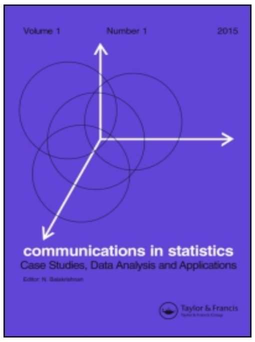


Taylor & Francis
Taylor & Francis Group

Communications in Statistics: Case Studies, Data Analysis and Applications

ISSN: (Print) 2373-7484 (Online) Journal homepage: www.tandfonline.com/journals/ucas20

Designing early detection and intervention techniques via predictive statistical models—A case study on improving student performance in a business statistics course

Sinjini Mitra & Zvi Goldstein

COMMUNICATIONS IN STATISTICS: CASE STUDIES, DATA ANALYSIS AND APPLICATIONS
2015, VOL. 1, NO. 1, 9–21
http://dx.doi.org/10.1080/23737484.2015.1063409

CASE STUDY

Designing early detection and intervention techniques via predictive statistical models—A case study on improving student performance in a business statistics course

Sinjini Mitra and Zvi Goldstein

Information Systems and Decision Sciences Department, California State University, Fullerton, California, USA

ABSTRACT
This article presents a comprehensive study of factors that potentially impact student performance and success in a "bottleneck" college-level course in Business Statistics, with the goal of devising effective intervention methods to provide additional support to students who are at risk of failing. The latter are based on statistical models that predict the probability of failure based on relevant factors identified earlier. These models report high accuracy in detecting at-risk students as assessed by cross-validation techniques. Moreover, implementation of our techniques yielded positive outcomes, indicating that those students who took advantage of the intervention significantly improved their performance.

ARTICLE HISTORY
Received 19 February 2015
Accepted 15 June 2015

KEYWORDS
College-level courses; detection; education; factors; intervention; prediction

1. Introduction

Students in Business Statistics courses are struggling. Data from across the United States reveal massive failure rates, high dropout rates as well as lack of degree completion. In fact, some universities report a 6-year graduation rate of only 50%, which is way below the national average. Several state universities receive a vast majority of transfer students from local community colleges. Analysis by the Community College Research Center at Columbia University's Teacher's College on groups of studies indicates that about two thirds of community college students enter college with academic skills weak enough to require remedial academic interventions along with comprehensive support services in order to progress in college-credit learning (Bailey 2008). It is thus necessary to design effective intervention techniques that will help students succeed in college and graduate in a timely manner. Toward this, it is necessary to form an understanding of the factors that contribute to student success, as well as the challenges that students face in such courses in order to help them achieve a higher level of success. In this article, we present a comprehensive study of factors that are potentially associated with student performance in a bottleneck course at the college of Business in a public university in California in order to implement successful intervention techniques.

Of all courses in the College of Business, the introductory course in Statistics is considered "bottleneck" since it historically has very high failure rates. Many of these students do not have a background that includes courses in quantitative topics, work full-time outside of school, and struggle with balancing their life in and outside school. For such students, early detection and intervention methods have the potential of improving performance and success rate by providing them with an opportunity to seek additional help and support from the instructor and other resources available on campus.

The rest of the article is organized as follows. Section 2 contains an overview of existing literature on factors contributing to student success and various retention and intervention methods being applied in college-level courses. Section 3 describes the data collected for the study and states our research goals. In Section 4, we introduce the statistical tools used to analyze the data, followed by our initial results and findings in Section 5. Our prediction models are introduced in Section 6 along with our proposed intervention technique. The results from practical imple-

CONTACT Sinjini Mitra smitra@fullerton.edu Information Systems and Decision Sciences Department, California State University, 800 N. State College Blvd., Fullerton, CA 92831, USA.

mentation of our intervention are presented in Section 7 and we finally conclude with a discussion in Section 8.

2. Literature review

The level of success students achieve in their years at college has far-reaching implications for students' personal and professional lives by influencing their academic self-esteem, persistence in elected majors, and perseverance in higher education. Success in early semesters at college and experience in introductory college courses also ultimately impact students' post-college experiences and adult life, such as career choice, personal income and level of success, and degree and nature of participation in community life. However, disaffection with low performance in introductory college classes is a serious problem at colleges and universities nationwide (Horn and Premo 1995; Horn et al. 2002). For more than two decades, research has shown that student success in STEM (Science, Technology, Engineering and Mathematics) disciplines is most negatively affected by students' lack of success in the gateway courses that develop essential skills and introduce students to disciplinary studies (Tobias 1990; Seymour and Hewitt 1997). The problem is especially evident in introductory business, mathematics, and science courses. Such courses are often required and are integral components of an undergraduate education, yet many students who enroll in these courses achieve moderate or low levels of success in them. This often results in significant attrition of talented students in these areas of study (Gainen 1995; Congress of the United States Office of Technology Assessment 1998).

An abundance of research has been performed in the last few decades to identify factors that contribute to student performance in bottleneck courses in Business, Mathematics, and other science-based disciplines. These range from demographic factors (such as gender, ethnicity) to academic factors (such as prior academic history, grade point average [GPA]). Brower and Ketterhagen (2004) and Herndon and Moore (2002) have shown that students belonging to some ethnic groups such as African Americans and Hispanics are more likely to drop from Business, Mathematics, and science-related majors. Furthermore, female students are also shown to be more likely to drop from these majors than males. All these clearly indicate a need to address these issues in order to bridge the existing gap.

Studies have shown different subsets of factors that are associated with student performance in different types of courses, thus indicating that the type of discipline plays a major role in determining factors contributing to students' success. Cognitive and academic variables have been shown to be only adequate predictors of success in introductory business, marketing, and economics courses. Sachdeva and Sterk (1982), Eskew and Faley (1988), Liesz and Reyes (1989), and Doran, Boullion, and Smith (1991) report that locally written and administered placement exams that measure student content knowledge and reasoning skills predict student performance in introductory finance courses. Eckel and Johnson (1983) report that the ACT score in math predicts success in introductory accounting courses. However, some studies contradict this conclusion and suggest that standardized entrance exam scores are not effective predictors in introductory accounting courses (Brown 1966; Ingram and Peterson 1987). The cognitive factors that have been most widely considered as potential predictors of college mathematics achievement are SAT and ACT scores. Another nearly universal variable that predicts students' success in freshman business, mathematics, and science courses is high school grade point average (GPA). GPA is neither a cognitive nor an affective variable; it is neither a measure of aptitude nor a state of mind. Instead it is a holistic measure of performance. Similar to data on mathematical skills, data on students' GPAs are widely available. High school and college performance seems to be a more reliable predictor of student success than are entrance exam scores in introductory courses in the business field. Brown (1966) reports that high school GPA adequately predicts success in accounting courses, and other investigators (Bellico 1972; Cohn 1972; Ingram and Peterson 1987; Borde 1998) report that college GPA is a valid predictor of success in economics and marketing courses.

