 to 300 students at any point in time, with classes taught in person, at locations in the Hudson Valley in New York State and with about 60% of students participating in the program partially or fully online. The focus of this analysis is the online courses.

Measures

Variables collected include a standard set of demographic characteristics, such as student age and gender, as well as static variables at the time of the course, such as their full-time/part-time status, the class size, whether or not they are on academic probation, and their cumulative GPA.

Dynamic variables related to course participation include a partial grade in the class, number of forum posts, number of times content is read, number of forums read, number of sessions opened, number of assignments submitted by the student, number of exams taken and submitted by the student, and number of assignments read by the student. See Table 1 for a full list of the input data set and the definition of variables.

In addition, variables were included in the analysis only if they had less than 20% of values missing or null, and only if 80% of the class was using a particular online tool. A few variables did not pass the latter criteria, such as number of exams taken and submitted by the student and number of assignments read by the student. For missing data in cases where variables passed the 20% threshold, imputations were used that replaced missing data with mean values (and medians or modes for discrete data).

Variables having to do with class participation were standardized to the norm of the class. Although the Marist MPA program has established best practices for online learning, not all courses use CMS tools identically. Course materials and instructor preferences introduce some variability in both the tools used and the extent to which tools are used. Variations in use of online tools can be due to course content (e.g., a course on statistics might use tests more than a policy course that emphasizes debates) or due to teaching style. To account for these differences in use of online tools, all variables having to do with class participation are standardized to the class mean. For example,

TABLE 1.
Input Data Set and Variable Definitions

Type	Variable Name	Description
Predictors	Enrollment	Course size in categories of 0–20, 21–30, and 30+
	DC_Age	Student's age, discretized into 20–29, 30–39, 40–49, and 50+
	Gender	Student's gender (1 = female, 0 = male)
	FTPT	Full-time or part-time student (FT = 1, PT = 0)
	Cum_GPA	Cumulative grade point average in the program
	Academic_Standing	MPA program academic standing (0 = probation)
	RMN_Score_Partial*	Score computed from partial contributions to the final grade submitted by instructor
	R_Content_Read*	Number of times a section in the lessons content is read by the student
	R_Forum_Post*	Number of times forum posts are made by the student
	R_Forum_Read*	Number of times the discussion forum threads are read by the student
	R_Lessons_View*	Number of times the lesson tool is accessed by the student
	R_Sessions*	Number of times the student has logged into the course
Target	Academic Risk	1 = at-risk (grade B– or below); 0 = good standing

Note. *Calculated as a ratio by dividing by the average course value.

instead of looking at the raw number of forum posts, we looked at how many forum posts a student contributes *relative to their peers in the class*. The same procedure is used to derive relative measures for amount of content read, number of forums read, and number of sessions opened. The main dependent variable for analysis, whether or not a student is at risk, is defined as earning a grade of B- or less in the course. Students are required to earn a 3.0 GPA or above in order to graduate from the program.

Statistical Analysis

For our analysis, we used logistic regression to derive coefficients for predicting at-risk status, and then evaluated how strong the models are in predicting whether or not students succeeded in the course. The data were split into two sub-samples to create a sample for training the data and a data set for testing it. The training data set of 557 cases (the initial data set for training was 743 cases, with a 75% re-sample size) was used to derive the logistic regression coefficients. The testing data set of 330 cases was used to calculate how well the model predicts at-risk status. This two-step procedure provides some protection against over-fitting the data in our prediction model. For training the models, we also use stratified sampling to oversample the at-risk group (and sub-sample the remaining group) to create a more balanced data set with enough predictive power for the target population. The testing data set is not balanced or sampled in any way.

Once we derived coefficients from the training data set, we tested the model's predictive power by deriving a "confusion matrix" (also known as an "error matrix") using the testing data set. This gives us an indication of the mode recall/sensitivity (the percentage of at-risk group identified), the number of false positives, and the accuracy of the model. A higher percentage of recall (and lower percentage of false positives) indicates a more powerful model. Increasing recall and decreasing false positives are some-what in conflict with each other—at a certain level, attaining higher recall will increase false positive rates. Likewise, steps taken to reduce false positives are likely to reduce the percentage of cases identified.

RESULTS

Students range in age from 20 to 60 years old, with 55% being female, and 56% being part-time students. Class size varies from 8 to 45 students per course, and courses represent all stages of the program, from the first introductory course to the final capping course. Cumulative GPAs range from 1.4 to 4.0, with 11% of students on academic probation.

