7-6-2015

Modeling Success: Using Preenrollment Data to Identify Academically At-Risk Students

Ann M. Gansemer-Topf
Iowa State University, anngt@iastate.edu

Jonathan I. Compton
Iowa State University, jcompton@iastate.edu

Darin R. Wohlgemuth
Iowa State University, darinw@iastate.edu

Gregory R. Forbes
Iowa State University, grforbes@iastate.edu

Katerina S. Ralston
Iowa State University, kroha@iastate.edu

Modeling Success: Using Preenrollment Data to Identify Academically At-Risk Students

Abstract
Improving student success and degree completion is one of the core principles of strategic enrollment management. To address this principle, institutional data were used to develop a statistical model to identify academically at-risk students. The model employs multiple linear regression techniques to predict students at risk of earning below a 2.0 grade point average (GPA) in their first semester of college. Data analysis from student cohorts starting in the Fall 2007 through Fall 2009 (N = 11,644) identified two groups of students—one predicted to earn less than a 2.0 and the other predicted to earn a 2.0 or higher. The first semester college GPA and retention rates of both groups of students were tracked to examine the accuracy of the model in predicting student success and subsequent retention rates. Multi-year analyses illustrates that the model can be used to identify students who are at risk of earning less than a 2.0 GPA. Additional analysis demonstrates there is a relationship between predicted and actual first semester GPA and retention rates. Since the data used to develop the model are commonly available at most institutions, this study provides a practical approach for the SEM research professional to identify potentially academically at-risk students, which subsequently can be used to assist students and improve student success and degree completion.

Disciplines
Education | Educational Assessment, Evaluation, and Research | Higher Education | Student Counseling and Personnel Services

Comments
This is the peer reviewed version of the following article: FULL CITE, which has been published in final form at [Link to final article using the DOI]. This article may be used for non-commercial purposes in accordance With Wiley Terms and Conditions for self-archiving

Strategic Enrollment Management Quarterly *Explore this journal >*

View issue TOC
Volume 3, Issue 2
July 2015
Pages 109–131

The Research Agenda

Modeling Success: Using Preenrollment Data to Identify Academically At-Risk Students

Ann M. Gansemer-Topf, Jonathan Compton, Darin Wohlgemuth, Greg Forbes, Ekaterina Ralston

First published:
6 July 2015 Full publication history

DOI:
10.1002/sem3.20064 View/save citation

Cited by:
0 articles Check for new citations

Abstract

Improving student success and degree completion is one of the core principles of strategic enrollment management. To address this principle, institutional data were used to develop a statistical model to identify academically at-risk students. The model employs multiple linear regression techniques to predict students at risk of earning below a 2.0 grade point average (GPA) in their first semester of college. Data analysis from student cohorts starting in the Fall 2007 through Fall 2009 (N = 11,644) identified two groups of students—one predicted to earn less than a 2.0 and the other predicted to earn a 2.0 or higher. The first semester college GPA and retention rates of both groups of students were tracked to examine the accuracy of the model in predicting student success and subsequent retention rates. Multi-year analyses illustrates that the model can be used to identify students who are at risk of earning less than a 2.0 GPA. Additional analysis demonstrates there is a relationship between predicted and actual first semester GPA and retention rates. Since the data used to develop the model are commonly available at most institutions, this study provides a practical approach for the SEM research professional to identify potentially academically at-risk students, which subsequently can be used to assist students and improve student success and degree completion.

Provide feedback or get help

Increasing retention and graduation rates continues to be a critical challenge facing institutions of higher education. From a global perspective, the 2011 report from the Organization for Economic Co-operation and Development (OECD) highlighted the United States' decline in the global share of college graduates, "slipping over 15 percentage points from 35.8% among 55–64 year olds to 20.5% among the 25–34 year-olds who have just entered the labor market" (2). Among this younger cohort of students, the United States ranks 15th among 34 OECD countries in college degree attainment which singles out the United States as "the only country where attainment levels among those just entering the labor market … do not exceed those about to leave the labor market…" (OECD **2011** , 2).

This finding is perplexing given that the decade prior to the 2011 report, college student enrollment increased 27% (NCES, **2011**). The United States ranked 9th in the percentage of college-aged students enrolling in higher education but 14th in completion rates (OECD **2011**). These data suggest that increasing degree attainment involves more than simply enrolling additional students; it involves the development of new strategies that increase the likelihood that enrolled students will graduate.

The National Student Clearinghouse reports that for the Fall 2008 entering cohort the overall 6-year completion rate was 55%, with 42.1% completing at the same institution they started and 13.0% at another institution. Additionally, for the students who enrolled full-time, 66.4% completed at the institution where they first enrolled and an additional 10.8% completed at a different institution (Shapiro et al. **2014** , 17).