In addition to the cognitive and quantitative factors, noncognitive factors or qualitative factors have been used successfully to predict grades in many gateway courses, in business, mathematics, and other areas of study. Although such factors are often overlooked (Glesne 1998), some studies have shown that these variables are more useful than cognitive variables in predicting the academic success of nontraditional students (e.g., Sedlacek 2002). Meece et al. (1982) found a relationship between student motivation and academic self-concept (a student's personal opinion toward his or her academic skills) in introductory Math courses. Academic self-concept was shown to be a strong predictor of persistence in undergraduate math programs (House 1995) and final grades in math courses (Gerardi 1990; Wilhite 1990; Astin 1993; House 1995). Interestingly, House (1995) found that academic self-concept specific to mathematical ability was a stronger predictor of final grade than any cognitive factors measured (including ACT scores), and that this academic self-concept was a stronger predictor of final grade for females than for males. But although predictive models are being deployed, details of implementation including the types of models used and ways to assess predictive power are difficult to find in the literature. Some reports are using percentage accuracies of competing models to select the optimal one (but no mention of any validation set).

Just as there have been several studies to determine factors that can predict student performance in particularly difficult courses, there have been studies also on intervention techniques that have been carried out with the purpose of improving academic success. Morrison and Schmit (2010) and Morrison (2012) tried to predict probabilities of success in a gateway Mathematics course at North Iowa Area Community College based on thresholds imposed on students' ACT Math scores and high school GPA. University of Maryland at Baltimore County (UMBC) has been implementing a "First Year Intervention" (FYI, for short) system to alert students who are in danger of failing a course based on performance on a screening quiz and first midterm (Baradwaj et al. 2012).

Moreover, with the advance of business intelligence software in the last few years, we have the emergence of a field called "learning analytics" in which sophisticated analytic tools are used to improve learning and education (Elias 2011). It uses tools and methods from business intelligence, web analytics, statistics, and data mining to analyze large repositories of data available at colleges and universities, mostly through integration with their Learning Management Systems (LMS). Such data may include the number of times lecture notes are viewed, quiz grades, and number of contributions to a discussion forum, and provide the instructor with a digital view of student performance and progress in real time (Educause 2010), commonly in the form of "dashboards" that are comprised of data visualization tools like graphs, charts, and tables. This helps the instructor to track and determine which students may need help and when so that he or she may reach out to those students via email as a way of intervention. This also helps the development of personalized learning environments to refine course offerings for users. Although dashboard technology is growing in popularity in education applications, there are associated challenges as well, such as providing the right information accurately and in a timely manner in order to be useful (Baker 2007). According to MacFayden and Dawson (2010), predictive models for monitoring student achievement should be built at the course level, which is consistent with prior studies that factors affecting student success vary significantly across courses and disciplines. Recently, an article in the Guardian (Ferguson 2014) discussed how analytics could help shape students' progress that shows that this topic is gaining popularity even in the media.

In our research study, we seek to obtain data on a wide range of cognitive and noncognitive variables consistent with prior studies that have been historically demonstrated to influence student performance in various courses. Moreover, research (Sellers et al. 2007) shows that active, collaborative learning approaches are more inclusive of students who come from backgrounds traditionally not well represented among those who are in STEM fields, such as Hispanic and Native American, low-income, and/or first-generation students. Since our university serves minority populations (Hispanics, Asians, and Pacific Islanders) and people with lower income status in the neighboring communities, its student population is very unique. Hence, as previous studies have pointed out, individual factors pertinent to our student population need to be identified in order to build effective intervention models that will be particularly beneficial to them. To the best of our knowledge, such models have not been previously deployed in bottleneck courses in Business, particularly for such a diverse student population as ours.

3. Data and research goals

In this article, we examine students' performance in a course on Business Statistics, one of the gateway courses in the College of Business at a certain public university in California during the academic year 2012–2013. The data used in the study consisted of grades (both midterm and final) and a host of external variables collected via an online survey. Data on academic background was obtained from the campus office of institutional research. These variables are categorized into five broad classes, namely:

i *Demographic factors:* age, gender, income, ethnicity, etc.
ii *Academic history and record-related factors*: GPAs—both high school and current college-level, pre-requisite course grades, scores on standardized tests like SAT, ACT, etc.
iii *Work-related factors*: whether they work part-time or full-time outside of school, how many hours they work, etc.
iv *Course-related factors*: number of attempts on the course, how many hours devoted to this course outside of lectures, how often they attend lectures, etc.
v *Academic self-concept factors*: general motivation, interest, and self-confidence in quantitative courses, level of preparedness at the start of the course, etc.

The online survey was administered to students in class via an online link posted on the course website. The survey was completely voluntary and no incentives were provided to encourage participation. Out of a total of almost 1,300 students enrolled in the Business Statistics course in Fall 2012, 796 completed the survey (response rate: 63%). The response rates are determined by the fact that they had submitted the survey in the final step. There were some missing data as students had the option of skipping questions but they were removed from the analysis (less than 10%). But note here that academic data were available for all students as they were obtained directly from campus offices, as mentioned above.

Our primary research goals in this article are enumerated below:

- Identify a subset of factors that contributes significantly to student success in this course among all those that are included in the study in order to build statistical models to predict student success.
- Design an effective multi-stage intervention method based on the predictive models to implement in the course by providing additional support to students who are at risk of failing the course.
- Assess and evaluate the intervention techniques, upon implementation.

4. Research methods

Our data obtained from survey responses yield categorical variables (like ethnicity, work status, etc.), as well as numerical ones (like GPA, SAT scores, number of units of classes taken, etc.). We use various descriptive statistics and statistical inference techniques (such as t-tests and analysis of variance, ANOVA) to identify variables that have statistically significant effects on student performance in this course. These exploratory univariate analyses will help us gain an understanding of our data so as to utilize the insight obtained into building suitable predictive models for intervention.