Using a logistic regression model to predict at-risk status (see Table 2), we find that course-related dynamic variables such as partial grade score, relative amount of content read, and relative number of forum posts are all statistically significant ($p < .05$ or less) and associated with lower odds of being at risk [OR = 0.95, 0.34, 0.38, respectively]. Relative number of forums read and online sessions opened are statistically significant and related to higher risks [OR = 2.3 and 1.9]. Cumulative GPA is highly significant with higher grades strongly associated with lower risk [OR = 0.04]. Static variables such as age and gender are not found to be statistically significant in the logistic regression model, nor are full-time versus part-time status, current academic standing, or class size.

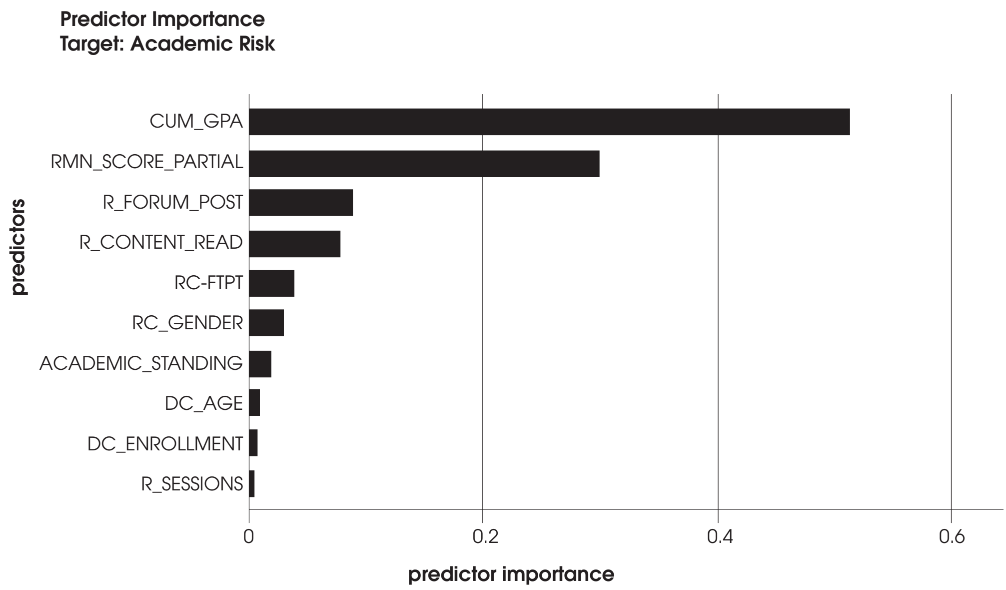
Looking at a graphic of the relative importance of each variable in the full model (see Figure 1), we see the strongest predictor of success in a course is the student's previous cumulative GPA. It is highly statistically significant, and its relative importance greatly overshadows any other factor, including how well they are doing in the course presently, all of the dynamic variables related to classroom participation, and all the static demographic variables.

TABLE 2.
Logistic Regression Model for At-Risk Students

Variable	B	Std. Error	Wald	Sig. *p*	Odds Ratio Exp(B)
Partial grades score	−0.052	0.01	24.189	< .001	0.95***
Content read	−1.078	0.271	15.784	< .001	0.34***
Forum posts	−0.963	0.382	6.35	.012	0.382**
Forums read	0.817	0.304	7.218	.007	2.264***
Sakai sessions opened	0.646	0.213	9.197	.002	1.907***
Cumulative GPA	−3.326	0.386	74.392	< .001	0.036***
Gender (F)	−0.499	0.282	3.13	.077	0.607*
Full-time (vs. part-time)	−0.402	0.268	2.245	.134	0.669
Good academic standing	−0.306	0.317	0.93	.335	0.737
Class size (0–20)	−0.236	0.372	0.403	.525	0.789
Class size (21–30)	−0.411	0.287	2.056	.152	1.509
Age (20–29)	−0.198	0.609	0.106	.745	0.82
Age (30–39)	−0.407	0.6	0.46	.497	0.666
Age (40–49)	−0.164	0.626	0.069	.793	0.849
Intercept	17.285	1.689	104.759	< .001	

Note. *** $p < .01$; ** $p < .05$; * $p < .10$.