Despite the nearly stagnant changes in retention and graduation rates in the past decade, pressure to improve degree attainment continues to increase (Bowen, Chingos, and McPherson **2009**). For example, the U.S. Department of Education plans to develop a federal rating system for colleges and universities that includes "outcomes, such as graduation and transfer rates, earnings of graduates, and completion of advanced degrees" (U.S. Department of Education **2015** , para. 3).

A significant amount of research has been dedicated to understanding and improving student degree attainment. Studies demonstrate that institutions can develop empirical models to identify students for interventions that improve student persistence (Herreid and Miller **2009** ; Miller and Herreid **2009**). Studies have investigated the relationship between student demographics and student academic achievement as a way to understand the potential for student success (Hirschy **2015** ; Murtaugh, Burns,

and Schuster 1999 ; Reason 2009). Since the majority of students who withdraw do so within their first year, focus has been given to investigating factors related to first-year success and programs and services that can decrease academic risk during this time (Tinto 1998).

Despite this significant body of research, college graduation rates remained relatively flat over the past decade (Bowen, Chingos, and McPherson 2009). Knapp, Kelly-Reid, and Ginder (2011) report that nationwide only 36.7% of students who start a degree program at a four-year institution graduate in four years, and 57.4% do so in six years.

While research on retention and graduation has borne useful results, the ways in which these results may be operationalized within an institutional setting may be limiting. The large body of research on retention and graduation reiterates a common theme: there are a myriad of reasons and factors that are involved in the degree attainment process, yet the compartmentalized approach to improving retention and graduation rates at many institutions seems to ignore this reality. Low-income students may be encouraged to seek assistance from financial aid. Multicultural centers provide support and assistance for students of color. Students with a lower academic profile, as measured by high school grade point average (GPA) or standardized test scores (ACT or SAT), may be directed to academic support centers. While such a single-dimension focus is important for helping specific groups of students, it can ignore the other factors that may contribute to student success across populations. Understanding how these factors (i.e., financial, academic, and demographic) are related to student success within a specific institutional context may provide new insights into the process of degree attainment.

A number of studies have demonstrated that college GPA in a student's first year has a significant impact on student retention and graduation (Bowen, Chingos, and McPherson 2009). Students with a first-term GPA below 2.0 are nearly three times more likely to withdraw after their first year (Ishitani 2008). Murtaugh, Burns, and Schuster (1999) found that students with a GPA below 2.0 in their first quarter were less likely to be retained than students with higher GPAs. Because of the positive relationship between college GPA within the first year and subsequent retention and graduation rates, focusing on factors related to first-term college GPA may provide insight into the variables that influence degree completion.

This study was designed to investigate these relationships. The study sought to develop a statistical model that examined precollege factors that contribute to student persistence. It focused on examining factors available *prior* to a student enrolling at the institution with the goal of identifying at-risk students early in the enrollment cycle so data could be provided to campus partners to proactively assist with their success. In addition, this study focused on students' first-semester college GPA and not first-year retention. By examining a combination of student attributes, identifying students prior to enrollment, and focusing on longer-term academic success, a model could be developed prior to enrollment that, when acted upon, could be useful in improving degree attainment.

Factors Contributing to Retention

Four specific categories of characteristics previously found to influence retention informed this study: academic characteristics, student demographics, financial aid variables, and program participation and involvement.

Academic Characteristics

Students with higher precollege academic performance (test scores, GPA, etc.) are more likely to persist in

college than students with lower test scores or high school GPAs. Astin, Korn, and Green's (**1987**) study of 8,000 students found that SAT scores and high school GPA were correlated with retention and graduation rates. A study by Levitz, Noel, and Richter (**1999**) a decade later illustrated a similar relationship. More recently, Bowen, Chingos, and McPherson (**2009**) also found that higher high school GPA, class rank, and test scores were positively associated with academic performance in college. Due to the importance of student's precollegiate academic characteristics in predicting academic success, this study included the variables of ACT/SAT test score; high school GPA; number of high school science, math, and English courses; and high school class rank.

Student Demographics

Research on the relationship between student demographics and retention and academic success justified the use of demographic variables in this study. In their study of 21 public institutions, Bowen, Chingos, and McPherson (**2009**) noted consistent patterns in the relationships among gender, race, and socioeconomic status and retention and graduation rates. Summarily, females have higher graduation rates than males, and White students graduate at higher rates than their non-White peers (Bowen, Chingos, and McPherson **2009** ; Jones **2015**). Additional studies supported the incorporation of other student demographic variables: age (Murtaugh, Burns, and Schuster **1999**), distance of the college from home (Ramist **1981**), and first-generation status (Ishitani **2006**).

Financial Aid Variables

The existing body of research illustrates that student financial aid is critical to a student's persistence toward graduation (Perna **1998** ; St. John **2000** ; Wohlgemuth et al. **1997**). The socioeconomic status (SES) of students and their families is positively correlated with retention and graduation rates; students with high SES backgrounds have higher graduation rates and take less time to complete their degree than students with low SES backgrounds (Bowen, Chingos, and McPherson **2009**). DesJardins, Ahlburg, and McCall (**2002**) found that different types of financial aid had varying impacts on student retention with scholarships more positively correlated with retention rates than grants or loans. This study investigated the impact of Pell eligibility, need-based aid, gift aid, and loan aid.