The prediction of the specific GPA for a student at the end of the course is difficult in terms of accuracy, and is a well-established finding in the literature over many years (Goldman and Slaughter 1976). On the other hand, the estimation of probability of success for any given student has been found to be more productive and useful. In our analysis, we thus use logistic regression (Hosmer and Lemeshow 2000) to assess the chances of students' success in this course by determining a subset of influential factors. In our binary logistic regression model, the outcome or the dependent variable (denoted by Y) represents failing the course (coded as "1") so that the other class denotes success (coded as "0"). These definitions are driven by the goal of the intervention models—detect students who are at risk of failing. Failure in the course is determined by a letter grade below C. Classification of a new observation is performed using a suitable cut-off value on the estimated probability of failure determined from the Receiver Operating Characteristic (ROC) curve. Students predicted to fail are identified as being at risk for potential intervention during the course. We tested the models on different subgroups of students by choosing a training set and a validation set. The models are fitted and coefficients estimated using the training set and tested on the validation set in each case. The models' performance is assessed using traditional metrics like "sensitivity" and "specificity." The final error rates are obtained by averaging over those from the different repetitions.

5. Analyses and results

First, we check whether there is any ground to motivate early detection and intervention in the course of interest. The course on Business Statistics is considered "bottleneck" in our university; hence the average

Table 1. Distribution of students by ethnicity for the course ($n = 795$).

Ethnicity	Frequency
American Indian/Alaskan native	2 (0.25%)
Asian/Pacific Islander	301 (38%)
Black or African American	29 (4%)
Hispanic	205 (26%)
White or Caucasian	246 (31%)
Missing responses	12 (1.5%)

failure rates (grades D and lower) are typically around 20%–22%. This is considerably high so that efforts to improve student success are required. Moreover, upon analyzing midterm and final grades of students in our dataset, we found that 70% of the students who received a grade of D or lower in the first midterm ultimately failed the course. The Pearson correlation coefficient of midterm grades with final grades is also strongly positive (0.78, significant with p-value < 0.0001). This implies that students scoring low on the first midterm typically have poor performance in the overall course (and as a result have higher probability of failing) as compared to those with high midterm scores. Thus, it is possible to detect majority of the students who are likely to fail early on in the course for the purpose of intervention.

Apart from midterm grades, we next look at a host of other background variables that may also help us detect students who are likely to have difficulty in being successful in the course on Business Statistics, as part of our exploratory analysis. Note that, here we only look at the different factors individually, and the significant factors that predict student success will be determined from the logistic regression model.

Demographic factors: The main demographic variable that had an effect on performance was ethnicity. Table 1 shows the ethnic group composition of students who responded to the survey. As is clearly evident, the course is dominated by Asian/Pacific Islanders followed by Caucasians and Hispanics, a fairly accurate representation of the composition of our student population in the entire university.

Statistically significant differences were observed in both midterm and final grades for students belonging to different ethnic groups (p-values: 0.0203 and 0.0405, respectively). In particular, Caucasians and Asians performed better than African Americans and Hispanics. Furthermore, students who were first generation in their family to attend college had significantly lower grades than the other students (p-value < 0.0001). Note here that the course had 27% first generation college attendees enrolled in that particular semester.

5.1. *Academic record and history*

Of all the factors that reflected the students' academic record and history, the most influential factor was college GPA. A very high positive correlation was observed between GPA and grades (midterm: 0.77 and final: 0.81), statistically significant as well. Other academic variables like high school GPA, SAT scores, number of hours devoted to the course, number of units of courses enrolled for, etc., did not have any significant relationship with student performance in this course. Moreover, students who repeat this course two to three times have significantly lower midterm and final grades than those who are taking it for the first time (p-values: < 0.0001).

The pre-requisite for the course on Business Statistics is a course on Business Calculus. Students with lower grades in the pre-requisite course are seen to be at a greater risk of failing the course. Hence, students who are already known to have performed poorly in the Business Calculus course are expected to have difficulty in the course, thus providing us with a way to identify at risk even before the start of the course on Business Statistics. Thus, our conclusion is that there is a very strong positive association of student performance and success in this course with that in the pre-requisite course. This pattern is clearly visible in Table 2. Note that we did not have Business Calculus grades for many students, especially those who transferred from another institution.

Thus, we found that many factors that constitute a student's academic record have significant effect on their performance in this course, which is consistent with expectations and prior studies described earlier. This leads to the conclusion that students with stronger

Table 2. Distribution of student grades in the Business Statistics course according to letter grades obtained in the pre-requisite course on Business Calculus ($n = 300$).

Letter grade in Business Calculus	Frequency	Average midterm grades in Business Statistics	Average final grades in Business Statistics
A (>90)	48 (16%)	86%	84%
B (80–89)	85 (28%)	78%	76%
C (70–79)	128 (43%)	76%	72%
D (60–69)	19 (6%)	76%	69%
F (<60)	20 (7%)	69%	64%

Table 3. Distribution of students by their work status outside of school and the average grades ($n = 795$).

Work status	Frequency	Average midterm grades	Average final grades
Full-time	152 (19%)	76.19%	72.94%
Part-time	401 (50%)	78.61%	75.54%
No work outside school	216 (27%)	79.82%	77%
Missing/No answer	26 (3%)	—	—

academic backgrounds are likely to have significantly better success in this course.

5.2. *Work-related factors*

Most of the students in the College of Business work outside of school, and the goal here is to determine how their work activities affect performance in this course on Business Statistics. Table 3 shows the distribution of students by their work status outside of school, along with average midterm and final grades. As we can see, majority of students work part time. Statistical tests reveal significant differences in both sets of grades for these students grouped by their work status (p-values, respectively, 0.049 and 0.019). Students who work less outside have significantly better performance than those who work more. This seems reasonable since the more time students spend on their work activities, the less time they spend on the course leading to poorer performance.

Moreover, 40% of students stated that external responsibilities always or often affect their success at school whereas approximately 40% said that they occasionally do so. The largest proportion of students mentioned work and financial situation as the most dominant factor, followed by interest and motivation in class and family obligations. Social/recreational activities, health-related factors, and athletics were not found to be the most influential factor in academic success according to students' self-perceived beliefs.