FIGURE 1.
Relative Importance of Factors (Full Model)



The second strongest predictor of ultimate performance is the partial grade in the course, which should come as little surprise, but which has much less influence than cumulative GPA. Consistent with analytic techniques for forecasting, predicting behavior, and optimizing performance, including partial grade as a predictor in the model is essential for identifying students at risk of performing poorly in any given course, despite methodological concerns in using an independent variable that constitutes part of the dependent variable.

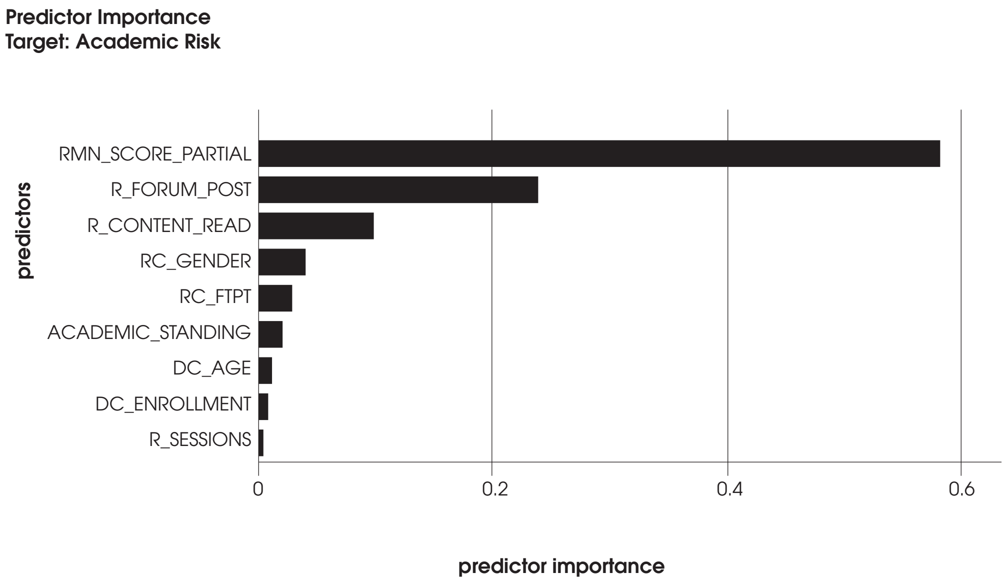
In addition to grades, dynamic variables related to class participation are all predictors of at-risk status. The most important factors are relative number of forum posts and amount of content read. Of less importance, but still statistically significant, are number of sessions opened.

The full model's overall accuracy in predicting at-risk cases is 84.84%. In terms of recall (accuracy within at-risk population), 80.43% of cases are correctly predicted, with a false positive rate of 13.5%. That gives precision of

TABLE 3.
Full Model Predictive Power

	Predicted At-Risk	Predicted Not At-Risk
At-Risk	37 or 80.4% (true positive)	9 or 19.6% (false negative)
Not At-Risk	38 or 13.5% (false positive)	243 or 86.5% (true negative)

FIGURE 2.
Relative Importance of Factors (Model Without Cumulative GPA)



49.33%. Using the testing sample, that means that 37 of the 46 cases would be correctly identified as at risk, with 38 false positives out of 281 not at risk (see Table 3).

Given the dominance of cumulative GPA in the model, we present the same graphic without that variable to highlight the relative impact of other factors. Looking at a graphic of predictor importance that drops cumulative GPA, we see that partial grade dominates (see Figure 2). Among the other class participation variables, relative number of posts is most important. Second, but less important, is the relative amount of content read. Of relatively little importance are, in decreasing order, gender, full-time or part-time status, class size, academic standing, age, number of online sessions opened, and number of forum posts read. (Note: None of the static demographic variables was statistically significant in the full model, while all the dynamic variables were statistically significant.)

The model without GPA explains less of the variance in academic at-risk status. In the prediction matrix, the overall performance of the model was 78.5%, with recall of 69.4% (accuracy among at-risk population), a false positive rate of 19.9%, and precision of 37.7%.

DISCUSSION

We find that a full set of online course participation variables are important second-tier predictors of at-risk status in an ongoing course, a result which is new to the literature for graduate education. Specifically, we find that students are more likely to perform poorly if they show low relative levels in the number of posts to online discussion forums and in the number of occasions course content is opened online, or if they show high relative levels of reading forums or opening online sessions multiple times. Previous work by Van der Merwe (2011) found only that more time in the online classroom was associated with success. Thus, the full set of variables collected from CMS potentially offers new indicators that can be used as early alerts for online course performance.