Program Participation and Involvement

Astin's (**1984**) theory of involvement and Tinto's (**1993**) interactionist theory argue that as students become more involved in their environment—both academically and socially—they are more likely to persist. A student's academic major, participating in a science field (Lam et al. **2005**), living in on-campus housing, participating in new students programs (Mayhew, Vanderlinden, and Kim **2010**), and joining a learning community (Braxton, Hartley, and Lyken-Segosebe **2015** ; Johnson **2000** ; Lenning and Ebbers **1999**) have been shown to positively influence student success.

Underlying themes in large scale studies such as Kuh, Kinzie, Schuh, Whitt, and Associates' (**2010**) *Student Success in College: Creating Conditions that Matter* and Pascarella and Terenzini's (**2005**) *How College Affects Students* reinforce the positive influence of college students' engagement on degree completion. Informed by this past research, this study included student engagement variables such as participation in learning communities and athletics, honors programming, on-campus living, and student orientation attendance.

Purpose of the Study

The purpose of this study was to determine if an analytic model can be developed to predict and accurately identify academically at-risk students. Academically at-risk students were defined as students who would earn less than a 2.0 GPA in their first semester in college. Since this institution, like many others, uses 2.0 GPA as a benchmark for satisfactory academic progress, this measure was an appropriate guideline as a proxy for defining academically at-risk students. Actual first-semester GPA, retention rates, and graduation rates were used to assess the effectiveness of the model in identifying academically at-risk students. The study has four research questions:

1. What is the influence of high school academic factors, demographic variables, financial variables, and program participation variables on first-semester GPA?
2. How accurately can a statistical model that uses precollege characteristics predict students who may earn above or below a 2.0 first-semester GPA?
3. Are there differences in first-semester GPAs between students who were predicted to earn less than a 2.0 and those who were predicted to earn a 2.0 or higher first-semester GPA?
4. Are there differences in retention and graduation rates between students who were predicted to earn less than a 2.0 and those who were predicted to earn a 2.0 or higher first-semester GPA?

Methods

Model Development

A statistical model to identify students who may earn less than a 2.0 in their first semester of college was created by examining U.S. domestic first-time freshmen enrolled at a large, public university from 2004 to 2006 (N = 11,912). Three years of data were utilized in developing the model to have a greater number of students and to minimize variance that could occur by only using one year of data.

The data collected for each student were based on available information one month prior to the student enrolling. Thus, for the 2004 cohort, data were from August 2004, data for the 2005 cohort were from August 2005, and data for the 2006 cohort were from August 2006. In each of the following three years (2007–2009), the model was rerun and modified, using the three most recent years, to capture any changes in practice and provide the most up-to-date model for predicting the first-term GPA of that entering cohort. This study examines six year graduation rates on the entering cohorts of 2007–2009.

Data Collection

All data for this study were collected from the university data systems. Data that were considered for the model included ethnicity, gender, first-generation status (defined here as neither parent having a four-year degree), residency status (in-state versus out of state), academic major (undecided versus not undecided), academic college, on-campus living, participation in learning communities, high ability scholarship recipient, honors program participation, student athlete status, financial aid data (including financial need, loan aid, gift aid, and Pell eligibility), precollege academic characteristics (such as ACT score and subscores; high school

rank and GPA; number of high school credits of math, science, English, and AP credits), and number of days remaining in the admission cycle when the student applied or paid the enrollment deposit. As a starting point for other institutions to develop comparable models, Figure 1 presents a set of variables that were considered in this model.

Academic Characteristics	Student Demographics
ACT or SAT Test Scores	Gender
High School GPA	Race & Ethnicity
High School Courses (Math, Science, English)	First-Generation College Student
High School Class Rank	Age
Admissions Score or Ranking	Distance from Institution
AP & College Course Work	
Financial Aid	**Program Participation & Involvement**
FAFSA Filer	Application Date & Type
Financial Need	Academic Program
Need-Based Aid	Learning Community Participation
Grants & Scholarships	Student Athlete
Student Loan Offer	Program Participant (Honors, etc.)
Current Economic Environment	Living on Campus
FAFSA Application Date (Priority Deadline)	Attended Orientation

Figure 1.

Open in figure viewer

Variables to Consider in Predicted GPA Model Development

For the second and third research questions, student GPA at the end of the first semester of college was collected. For the fourth research question, first-year, second-year, third-year, and fourth-year retention rates and four-year, five-year, and six-year graduation rates were collected. Retention rates were based on census data. For example, students who started in Fall 2007 would be listed as retained for the first year if they were enrolled in the institution on census day of Fall 2008; they would be listed as retained for the second year if they were enrolled in the institution on census day of Fall 2009, etc.