5.3. *Academic self-concept factors*

As in all educational studies, academic self-concept factors are very important in assessing student performance and success in major courses. They represent students' self-perceived notions about certain factors that affect learning. So we included in our survey some such factors that were aimed at understanding students' own perception about motivation, interest, intellectual ability in topics of a quantitative nature, and their self-perceived level of preparedness at the beginning of the course. Statistical tests indicated that those students who were more motivated, had more interest in Mathematics, had more self-confidence in their intellectual ability in Mathematics, and believed that they were better prepared before the course had significantly better performance overall (p-values close to 0 in all cases). 75% of the students were motivated to some extent, while 35% were slightly or not interested in Mathematical topics. Moreover, only about 6% of the students had low self-confidence in their intellectual ability in Mathematics and 52% believed that they were somewhat prepared for the course but were lacking some important skills and knowledge.

These findings are consistent with earlier studies and expectations (Wilhite 1990; House 1995). Students who have higher academic self-concept are expected to have more success in courses than students with relatively lower academic self-concept. Thus, efforts are required to help improve students' academic self-concept via novel and innovative pedagogical methods to increase motivation and interest, not only for the Business courses but even earlier for the pre-requisite courses that prepare them for these courses. These, along with intervention techniques, will be helpful in enhancing the overall intellectual and cognitive abilities of students, thus improving their chances of succeeding in the course on Business Statistics.

6. Prediction models

Although several factors were found to affect student success in the Business Statistics course from the exploratory analyses outlined above, the goal now is to identify those that are useful for predicting whether a student will ultimately succeed in this course while controlling for the others. After fitting several models using different subsets of the variables included in our study, the variables that yielded the best models of students' success in terms of prediction accuracy are college GPA, pre-requisite course grade, and the midterm grade. Note that these data are obtained from the campus and are hence reliable. Based on these, two models are proposed:

1. *"Early intervention" model*—this model produces the probability of failure in the course based on predictors generated prior to the start of the semester, but after the course pre-requisite

Table 4. Model output for the early intervention model.

	Coefficient	Standard error	*p*-Value
MATH	− 0.936	0.3606	0.0094
GPA	− 1.3912	0.3168	0
Intercept	2.3788	0.7754	0.0022

was completed. We use this model to identify a group of students at-risk at the beginning of the course. Of the 647 complete records available (recall, that we did not have pre-requisite course grades for many students), we use 400 for training (or, fitting the model). The rest 247 records form the validation set for assessing predictive accuracy.

2. *"Late intervention" model*—this model produces the probability of failure in the course based on a combination of predictors generated before the start of the course in question and after the first midterm is given in that course. For this model, we used 751 records for training the model and the rest 482 for validation purposes (total sample size = 1,233).

6.1 *Early intervention model*

The two predictors used in this model are the letter grade in the pre-requisite course on business calculus that we refer to as "Math" (grades are divided into two categories: B and above vs. below B because only letter grades were available) and the cumulative GPA prior to enrolling in the Business Statistics course. The model is

$$\text{Probability} = \frac{\exp(2.3788 - 0.936\text{Math} - 1.3912\text{GPA})}{1+\exp(2.3788 - 0.936\text{Math} - 1.3912\text{GPA})}.$$

The model output along with fitted coefficients, *p*-values are shown in Table 4.

Figure 1 shows that the "Math" variable expresses the difference in performance in a quantitative course between students with quantitative skills and those without quantitative skills. Note that, the difference in the probability of failing the course between students who scored high in Math and those who did not reduces when the GPA increases. A list of other competing models is considered along with the Akaike's Information Criterion (or AIC) values appear in the Appendix. Our chosen model has the lowest AIC among all of these.

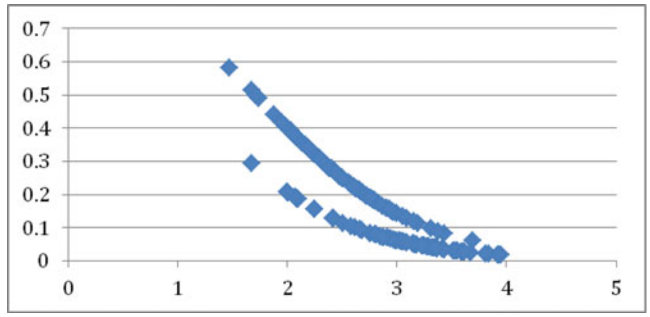


Figure 1. The relationship between GPA and the probability of failing the course for two different categories of the course grade on Business Calculus. The upper line belongs to students with prerequisite grade < B, while the lower line belongs to those with prerequisite grade ≥ B.

Table 5. Model output for the late intervention model.

	Coefficient	Standard error	*p*-Value
MATH	−0.0732	0.0072	0.0000
GPA	−0.6742	0.1981	0.0007
Intercept	5.8368	0.7033	0.0000

6.2 *Late intervention model*

The two predictors for this model were midterm grade (percentage score) and GPA, chosen again using the AIC criterion. The model is

$$\text{Probability} = \frac{\exp(5.8368 - 0.0735\text{midterm} - 1.3912\text{GPA})}{1+\exp(5.8368 - 0.0735\text{midterm} - 1.3912\text{GPA})}.$$

The model output along with fitted coefficients, *p*-values are shown in Table 5. Other competing models that we explored, along with the AIC values, are included in the Appendix.

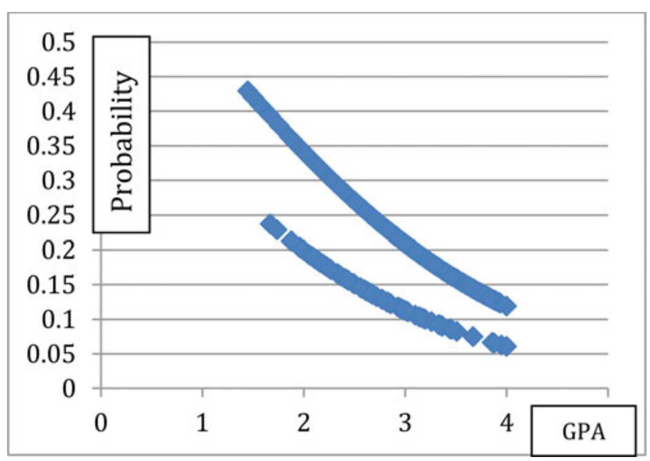


Figure 2. The role of GPA in determining the probability estimate (*y*-axis). The upper line belongs to students with midterm grade < B, while the lower line belongs to those with midterm grade ≥ B.