Similar to the existing literature from undergraduate hybrid courses, we find that cumulative GPA and partial grades are the biggest predictors of risk status. Learning analytics projects at the University of Alabama and Sinclair Community College also found that overall GPA was a major determinant of success in courses (J. P. Campbell, 2007). We do not find that characteristics of students, in our case age, gender, full-time or part-time status, or academic probation, are strong predictors of being at risk for students in an online graduate MPA program. Earlier studies found characteristics such as race, family income level, and distance from home to be important factors (J. P. Campbell, 2007; Morris, Wu, & Finnegan, 2005).

Our models for predicting at-risk status correctly identify over 80% of cases, with a relatively low false positive rate. That compares favorably with the existing literature, with a 74.5% accuracy rate for an online general education course (Morris et al., 2005), and 67% overall (80% for freshmen) in Purdue University's Course Signals project. Our prediction rates are similar to the Marist OAAI project with undergraduates at five institutions (Jayaprakash et al., 2014). Given the strength of the prediction model, if the cost of an intervention is relatively low (involving slight additional work by the instructor), and if the impact of false positives is of little harm, then this is a good target group for early intervention.

A limitation of our models that has implications for interventions is that they indicate association between factors but do not necessarily show causal pathways. This becomes less of an issue if interventions are low cost and at low risk of causing harm. In addition, the findings give a sense of the respective strength of tools that are systematically used by the Marist MPA online program—and little insight into effectiveness of tools that are not used much. Consequently, the actual coefficients and information on relative importance of various online approaches would be useful to master's programs using similar tools. However, this *process* could be adopted by other programs with their own predictors.

We are also limited by the variables that were immediately available to us, and we plan to collect a richer data set in further models, such as (a) previous degree GPA, (b) years since completed last degree, and (c) number of credits earned.

CONCLUSION

In the past few decades, we have seen a large growth in online learning—in public administration and public affairs graduate education as well as across all of higher education (Allen & Seaman, 2011; Ginn & Hammond, 2012). Online education holds tremendous promise not only to reach new students, but also to engage them in ways that traditional classrooms may not be able to do (Austin, 2009; Brower & Klay, 2000). The promises of online education are neither universal nor guaranteed, as online courses can require greater workload, from students and instructors alike, and can suffer from lower retention rates (Barth, 2004; Kolowich, 2013).

Creating a quality online course or program goes beyond simply getting faculty onboard or even implementing learning analytics. The quality of instruction and content remain the central factors. But in terms of adapting a quality course to the online environment, other research has documented the need for training instructors, providing time and institutional supports to design courses for online audiences, building in considerations of how to increase instructor-to-students or peer-to-peer interactions by using various types of instructional design methods, setting departmental limits on course enrollments so instructors are more focused on communicating and interacting with online students, and providing instructors administrative and technical support as well as guidance from experienced online professors. (See Yang & Cornelious, 2005 for a review.)

Given a program with a good foundation of administrative and institutional supports, the findings of this article illustrate one example of how to enrich online education. Putting multivariate data with direct measures to use in a way that provides real-time information to course instructors gives them another tool for effectively monitoring student progress and helping students persist toward course and degree completion, at little additional time or money cost. Consistent with analytics literature, the research seeks to move beyond management analysis toward predicting future behaviors, automating an alert system in an attempt to optimize the educational experience of students and increase efficiency of teaching. Having specific, directly measured behavior variables with high validity gives instructors a valuable tool for calling attention to at-risk students early enough in the course that these students can take effective corrective action and still succeed.

When models produce high accuracy and involve low impact from false negatives, there is also an opportunity to test an early-alert intervention. Such strategies have been tried in higher education at the undergraduate or community college level, with the Purdue University Course Signals and the Marist College OAAI projects, in which messages were sent to students flagged as being at risk by the learning analytics data. The interventions involved alerting the students about concerns with their performance and providing steps the student should take to improve, such as meeting with the instructor or getting a tutor, or links to help desks, existing student services, or similar programs. The Course Signals and OAAI projects demonstrate that interventions using learning analytics can successfully move many students up the hierarchy in terms of course performance and decrease the number of students failing the courses compared to similar students who do not receive an intervention. Based on experience with the OAAI project for undergraduate students, the Marist College MPA program will test giving students alerts about their at-risk status during the course, along with direction and resources for improving their performance.