Data Analysis

For Research Question 1, data were analyzed using ordinary least squares regression on the 2004–2006 data. The goal of the analysis was to predict first-semester GPA. Individuals who did not have a first-term GPA were omitted from the analysis. The resulting regression coefficients were used to predict the first-term GPA for each student in the Fall 2007 entering cohort. Based on this predicted first-term GPA, students were placed in one of two categories: students predicted to earn less than a 2.0 first-semester GPA, and students who were predicted to earn a 2.0 or higher GPA. In 2008 and 2009 a similar approach was used.

For Research Question 2, students were divided into two groups: those who were predicted to earn less than a 2.0 first-semester GPA and those who were predicted to get a 2.0 or higher first-semester GPA. Actual first-semester GPAs for students were collected and four groups emerged: Group 1 included students who were predicted to earn less than a 2.0 and did earn less than a 2.0. Group 2 included students who were predicted to earn less than a 2.0 but earned a 2.0 or higher. Group 3 included students who were predicted to earn a 2.0 or higher but earned less than a 2.0, and Group 4 included students who were predicted to earn a 2.0 or higher and did earn a 2.0 or higher. Percentages of students in each category were recorded.

The four groups mentioned above were also used in the analyses of Research Questions 3 and 4. For Research Question 3, the mean first-semester college GPA for each category of students was calculated. Due to unequal sample sizes and variances, t-tests for unequal variances were conducted to assess if the means between groups were statistically significant (Mertler and Vannatta **2001**). For Research Question 4, retention and graduation rates for each category of students were calculated. Chi-square tests were performed to determine if the retention rates between groups were statistically significant.

T-tests and chi-square analyses were conducted between the two predicted groups (i.e., those predicted a 2.0 or higher and those predicted below a 2.0), between Group 1 and Group 3 (both groups who earned below a 2.0), and between Group 2 and 4 (both groups who earned a 2.0 or higher). The intent was to investigate if the predictive model may be useful, even if the students were incorrectly identified. So, for example, if students who were predicted to earn a 2.0 or higher did not, but their grades and retention rates were statistically significantly higher than those who were predicted to earn less than a 2.0, the predictive model, although not always accurate, still may be helpful in identifying at-risk students.

Results

Research Question 1

What is the influence of high school academic factors, demographic variables, financial variables, and program participation variables on first-semester college GPA?

Using data from cohort years 2004, 2005, and 2006 and ordinary least squares regression, a model was developed to identify students who were predicted to earn less than a 2.0 GPA. Prior to performing multiple regression techniques, the data set was analyzed to ensure it met the three general assumptions of multiple regression: normality, linearity, and homoscedasticity (Mertler and Vannatta **2001**). Evaluation of linearity led to data transformations for deposit days, high school rank, and number of days in admission cycle. Correlations among variables were also reviewed to ensure that multicollinearity was not a problem. Regression results indicated the model was statistically reliable in predicting first semester college GPA, $R^2 = .380$, $F(14,5941) = 259.83$, $p < .001$. Results indicated that the independent variables explain 38.0% of the variation in first-term GPA.

Regression coefficients specified that eleven variables significantly contributed to the model. There was a

positive correlation between high school GPA, ACT score (or converted SAT score), deposit days (the number of days between the start of the semester and when the student paid the enrollment deposit), being female, having an undecided major, being in the College of Human Science, having AP credits, number of days between the start of the semester and when the student submitted their application for admission (App Days), high school rank, and first-semester GPA. There was a negative relationship between financial need amount and student loan aid. A summary of means, standard deviations, and regression coefficients for the model predicting GPA is presented in Table 1.

Table 1. Means, Standard Deviations, and Regression Coefficients for Variables Predicting First Semester GPA

Variables	Mean	SD	b (Coefficient)	β	SE B	t-statistic
High School GPA	3.49	0.45	0.965**	.457	0.049	19.8
ACT Score	24.59	3.96	0.012**	.051	0.003	3.91
Financial Need	6,528.35	7,926.11	−0.000**	–.055	0.000	−5.04
Deposit Days	205.28	350.07	0.001**	.278	0.000	4.24
Female	0.45	0.50	0.087**	.046	0.020	4.27
Undecided	0.11	0.31	0.139**	.046	0.031	4.43
Loan Aid	12,769.68	7,851.84	−0.000**	–.031	0.000	−2.97
Human Sciences	0.10	0.29	0.190**	.059	0.034	5.68
AP Credit	1.40	4.54	0.063**	.073	0.010	6.51
App Days	280.55	72.54	0.057**	.034	0.022	2.58
High School Rank	74.70	17.17	0.000**	.074	0.000	3.27

R^2 = .380

** Statistically significant at the $p < .01$ level

Research Question 2

How accurately can a statistical model that uses precollege characteristics predict students who may earn above or below a 2.0 first-semester GPA?