In Fig. 2, we demonstrate the behavior of the probability of failure with respect to GPA for different scores in the midterm. Note that the midterm scores are more consequential in predicting failure in the course for lower levels of GPA, thus indicating that those students with lower midterm grades are at a greater risk of failing in the course if they also have low GPAs.

An important observation from all the models is that none of the background information like ethnicity or work-related factors that were found to have a significant effect in the univariate analyses earlier was found to have a significant effect on the target outcome in presence of the academic predictors (GPA, midterm, and pre-requisite grades). This does not necessarily imply that the background factors are not relevant to student success, but that their effects are not as significant in the presence of the stronger academic record-based factors.

6.3. *Assessing predictive accuracy of the models*

The model accuracy or predictive power is assessed via performance on the validation sets (treated as "new data" at each stage) measured by specificity and sensitivity as mentioned earlier. The data partition is done via random selection, and is repeated 25 times in order to remove selection bias. Specificity was calculated as the percentage of the number of students who were predicted to pass among those who indeed passed the course. Sensitivity was calculated as the percentage of the actual number of students who failed the course that were predicted to fail. If the predicted probability of failure for a particular student was greater than a pre-determined cut-off probability, he or she was determined to be "at risk of failing the course." The latter was selected by observing the ROC curve (which plots "sensitivity" vs. 1− "specificity") shown in Fig. 3.

The graphs demonstrate the trade-off between sensitivity and specificity for different values of the cut-off probability. For early intervention, we focus on reaching out only to those students who are at a high risk of failing the course (this is because, many students get confused by a letter early on in the course before the first exam). Thus, we consider "specificity" a more important criterion than "sensitivity" to select the optimal cut-off probability at this stage. This situation is reversed for late detection because at that stage our objective is to intervene upon as many students as possible and not miss anybody who might have benefitted from it. Therefore, we consider "sensitivity" a more important criterion, thus resulting in a low cut-off probability that yields a lower specificity value. However, in both cases we need to ensure that the other metric also does not fall below a certain value. Such choices are further supported by the fact that our models predict success in the course better than failure (possibly because the data are dominated by success, that is, more students pass than fail the course).

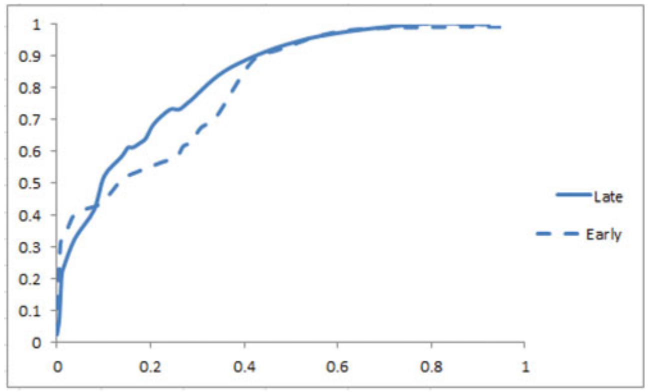


Figure 3. ROC curves for the early and the late Intervention models.

The average sensitivity and specificity values for each model are calculated by averaging over the validation sets from 25 repetitions using the chosen cut-off probability, along with the estimated standard errors. This information is summarized in Table 6. The standard errors help conclude that these accuracy metrics are quite robust, that is, they are not much affected by specific training and test set combinations.

The late detection model is more accurate than the early detection one because it incorporates the midterm grades (which being part of the course itself is expected to be a stronger predictor of final scores than the pre-requisite course grade that is included in the early intervention model). The early detection model, despite not using any information from the course itself, is quite effective in predicting the success rates. This is very crucial for intervention purposes because

Table 6. Prediction performance of the models in the course. The entries in this table represent the "average accuracy rate" ± 2 standard errors.

Models	Sensitivity	Specificity	Cut-off probability
Early intervention	60% ± 3.4%	73% ± 1.7%	0.25
Late intervention	75% ± 5.1%	76% ± 0.3%	0.2

the earlier you identify at-risk students and provide them with help the better are their chances of succeeding eventually in the course.

We also compared the predictive performance of our models with other commonly used classification methods, such as Classification Trees and Neural Networks in order to establish their relative efficiency for our purpose. Using the same cross-validation techniques, we found that the same subset of predictors came out to be influential (i.e., GPA and Math grades for early intervention and GPA and midterm grades for the late intervention) and the prediction accuracies of these models were significantly poorer than that of the logistic regression models that we employed. Both the sensitivity ratio and the specificity ratio were quite low, the range being 45%–52%. This clearly indicates that a large proportion of students who failed the course would not receive the additional help early on through intervention that they needed, should any of these two models were deployed in practice. Logistic regression thus proved to be the superior method compared to both these tools, with significantly higher predictive power.

However, the relatively newer classification technique of Random Forests (Breiman 2001), which considers an ensemble of classification trees to improve accuracy, yielded considerably better results that were at par with those we obtained from the logistic regression, with sensitivity values of 62% and 72% for early and late intervention models, respectively, and specificity values of 75% and 76% for the two models, respectively. Note that these values are close and not significantly different from those of our models. So we conclude that logistic regression does an overall good job of predicting student failure and we adopt it for our intervention studies. We wish to explore other classification methods such as boosted trees, in our future work to determine whether further improvement in prediction accuracy can be achieved.

6.4. *Devising intervention based on the predictive models*

Our overall recommendations based on these predictive models consist of a two-stage intervention plan for students in the Business Statistics course. This is done as follows:

- **Stage 1:** The early detection model is used to identify at-risk students based on their academic record before the start of the course. Actions are then taken by the school to approach these students and urge them to receive support made available by the school.
- **Stage 2:** The late detection model is used to identify at-risk students after the first midterm. These students are also encouraged to avail resources to help them with the course. Note that, those students selected at this stage who were already identified in Stage 1 need to be provided with greater additional support in order to improve their odds of passing.

The implementation of these intervention models is based on the probability estimate of failure obtained from the logistic regression model selected at each stage, say $P(F)$. With the cut-off probability selected before (say, T), we classify a student as “failing” if $P(F) \geq T$ or “passing” if $P(F) < T$.