One of the central challenges of optimizing online programs is how to leverage the good (increase quality and richness of education and expanding enrollments), while preventing the bad (low retention, increased faculty and administrative burden). Efforts at supporting students often come after the fact (e.g., probation and/or remedial schooling for poor performance following a course, or indirect student measures) or are based on limited data (e.g., reports from faculty, course attendance, or midterm grades).

Grades have long been considered an insufficient indicator of student success by learning outcomes assessment practitioners. A grade is an indicator of how a student is doing, but it gives no information on why a student is performing at that level. This study shows that a variety of directly measured behaviors give insight for answering the question of why a student is earning a particular grade, not just how they are earning it, demonstrating that it is worthwhile to track these behaviors. If a college or university is not yet equipped to provide instructors with reports on key analytics, faculty and administrators can approximate the effect by using course statistics tools as an adjunct to the grades. Scanning such information as recency of log-ins and course component access, frequency of log-ins and course component access, and patterns of interaction between students on discussion forums can provide important insight into specific behavioral changes that will help student persistence and retention. Use of these types of tools can be further aided by faculty discussions that help benchmark common expectations for participation and classroom engagement, for both students and faculty members.

NOTES

1. See http://www.confluence.sakaiproject.org/pages/viewpage.action?pageId=75671025.

2. Separate from the OAAI project, the MPA program began a process in 2009 to deliver its hybrid and online courses in a more uniform and systematic way across courses and modalities of teaching, by following best practices of online teaching, particularly those promoted by the national organization Quality Matters. Subsequently, MPA online courses have provided a common set of guidelines, expectations, and tools for online experiences. For example, all courses require use of online lesson notes and participation by students in a certain number of online discussion forums each week. The use of consistent tools and approaches to online learning, along with the relatively large sample sizes afforded by the MPA program and the availability of an expanded data set on students thanks to the OAAI initiative, makes this a propitious group for review.

3. See www.sakaiproject.org.

REFERENCES

Allen, I. E., & Seaman, J. (2006). *Making the grade. Online education in the United States.* Needham, MA: Sloan Consortium.

Allen, I. E., & Seaman, J. (2011). *Going the distance: Online education in the United States, 2011.* Needham, MA: Sloan Consortium.

Arnold, K. (2010). Signals: Applying academic analytics. *EDUCAUSE Review*, http://www.educause.edu/ero/article/signals-applying-academic-analytics.

Austin, E. K. (2009). Limits to technology-based distance education in MPA curricula. *Journal of Public Affairs Education*, *15*(2), 161–176.

Barth, T. J. (2004). Teaching PA online: Reflections of a skeptic. *International Journal of Public Administration, 27*(6), 439–455.

Bocchi, J., Eastman, J. K., & Swift, C. O. (2004). Retaining the online learner: Profile of students in an online MBA program and implications for teaching them. *Journal of Education for Business, 79*(4), 245–253.

Bonk, C. J., & Reynolds, T. H. (1997). Learner-centered Web instruction for higher-order thinking, teamwork, and apprenticeship. In B. H. Khan (Ed.), *Web-based instruction* (167–178). Englewood Cliffs, NJ: Educational Technology Publications.

Brower, R. S., & Klay, W. E. (2000). Distance learning: Some fundamental questions for public affairs education. *Journal of Public Affairs Education, 6*(4), 215–231.

Campbell, H. E. (2006). Cheating, public administration education, and online courses: An essay and call to arms. *Journal of Public Affairs Education, 12*(1), 33–47.

Campbell, J. P. (2007). *Utilizing student data within the course management system to determine undergraduate student academic success: An exploratory study* (Doctoral dissertation). Retrieved from ProQuest (ISBN 054930293X).

Conrad, R., & Donaldson, J. A. (2004). *Engaging the online learner: Activities and resources for creative instruction.* San Francisco: Jossey-Bass.

Davenport, T., & Harris, J. (2007). *Competing on analytics.* Boston, MA: Harvard Business School Press.

Dringus, L. P., & Ellis T. (2005). Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education*, *45*(1), 141–160.

Ebdon, C. (1999). Teaching public finance administration online: A case study. *Journal of Public Affairs Education, 5*(3), 237–246.

Elwood, J., & Klenowski, V. (2002). Creating communities of shared practice: The challenges of assessment use in learning and teaching. *Assessment & Evaluation in Higher Education, 27*(3), 243–256.