The statistical model was applied to Fall 2007, Fall 2008, and Fall 2009 cohorts. The model calculated a predicted GPA for each student. Students were divided into two groups: those who were predicted to earn less than a 2.0 GPA and those who were predicted to earn a 2.0 or higher. At the end of the semester, the actual GPA was compared to the predicted GPA to examine the accuracy of the predicted model. For Fall

2007, 472 (13.9%) of the incoming students were predicted to earn less than a 2.0 GPA. Of those predicted to have less than a 2.0, 59.3% (N = 280) did earn less a 2.0 GPA; for Fall 2008, of the 580 (13.6%) students who were predicted to earn less than a 2.0 GPA, 58.3% (N = 338) did earn below a 2.0 GPA; for Fall 2009, of the 481 (12.1%) incoming students who were predicted to earn below a 2.0 GPA, 52.0% did earn below a 2.0 GPA (see Table 2).

Table 2. Comparison of the Predicted and Actual GPAs (Below 2.0 and 2.0 or Higher) for Fall 2007–Fall 2009

		All Years (N = 11644)		Fall 2007 (N = 3407)		Fall 2008 (N = 4256)		Fall 2009 (N = 3981	
		Actual GPA		Actual GPA		Actual GPA		Actual GPA	
		<2.0	≥2.0	<2.0	≥2.0	<2.0	≥2.0	<2.0	≥2.0
Predicted < 2.0	**N**	868	665	280	192	338	242	250	231
	%	56.6	43.4	59.3	40.7	58.3	41.7	52.0	48.0
Predicted ≥ 2.0	**N**	1,587	8524	479	2,456	594	3,082	514	2,986
	%	15.7	84.3	16.3	83.7	16.2	83.8	14.7	85.3

For Fall 2007, of the 2,935 students who were predicted to earn a 2.0 or higher GPA, 83.7% (N = 2,456) did earn 2.0 or higher GPA; for Fall 2008, of the 3,676 students who were predicted to earn a 2.0 or higher GPA, 83.8% (N = 3,082) did earn 2.0 or higher GPA; for Fall 2009, of the 3,500 students who were predicted to earn a 2.0 or higher GPA, 85.3% (N = 2,986) did earn 2.0 or higher GPA (see Table 2).

Research Question 3

Are there differences in first-semester GPAs between students who were predicted to earn less than a 2.0 and those who were predicted to earn a 2.0 or higher first-semester GPA?

Students were categorized into four groups. Group 1 included students who were predicted to earn less than a 2.0 and did earn less than a 2.0. Group 2 included students who were predicted to earn less than a 2.0 but earned a 2.0 or higher. Group 3 included students who were predicted to earn a 2.0 or higher but earned less than a 2.0, and Group 4 included students who were predicted to earn a 2.0 or higher and did earn a 2.0 or higher.

Overall, students who were predicted to earn less than a 2.0 had a lower mean first-semester GPA ($\bar{x}$ = 1.76) than those who were predicted to earn a 2.0 or higher ($\bar{x}$ = 2.82). This pattern was consistent throughout all three years: Fall 2007 ($\bar{x}$ = 1.69 vs $\bar{x}$ = 2.81), Fall 2008 ($\bar{x}$ = 1.75 vs. $\bar{x}$ = 2.80), Fall 2009 ($\bar{x}$ = 1.85 vs. $\bar{x}$ = 2.84). Of the four categories of students, students who were predicted to earn less than a 2.0 and did earn less than a 2.0 (Group 1) had the lowest mean first-semester GPAs ($\bar{x}$ = 1.11). Students who were predicted to earn a 2.0 or higher but earned less than a 2.0 (Group 3) had the second lowest GPA ($\bar{x}$ = 1.31). Students who were predicted to earn less than a 2.0 but earned a 2.0 or higher (Group 2) had the second highest mean first-semester GPA ($\bar{x}$ = 2.61), and students who were predicted to earn a 2.0 or higher and did (Group 4) had the highest mean first-semester GPA ($\bar{x}$ = 3.10). This pattern was consistent for each year (see Table 3).

Table 3. Sample Sizes, Mean Differences in First-Semester GPAs, and Standard Deviations and t-test Results for Predicted < 2.0 GPA, Predicted ≥ 2.0 GPA, and Four Groups

	All Years				Fall 2007			
	N	**Mean**	**SD**	***t***	**N**	**Mean**	**SD**	***t***
Predicted < 2.0	1,533	1.76	0.92		472	1.69	0.91	
Predicted ≥ 2.0	10,111	2.82	0.86	−42.18**	2,935	2.81	0.86	−24.83**
Group 1	868	1.11	0.62		280	1.09	0.61	
Group 3	1,587	1.31	0.57	−7.54**	479	1.31	0.56	−5.13**
Group 2	665	2.61	0.44		192	2.58	0.41	
Group 4	8524	3.10	0.56	−27.09**	2,456	3.10	0.56	−16.24**

Note : Group 1: Students predicted to earn less than a 2.0 and did earn less than a 2.0