During each semester, $P(F)$ is computed for each student based on the relevant variables included in each of the models—GPA and pre-requisite course grade for early detection, and GPA and midterm grades for late detection. Using a threshold as described above, we identify subgroups of students who are at risk, an example being:

Early detection:

1. Students with a letter grade of below B in the pre-requisite course and GPA lower than 2.72.
2. Students with a letter grade of A or B in Business Calculus but a GPA lower than 2.49.

Late detection:

1. Students with midterm grade below 72% and GPA lower than 2.67.
2. Students with midterm grade above 72% but a GPA lower than 2.45.

7. Implementation of the intervention methods

We deployed the two-stage intervention model for the Business Statistics course in Spring 2014, wherein students detected to be at risk at each stage received a letter from the college dean asking them to seek additional help and support for the course. The particular intervention consisted of *supplemental instruction (SI),* which is an academic assistance program that utilizes peer-assisted study sessions through hands-on practice and collaborative group exercises. These are known to increase retention and improve student grades in

Table 7. Performance of students detected for early intervention broken down by whether they attended SI sessions or not.

Attend SI	Failure rates	Midterm grades (mean)	Final grades (mean)
Yes ($n = 27$)	2 (7.4%)	74.86	77.61
No ($n = 35$)	8 (22.9%)	75.47	72.83

Table 8. Performance of students detected for late intervention broken down by whether they attended SI sessions or not.

Attend SI	Failure rates	Midterm grades (mean)	Final grades (mean)
Yes ($n = 67$)	14 (20.9%)	63.51	71.65
No ($n = 297$)	135 (45.5%)	58.63	58.77

targeted historically difficult courses (UMKC website, http://www.umkc.edu/ASM/si/overview.shtml).

Stage 1: Early intervention

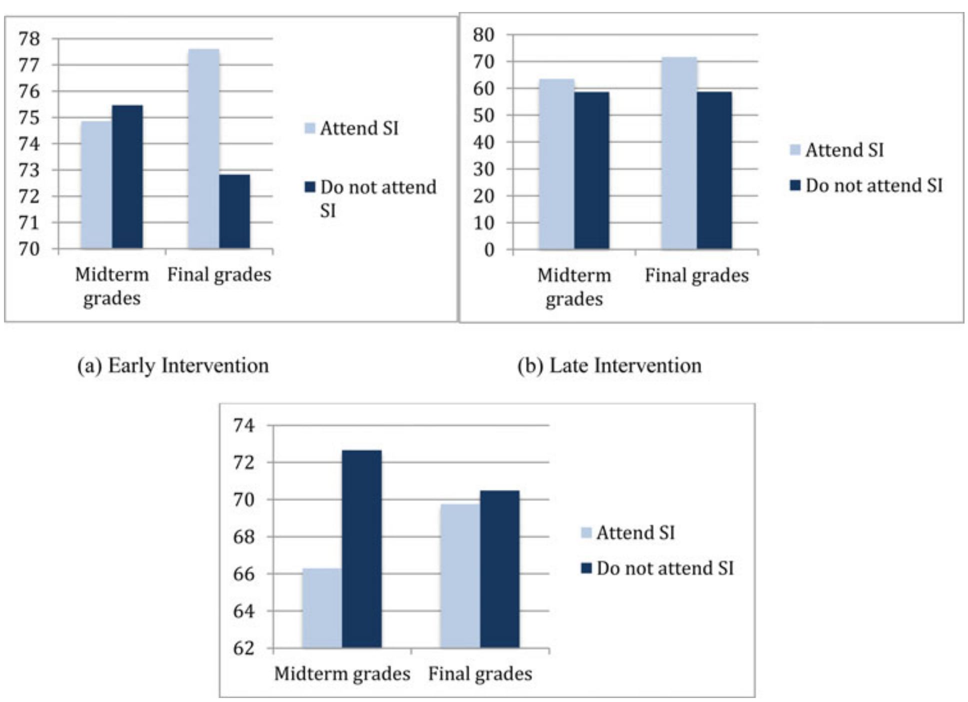
Of the 62 students included in the early detection phase, 10 failed (16.13%). Table 7 shows the failure rates among these students, broken down by those who attended SI sessions and those who did not. Clearly, there is a significantly lower percentage of students who were intervened upon and took advantage of it (by attending SI sessions), failed compared to those did not, and a p-value of 0.0375 was observed when testing the difference in the two proportions of failure rates. The average midterm and final grades also clearly indicate that performance of students improved significantly if they attended SI sessions after the intervention and it deteriorated if they did not. The average final grade of students who attended SI is significantly higher as well (p-value $= 0.035$).

Stage 2: Late intervention

Of the 364 students who received the late intervention letter (after midterm), 149 failed with D/F/W grades (40.93%) and 209 passed (57.4%). The failure rates in Table 8 indicate that a significantly lower percentage of students who took advantage of the intervention (by attending SI sessions) failed compared to those did not (p-value < 0.0001), just as in the case of the early intervention. The average midterm and final grades also demonstrate that performance of students improved significantly if they attended SI sessions after the intervention, and it did not otherwise. There is a statistically significant difference in the final grades of these two groups (p-value $<$ 0.0001).

7.1. *Both early and late interventions*

Let us now look at those students who were selected in both stages of intervention. Clearly these students are in greater need for additional help in order to succeed in the course. There were 33 students who were chosen for both interventions. Of these 33, 17 (51.5%) attended SI sessions and the rest 16 (48.5%) did not. The failure



(a) Early Intervention

(b) Late Intervention

Figure 4. Average midterm and final grades of students (shown on the y-axes) who received interventions at the different stages.

Table 9. Performance of students detected for both early and late interventions broken down by whether they attended SI sessions or not.

Attend SI	Failure rates	Midterm grades (mean)	Final grades (mean)
Yes ($n = 17$)	2 (11.8%)	66.29	69.75
No ($n = 16$)	7 (43.75%)	72.65	70.48

rates in Table 9 indicate that a significantly higher proportion of students who did not avail of the interventions failed the course (p-value $= 0.0146$).

Table 9 demonstrates the performance of these students in the course, both in terms of midterm and final grades. ANOVA revealed significant differences between the midterm grades (p-value $= 0.039$), thus suggesting that students who received both interventions had poorer performance up to that point than students who did not (as expected). On the other hand, there is no statistically significant difference in the final grades of students who received the interventions and those who did not (p-value $= 0.857$). In fact, average grades of the students who were intervened upon twice improved from the midterm to the final whereas for the rest of the students, the grades deteriorated. The typical trend in this course is that performance deteriorates from the midterm to the final (as discussed earlier).