Feenberg, A. (1999). Distance learning: Promise or threat. *Crosstalk, 7*(1), 12–14.

Gibson, P. A., & Dunning, P. T. (2012). Creating quality online course design through a peer-reviewed assessment. *Journal of Public Affairs Education, 18*(1), 209–228.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57*(4), 2333–2351.

Ginn, M., & Hammond, A. (2012). Online education in public affairs: Current and emerging issues. *Journal of Public Affairs Education, 18*(2), 247–270.

Grandzol, J. R. (2004). Teaching MBA statistics online: A pedagogically sound process approach. *Journal of Education for Business, 79*(4), 237–244.

Jacobi, M., Astin, A., & Ayala, F. (1987). College student outcomes assessment. Washington, DC: Clearinghouse on Higher Education.

Jayaprakash S., Moody E., Lauría E., Regan J., & Baron J. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics, 1*(1), 6–47.

Karaksha, A., Grant, G., Anoopkumar-Dukie, S., Nirthanan, S. N., & Davey, A. K. (2013). Student engagement in pharmacology courses using online learning tools. *American Journal of Pharmaceutical Education*, *77*(6), 125.

Kolowich, S. (2013). The professors who make the MOOCs. *The Chronicle of Higher Education,* Retrieved from http://chronicle.com/article/The-Professors-Behind-the MOOC/13790/#id=overview

Kuh, G. D. (2003). What we're learning about student engagement from NSSE. *Change, 35,* 24–31.

Lauría E., Moody E., Jayaprakash S., Jonnalagadda N., & Baron J. (2013). Open academic analytics initiative: Initial research findings. In Suthers, Verbert, Duval, and Ochoa (Eds.), *Proceedings of the 2013 Learning Analytics and Knowledge Conference* (pp. 150–154), Leuven, Belgium.

Lauría E., Baron J., Devireddy, M., Sundararaju V., & Jayaprakash S. (2012). Mining academic data to improve college student retention: An open source perspective. In Shum, Gasevic, and Ferguson (Eds.), *Proceedings of the 2012 Learning Analytics and Knowledge Conference* (pp. 139–142), Vancouver, BC, Canada.

Leavitt, W. M., & Richman, R. S. (1997). The high tech MPA: Distance learning technology and graduate public administration education. *Journal of Public Administration Education, 3*(1), 13–27.

Ma, X., & Klinger, D. A. (2000). Hierarchical linear modelling of student and school effects on academic achievement. *Canadian Journal of Education, 25*(1), 41–55.

Mandinach, E. (2005). The development of effective evaluation methods for e-learning: A concept paper and action plan. *The Teachers College Record, 107*(8), 1814–1836.

Mingus, M. S. (1999). Toward understanding the culture of Internet-mediated learning. *Journal of Public Affairs Education, 5*(3), 225–235.

Morris, L. V., Wu, S.-S., & Finnegan, C. L. (2005). Predicting retention in online general education courses. *The American Journal of Distance Education, 19*(1), 23–36.

Nelson-Le Gall, S. (1981). Help-seeking: An understudied problem-solving skill in children. *Developmental Review, 1,* 224–226.

O'Leary, R., & Stowers, G. N. L. (1999). Computer conferencing in the public affairs classroom. *Journal of Public Affairs Education*, *5*(1), 57–66.

Palloff, R. M., & Pratt, K. (2007). *Building online learning communities: Effective strategies for the virtual classroom.* San Francisco, CA: John Wiley & Sons.

Perez-Pena, R. (2012, July 17). Top universities test the online appeal of free. *The New York Times,* p. A15.

Rahm, D., & Reed, B. J. (1997). Going remote: The use of distance learning, the World Wide Web, and the Internet in graduate programs of public affairs and administration. *Public Productivity & Management Review, 20*(4), 459–474.

Rahm, D., Reed, B. J., & Rydl, T. L. (1999). Internet-mediated learning in public affairs programs: Issues and implications. *Journal of Public Affairs Education, 5*(3), 213–223.

Robinson, C. C., & Hullinger, H. (2008). New benchmarks in higher education: Student engagement in online learning. *Journal of Education for Business, 84*(2), 101–109.

Rogers, E. M. (2010). *Diffusion of innovations.* New York, NY: Simon and Schuster.