Group 2: Students predicted to earn less than a 2.0 but earned a 2.0 or higher

Group 3: Students predicted to earn a 2.0 or higher but earned less than a 2.0

Group 4: Students predicted to earn a 2.0 or higher and did earn a 2.0 or higher

** Statistically significant at the p < .001 level

Due to unequal sample sizes and variances, independent sample t-tests for unequal variances were conducted to compare first-semester GPA between the predicted less than 2.0 and predicted 2.0 or higher groups, between Group 1 and Group 3, and between Group 2 and 4. Overall, students who were predicted to earn less than a 2.0 ($\bar{x}$ = 1.76, SD = 0.92) had significantly lower first-semester GPAs than students who were predicted to earn a 2.0 or higher ($\bar{x}$ = 2.82, SD = 0.86), $t(1959) = -42.18$, $p < .001$. Of those students who earned less than a 2.0 GPA, t-tests were conducted between the group that was accurately predicted to earn less than a 2.0 (Group 1) and those that had been predicted to earn a 2.0 or higher but did not (Group 3). There were significant differences in first-semester GPA between these groups. Similarly, of those students who earned a 2.0 GPA or higher, t-tests were conducted between the group that was accurately predicted to earn a 2.0 and higher (Group 4) and those that had been predicted to earn less than a 2.0 but did not (Group 2). There were significant differences in first-semester GPA between these groups. The t-test results for each year (2007, 2008, and 2009) are presented in Table 3.

Research Question 4

Are there differences in retention and graduation rates between students who were predicted to earn less than a 2.0 and those who were predicted to earn a 2.0 or higher first-semester GPA?

This study compared retention and graduation rates for students who were predicted to earn less than a 2.0 GPA and students who were predicted to earn a 2.0 or higher. Students who were predicted to earn less than a 2.0 GPA had lower retention and graduation rates than the other group. Of the four categories of students,

students who were predicted to earn less than a 2.0 and did earn less than a 2.0 (Group 1) had the lowest retention rates. Students who were predicted to earn a 2.0 or higher GPA but earned less than a 2.0 (Group 3) had the second lowest retention and graduation rates. Students who were predicted to earn less than a 2.0 but earned a 2.0 or higher (Group 2) had the second highest retention and graduation rates, and students who were predicted to earn a 2.0 or above and did (Group 4) had the highest retention and graduation rates (see Table 4).

Table 4. Retention and Graduation Rates and Chi-square Results for Predicted < 2.0 GPA, Predicted ≥ 2.0 GPA, and Four Groups

	All Years		Fall 2007		Fall 2008		Fall 2009	
	%	χ^2	%	χ^2	%	χ^2	%	χ^2
First-Year Retention								
Predicted < 2.0	68.5		65.9		69.0		70.6	
Predicted ≥ 2.0	87.8	391.47**	86.9	133.59**	87.1	126.14**	89.2	129.35**
Group 1	56.7		55.0		57.7		57.4	
Group 3	62.5	7.65**	61.3	2.89*	60.9	0.95	95.3	4.44*
Group 2	83.9		81.8		84.7		84.8	
Group 4	92.5	60.74**	91.9	22.64**	92.1	15.93**	93.4	23.11**
Second-Year Retention								
Predicted < 2.0	54.4		49.6		55.0		58.3	
Predicted ≥ 2.0	80.9	533.02**	80.0	204.17**	80.4	181.26**	82.2	145.73**
Group 1	40.7		38.9		40.5		43.0	
Group 3	47.9	11.63**	47.5	5.24*	48.5	5.49*	47.6	1.42
Group 2	72.2		65.1		75.2		74.9	

* $p < .05$

** $p < .001$

Note : Group 1: Students predicted to earn less than a 2.0 and did earn less than a 2.0

Group 2: Students predicted to earn less than a 2.0 but earned a 2.0 or higher

Group 3: Students predicted to earn a 2.0 or higher but earned less than a 2.0

Group 4: Students predicted to earn a 2.0 or higher and did earn a 2.0 or higher

	All Years		Fall 2007		Fall 2008		Fall 2009	
	%	χ^2	%	χ^2	%	χ^2	%	χ^2
Group 4	87.0	111.90**	86.3	61.70**	86.5	23.31**	88.1	33.40**
Third-Year Retention								
Predicted < 2.0	49.4		44.1		50.3		53.5	
Predicted ≥ 2.0	78.4	591.52**	77.2	224.84**	77.9	197.99**	80.0	166.43**
Group 1	35.2		32.1		37.0		36.1	
Group 3	44.4	19.75**	45.2	12.48**	44.3	4.72*	43.9	4.11*
Group 2	68.0		61.5		69.0		72.3	
Group 4	84.8	126.56**	83.5	58.08**	84.4	37.88**	86.3	33.24**
Fourth-Year Retention								
Predicted < 2.0	46.3		41.7		46.7		50.3	
Predicted ≥ 2.0	77.4	653.42**	76.3	237.77**	76.7	227.15**	78.9	185.48**
Group 1	31.6		30.4		32.8		31.3	
Group 3	42.2	26.45**	42.1	10.28**	42.4	8.32**	42.0	8.05**
Group 2	65.6		58.4		66.1		70.9	
Group 4	83.9	143.59**	83.0	69.64**	83.4	45.45**	85.2	32.99**
Four-Year Graduation Rate								
Predicted < 2.0	15.8		13.3		17.2		16.5	
Predicted ≥ 2.0	43.9	437.72**	42.4	145.53**	43.9	148.47**	45.1	142.81**
Group 1	6.8		6.1		7.4		6.8	