The results from the different stages of interventions are summarized in Fig. 4 for easy comparison of the outcomes. It is easily seen that the biggest impact is seen on the final grades after attending a significant number of SI sessions, and for those students who were selected for the early intervention. Note that the latter students are the ones with the weakest background as determined by their GPA and pre-requisite course grade before they took the Business Statistics course.

8. Conclusions

This study was the first of its kind in the College of Business at this university wherein the focus was to develop an understanding of the factors that affect student success in two bottleneck core courses with the objective of devising effective intervention. It was a large-scale study involving over 1,000 students; hence the results are very promising. We identified several factors associated with student performance, as well as devise two rigorous statistical models to be used for early detection and intervention in this course. Our intervention implementation also proved to be a major success as our results clearly indicated that a significantly lower number of students who availed supplemental instruction failed the course than those who did not. This suggests that timely and effective intervention techniques have the potential to improve student success rates in difficult courses. Development of these intervention strategies based on the predictive models is thus the major contribution of this work. We note here that the factors studied in this project cover a wide range of information about the students and are quite general across disciplines. Thus although developed for a business course, our methods may be easily adapted in building similar intervention for any course in any field and at any institution of higher education.

Moreover, our university particularly has a large minority population dominated by Hispanics, many of who are first generation college attendees. Historical data have shown that many of these students have difficulty in quantitative courses in the area of Business, and hence can benefit greatly from the additional support offered in the form of intervention, both before and during the course. Steps can be taken to develop special curriculum with preparatory courses in quantitative disciplines so that students graduating from high schools could acquire the skills necessary to be successful in college. This considerably broadens the scope of our research. All these insights, along with the proposed intervention tools have the potential of improving student success in bottleneck courses in the Business curricula, and as a result, have far-reaching impact of improving graduation rates at colleges and universities, specifically for under-represented minority groups.

Another aspect is that since most of the factors affecting student performance are usually identifiable before the course starts, they can aid the administration in the respective departments in the universities with logistical planning and decision-making in the form of allocating resources for supplemental instruction (for instance, how many sessions to offer in a particular semester?), tutoring, and other student-centered services (tutors, space, equipment, etc.). For example, based on GPA and pre-requisite course grades, they can estimate ahead of time how many students enrolled in the course in a particular semester are likely to need extra help and support. In addition, making successful intervention should shorten the average time till graduation (because the course of interest is a bottleneck course), and this has an impact on cost, seat occupancy, faculty recruitment, and other measures of program efficiency. This study and the results obtained, therefore, have extensive ramifications in terms of budgetary and resource management as well. We also wish to integrate a cost analysis into our modeling for assessing predictive power, provided we are able to obtain the relevant data from campus offices.

We are currently refining our models for even better outcomes in the future semesters, as well as studying those students who were selected for interventions more closely with respect to the background factors. Since our predictive models that are the main goals of the study only required academic factors, we do not conduct the survey any more, but instead obtain all our background data from the campus office of Institutional research. This has considerably streamlined the entire process and ensured that the data received are trustworthy so that the accuracy and validity of our methods can be established. In the future, we also hope to obtain some background data such as transfer status from a Community College (Yes/No), number of units of courses enrolled in, etc., from the campus in order to incorporate them into our intervention models for potentially greater accuracy is detecting students at risk of failing.

References

Astin, A. W. **1993**. An Empirical Typology of College Students. *Journal of College Student Development* 34:36–46.

Bailey, T. **2008**. "Transition Matters: Community College to Bachelor's Degree." Proceedings Report of the Advisory Committee on Student Financial Assistance. pp. 46–49. https://protect-us.mimecast.com/s/rx4mBRUDx1VgCW

Baker, B. **2007**. "A Conceptual Framework for Making Knowledge Actionable Through Capital Formation." D. Mgt. diss., University of Maryland University College.

Baradwaj, R., K.Baron., C. Bichy, M. Dilton, J. Fritz, and J. Joseph. **2012**. *Using Learner Analytics to Scale Feedback in Gateway Courses*, Baltimore, MD: University of Maryland, Baltimore County (UMBC) Accessed April 9, 2015. http://www.jngi.org/wordpress/wp-content/uploads/2013/04/Using-Learner-Analytics-to-Scale-Feedback-in-Gateway - Courses-Bichy8.pdf.

Bellico, R. **1972**. Prediction of Undergraduate Achievement in Economics. *Journal of Economic Education* 3(1):54–55.

Borde, S. F. **1998**. Predictors of Student Academic Performance in the Introductory Marketing Course. *Journal of Education for Business* 73(5):302–306.

Breiman, L. **2001**. Random Forests. *Machine Learning* 45:5–32.

Brower, A. M., and A. Ketterhagen. **2004**. Is There an Inherent Mismatch Between how Black and White Students Expect to Succeed in College and What Their Colleges Expect from Them? *Journal of Social Issues* 60(1):95–116.

Brown, R. O. **1966**. Prediction of Accounting Grades. *The Accounting Review* 41(2):340–343.

Cohn, E. **1972**. Students' Characteristics and Performance in Economic Statistics. *Journal of Economic Education* 3(2):106–111.

Congress of the United States, Office of technology Assessment. **1988**. *Educating Scientists and Engineers: Grade School to Grad School*. Washington DC: US Government Printing Office, pp. 128.

Doran, B. M., M. L.Boullion, and C. G. Smith. **1991**. Determinants of Student Performance in Accounting Principles I and II. *Issues in Accounting Education* 6(1):74–84.

Eckel, N., and W. A. Johnson. **1983**. A Model for Screening and Classifying Potential Accounting Majors. *Journal of Accounting Education* 1(2):57–65.

Educause **2010**. "7 Things You Should Know About Analytics, EDUCAUSE 7 Things You Should Know Series." Accessed April , **2015**. http://www.educause.edu/ir/library/pdf/ELI7059.pdf

Elias, T. **2011**. *Learning Analytics: Definitions, Processes and Potential*. Edmonton, AB, Canada: Athabasca University. Accessed April 9, 2015. http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf

Eskew, R. K., and R. H. Faley. **1988**. Some Determinants of Student Performance in the First College-level Financial Accounting Course. *The Accounting Review* 63(1):137–147.