Stowers, G. N. (1995). Getting left behind? Gender differences in computer conferencing. *Public Productivity & Management Review, 19*(2), 143–159.

Swan, K., Shen, J., & Hiltz, S. R. (2006). Assessment and collaboration in online learning. *Journal of Asynchronous Learning Networks, 10*(1), 45–62.

Van der Merwe, A. (2011). Can online learning boost academic performance? A microeconomics study. *International Business & Economics Research Journal (IBER), 10*(8), 45–56.

Vonderwell, S. K., & Boboc, M. (2013). Promoting formative assessment in online teaching and learning. *TechTrends, 57*(4), 22–27.

Vonderwell, S., Liang, X., & Alderman, K. (2007). Asynchronous discussions and assessment in online learning. *Journal of Research on Technology in Education, 39*(3), 309.

Waldrop, M. (2012, March 13). Massive open online courses, aka MOOCs, transform higher education and science. *Scientific American.* Retrieved from http://www.scientificamerican.com/article/massive-open-online-courses-transform-higher-education-and-science/

Yang, Y., & Cornelious, L. F. (2005). Preparing instructors for quality online instruction. *Online Journal of Distance Learning Administration, 8*(1). Retrieved from http://www.westga.edu/~distance/ojdla/spring81/yang81.htm.

Zaiane, O. R., & Luo, J. (2001). Web usage mining for a better web-based learning environment. In *Proceedings of conference on advanced technology for education, Banff, Alberta* (pp. 60–64).

ABOUT THE AUTHORS

Jay Bainbridge is an assistant professor of public administration in the School of Management at Marist College. His research applies statistical and research methods to public sector problems. His current research focuses on trends, causes, and solutions to homelessness, especially with respect to the street homeless, and use of performance measurement in the public sector. Prior to joining Marist College, Bainbridge was assistant commissioner of policy and research at the NYC Department of Homeless Services.

James Melitski is an associate professor of public administration in the School of Management at Marist College. His teaching and research interests include organizational theory and behavior, performance measurement, public strategic management, and digital government. His research assesses the use of technology to engage citizens, implement successful programs, and improve the performance of public organizations. His work has appeared in journals such as *Public Performance and Management Review, American Review of Public Administration, Public Administration Quarterly, International Journal of Electronic Government Research, International Journal of Public Administration,* and *International Journal of Organization Theory and Behavior.* He is a leading authority on e-government and continues to write and consult throughout the United States and internationally on the subject.

Anne Zahradnik is an assistant professor of health care administration in the School of Management at Marist College, and a Teagle Scholar. Her research focuses on health care communication and learning outcomes assessment.

Eitel J. M. Lauría is a professor and graduate director of information systems at the School of Computer Science & Mathematics at Marist College. His broad research interests cover the fields of data and decision science, business intelligence, data mining and predictive analytics, and statistical machine learning, focusing on the application of these disciplines in a variety of domains. His research has been published in a number of prestigious journals, including *Decision Support Systems, European Journal of Operational Research, ACM Journal of Data and Information Quality,* and *Expert Systems with Applications.* He is coauthor of a textbook on data and information quality published by MIT/IQ. He has served as an information systems and technology consultant to IBM, Microsoft, Exxon Mobil, Reuters, Hewlett-Packard, Stet-France Telecom, GE Global Research, Ryder, and the World Bank, among other global corporations. Lauría served as the analytics lead of the Open Academic Analytics Initiative (OAAI).

Sandeep Jayaprakash is a learning analytics specialist in the Academic Technology group at Marist College. He works on applying big data and data-mining techniques in education to identify trends and patterns among learners in a variety of academic settings and their learning methodologies. The insights gained are used to further enrich the learning experience of faculty, students, and others, through building practical and innovative applications to supplement their learning efforts. He has a master's degree in software development, and his current research focuses on predictive modeling and building large-scale academic early alert and retention systems to identify at-risk student populations.

Josh Baron is senior academic technology officer at Marist College, responsible for supporting instructional technology initiatives, including distance learning, faculty training, and student support. He plays a leadership role on campus in the area of strategic planning for instructional technology, is currently serving on the Apereo Foundation Board of Directors and was principal investigator on the Open Academic Analytics Initiative (OAAI). He graduated from the University of Michigan with a BS in aerospace engineering, holds an MS in educational technology leadership from the George Washington University, and has successfully completed the Institute for Management and Leadership in Education at Harvard University's Graduate School of Education.