* $p < .05$

** $p < .001$

Note : Group 1: Students predicted to earn less than a 2.0 and did earn less than a 2.0

Group 2: Students predicted to earn less than a 2.0 but earned a 2.0 or higher

Group 3: Students predicted to earn a 2.0 or higher but earned less than a 2.0

Group 4: Students predicted to earn a 2.0 or higher and did earn a 2.0 or higher

	All Years		Fall 2007		Fall 2008		Fall 2009	
	%	χ^2	%	χ^2	%	χ^2	%	χ^2
Group 3	12.2	17.88**	11.1	5.30**	11.6	4.23*	14.0	8.44**
Group 2	27.5		24.0		31.0		26.8	
Group 4	49.8	122.85**	48.6	43.27**	50.2	33.06**	50.5	47.98**
Five-Year Graduation Rate								
Predicted < 2.0	35.6		31.1		35.7		39.9	
Predicted ≥ 2.0	71.2	759.34**	69.4	258.66**	70.5	270.31**	73.5	225.49**
Group 1	20.6		19.6		20.4		22.0	
Group 3	31.5	33.22**	29.6	9.20**	30.0	10.09**	35.0	13.39**
Group 2	55.2		47.9		57.0		59.3	
Group 4	78.6	191.43**	77.2	81.29**	78.4	57.32**	80.2	55.65**
Six-Year Graduation Rate (Fall 2007 & Fall 2008)								
Predicted < 2.0	39.2		36.2		41.6		n.a.	
Predicted ≥ 2.0	74.4	531.87**	73.7	265.19**	74.9	267.50**	n.a.	
Group 1	25.7		24.3		26.9		n.a.	
Group 3	36.6	21.18**	36.1	11.41**	37.0	9.91**	n.a.	
Group 2	58.3		53.6		62.0		n.a.	
Group 4	81.7	137.92**	81.1	80.93**	82.2	59.10**	n.a.	
N								
Predicted < 2.0	1,533		472		580		481	

* $p < .05$

** $p < .001$

Note : Group 1: Students predicted to earn less than a 2.0 and did earn less than a 2.0

Group 2: Students predicted to earn less than a 2.0 but earned a 2.0 or higher

Group 3: Students predicted to earn a 2.0 or higher but earned less than a 2.0

Group 4: Students predicted to earn a 2.0 or higher and did earn a 2.0 or higher

	All Years		Fall 2007		Fall 2008		Fall 2009	
	%	χ^2	%	χ^2	%	χ^2	%	χ^2
Predicted ≥ 2.0	10,111		2,935		3,676		3,500	
Group 1	868		280		338		250	
Group 3	1,587		479		594		514	
Group 2	665		192		242		231	
Group 4	8,524		2,456		3,082		2,986	

* $p < .05$

** $p < .001$

Note : Group 1: Students predicted to earn less than a 2.0 and did earn less than a 2.0

Group 2: Students predicted to earn less than a 2.0 but earned a 2.0 or higher

Group 3: Students predicted to earn a 2.0 or higher but earned less than a 2.0

Group 4: Students predicted to earn a 2.0 or higher and did earn a 2.0 or higher

Pearson chi-square analyses were performed to examine if there was a statistically significant difference in retention and graduation rates for predicted student achievement groups. Retention and graduation rates were significantly different between students who were predicted to earn less than a 2.0 and those predicted to earn a 2.0 or higher. The difference in retention and graduation rates was significant between each tested group except for first-year retention rates between Group 1 and Group 3 for Fall 2008 cohort and second-year retention rates between Group 1 and Group 3 for Fall 2009 (see Table 4).

Discussion and Implications

This study illustrates that a predictive model that uses institutional data commonly collected before a student arrives on campus can have wide practical application for the SEM manager identifying academically at-risk students. Over 56% of students predicted to earn less than 2.0 did earn less than a 2.0, and over 84% of students predicted to earn a 2.0 or higher did earn a 2.0 or higher. Understanding this allows the enrollment manager, faculty, and staff to focus intervention efforts on those most in need of assistance. As presented in Tables 3 and 4, students who were predicted to earn less than 2.0 GPA had lower first-semester GPA and lower retention and graduation rates than students who were predicted to earn a 2.0 or higher GPA. These findings support previous studies that have found college GPA to be an important predictor in retention (Gifford, Briceno-Perriott, and Mianzo **2006**). While other studies have looked at overall college GPA or GPA at the end of a student's first year of college (Chen and St. John **2011**), this study shows that predicted GPA prior to enrollment can be used to identify at-risk students, and also allows for earlier intervention.