Ferguson, R. **2014**. Learning Analytics don't just Measure Students' Progress – They Can Shape it. *The Guardian*, 24 March 2014. Retrieved April , **2015**. http://www.theguardian.com/education/2014/mar/26/learning-analytics-student-progress

Gainen, J. **1995**. Barriers to Success in Quantitative Gatekeeper Courses. In *Fostering Student Success in Quantitative Gateway Courses*, edited by J. Gainen and E. W. Williamson, 5–14. San Francisco: Jossey-Bass.

Gerardi, S. **1990**. Academic Self-Concept as a Predictor of Success Among Minority and Low Socio-economic Status Students. *Journal of College Student Development* 31:402–407.

Glesne, C. **1998**. *Becoming Qualitative Researchers: An Introduction*. New York: Longman.

Goldman, R. D., and R. E. Slaughter. **1976**. Why College Grade Point Average is Difficult to Predict. *Journal of Educational Psychology* 68(1):9–14.

Herndon, M. K., and J. L. Moore. **2002**. African American Factors for Student Success: Implications for Families and Counselors. *The Family Journal: Counseling and Therapy for Couples and Families*. 10(3):322–327.

Horn, L., K. Peter, and K. Rooney. **2002**. *Profile of Undergraduates in U.S. Postsecondary Institutions: 1999–2000 (NCES 2002–168)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics, U.S. Government Printing Office.

Horn, L. J., and M. D. Premo. **1995**. *Profile of Undergraduates in U.S. Postsecondary Education Institutions: 1992–93,* *With an Essay on Undergraduates at Risk (NCES 96–237).* Washington, DC: U.S. Department of Education, National Center for Education Statistics, U.S. Government Printing Office.

Hosmer, D. W., and S. Lemeshow. **2000**. Applied Logistic Regression. *Wiley Series in Statistics and Probability*. Hoboken, NJ: Wiley.

House, J. D. **1995**. Non-cognitive predictors of achievement in introductory college Mathematics. *Journal of College Student Development* 36(2):171–181.

Ingram, R. W., and R. J. Peterson. **1987**. An Evaluation of AICPA Tests for Predicting the Performance of Accounting Majors. *The Accounting Review* 62(1):215–223.

Liesz, T. J., and Reyes, M. G. C. **1989**. The use of Piagetian Concepts to Enhance Student Performance in the Introductory Finance Course. *Journal of Finance education* 18:8–14.

MacFayden, L. P, and S. Dawson. **2010**. Mining LMS Data to Develop an Early Warning System for Educators: A Proof of Concept. *Computers and Education* 54(2):588–599.

Meece, J. L., J. E. Parsons, C. M. Kaczala, S. B. Goff, and R. Futterman. **1982**. Sex Differences in Math Achievement. *Journal of College Science Teaching* 28(1):307–310.

Morrison, M. **2012**. "Improving Probabilities for Success in a Gateway Math Course, Decisions Based on Evidence." Accessed November 20, 2014. http://www.decisionsonevidence.com/2012/04/improving-probabilities-for-success-in-a-gateway-math-course/

Morrison, M. C., and S. Schmit. **2010**. *Predicting Success in a Gateway Mathematics Course*. North Iowa Area Community College. Accessed from Institute of Education Sciences (IES) ERIC database. http://eric.ed.gov/?id=ED511033

Sachdeva, K. S., and W. E. Sterk. **1982**. Projecting Finance Student Final Course Scores Based on Initial Exam Scores. *Journal of Financial Education* 11:55–60.

Sedlacek, W. E. **2002**. "Employing Non-cognitive Variables in the Admission and Retention of Non-traditional Students." Unpublished Technical Report, Counseling Center, University of Maryland at College Park.

Sellers, S. L., J. Roberts., L. Giovanetto, K. Friedrich, and C. Hammargren. **2007**. *Reaching All Students, a Resource for Teaching in Science, Technology, Engineering, and Mathematics*.2nd ed. Madison: University of Wisconsin Center for the Integration of Research, Teaching, and Learning.

Seymour, E., and N. Hewitt. **1997**. *Talking About Leaving: Why Undergraduates Leave the Sciences*. Boulder, CO: Westview.

Tobias, S. **1990**. *They're not Dumb, They're Different: Stalking the Second Tier*. Tucson, AZ: Research Corporation.

Wilhite, S. C. **1990**. Self-efficacy, Locus of Control, Self-assessment of Memory Ability, and Study Activities as Predictors of College Course Achievement. *Journal of Educational Psychology* 82:696–700.

Appendix: Selecting the best predictive model

While looking for the best predicting model, we used many subsets of the available variables. In the partial list that follows, we summarize some of the models attempted when finding the best predictive models. Tables A1 and A2 show some of these models for the early and late intervention stages, respectively.

One can observe the "strength" of the variables GPA and MATH for the early detection model. Once these two variables are included in the regression, the rest of the variables become insignificant, that is, they do not add anything to the information we already have. We pick model 8, which gives the best *p*-values for the variables, and the lowest AIC value.

A similar process was followed for model selection in case of the late intervention where we considered several variables as well as variations of the midterm grade and the GPA variables, shown in Table A2. Again, the model with GPA and midterm grade proved to be the best as assessed by the AIC values.

Table A1. The entries are the *p*-values, and AIC values for competing models of early intervention. Model 8 is the selected model.

	Workhrs/ week	Transfer Yes/No	Ethnicity	Units taken	GPA	Math	AIC
Model 1	0.0598	0.6836		0.0318			484.6
Model 2		0.1481	0.4903				439.7
Model 3	0.164	0.0018	0.4002				412.6
Model 4	0.0642			0.289			482.3
Model 5	0.342				0.0001		418.5
Model 6	0.2904					0.034	388.1
Model 7			0.55			0.014	397.7
Model 8*					0.0000	0.0094	362.5

Table A2. The entries are the *p*-values, and AIC values for competing models of late intervention. Model 1 is the selected model.

	Midterm	Midterm-square	GPA	1/GPA	Transfer Yes/No	Ethnicity	AIC
Model 1*	0.0000		0.0001				372.2
Model 2	0.0000			0.0026			392.4
Model 3	0.64	0.0040	0.0028				377.6
Model 4	0.5976	0.0164		0.0063			374.4
Model 5	0.0000		0.0000		0.7936		387.1
Model 6			0.0000		0.8831	0.7126	474.2