In addition, students who were predicted to earn below a 2.0 their first semester but earned a 2.0 GPA or higher (Group 2) still had significantly lower retention rates and grade point averages than those who were predicted to earn a 2.0 or higher and did earn a 2.0 or higher (Group 4), indicating that actual first-term

achievement can serve as an important element of future persistence. Similarly, students who were predicted to earn a 2.0 or higher but did not (Group 3) had significantly higher grade point averages and retention rates than students who were predicted to earn less than a 2.0 and did earn less than a 2.0 (Group 1). Therefore, even when the model did not accurately predict whether a student would earn above or below a 2.0 the first semester, the model may still be useful in identifying at-risk students. This finding supports Tinto's (**1998**) recommendations that institutions “front-load” resources in the first year. It encourages institutions to be proactive in reaching out to students prior to enrollment as well as during their first semester as opposed to waiting until students do poorly.

The results of this study also reinforce a common theme in the discussions regarding student success: student success and persistence is a complex phenomenon and one that requires institution-wide attention. Academic success does not rely simply on one, two, or three variables. Therefore, a university-wide effort is required to improve student academic success. For instance, if only financial aid variables were significant, the institution may focus on financial aid policies and strategies. If precollege characteristics and academic preparation were the sole factors, institutions could focus efforts on developing academic support courses or programs. However, identifying students early and working individually with smaller groups of students may be most useful. Providing information to academic advisors for early interventions or first-year seminar instructors, for example, would be useful as these professionals work and interact with their students. It may give academic advisors an additional signal for at-risk students. The results of this study may help institutions identify and target smaller groups of students who are more likely to be at risk earlier, allowing the development of narrowly tailored programming that can more efficiently use staff time and financial resources.

Some of the variables used in this study are similar to ones used in other predictive models (i.e., financial aid variables, high school academic characteristics) but other variables are specific to the institution where the research was conducted. For example, the role of the academic college and participation in certain programs may be unique to this institutional setting. Although some of these factors, such as structure or organization of the college, may not be within the SEM professionals’ purview to change, it is nevertheless informative to know the relationship between these variables and predicted GPA and retention.

The results of this study suggest that institutions have data available that can assist in identifying students who may be academically at-risk prior to the student enrolling. Institutional leaders do not need to spend additional time or money to disseminate new surveys or collect data from other sources. Powerful information about a student's success is already accessible.

Consistent with Hirschy (**2015**), this finding reiterates the importance of developing models that account for unique institutional contexts. In order to select feasible models for improving student persistence and retention, SEM professionals need to rely on their professional judgment and experience as a guiding principle, modifying and customizing the approach in light of their institutional mission and vision. Engaging faculty, staff, and the enrollment management team with both empirically based models and institutional context will help accomplish student success goals. For example, while the data utilized for this study are from August (about a month prior to enrollment), through discussion with colleges it was determined that the operational model would utilize data from May (prior to summer orientation) so that academic advisors have the information available when students are first registering for courses.

This study also emphasizes the need for robust data management systems that capture historical data. In order to develop a model, past data, collected at similar points in time, need to be utilized (Bontrager **2004** ; Wohlgemuth et al. **2009**). While some variables, such as student background characteristics (gender, ethnicity, residence), are relatively stable, other variables are time sensitive (student's decision to live on campus, choice of major). Therefore, this study requires institutions to have “point-in-time” data. It reiterates

the need for institutions to invest in and maintain data management systems and to dedicate staff time and resources in this area (Wohlgemuth 2015).

Limitations and Recommendations for Future Research

Although the results of this study demonstrate the ability of a model to predict academically at-risk students, limitations exist. Since this study was conducted at one large, public 4-year institution, some of the variables (i.e., participation in specific institutional programs) are unique to the institution and the results may not be generalizable to all institutions.

Additionally, academic success can be influenced by students' attitudes, expectations, and motivations as well as other noncognitive measures of the students ability to persist, often referred to as "grit" (Cortes and Kalsbeek 2012 ; Duckworth et al. 2007). This study did not account for these factors; therefore, although the model can be used to identify students likely to earn less than a 2.0, the study also illustrated that the available information does not allow for a perfect prediction.

Finally, this model considered all entering freshman students across the university. Examining smaller cohorts of students, such as those entering closely related majors (e.g., STEM fields or majors in a particular college), may allow a more accurate prediction of academic success.

Conclusion

This study sought to determine if a statistical model, using preexisting institutional data, could be developed to predict academically at-risk students and whether this model was accurate in predicting student retention and graduation. Results suggest that it is possible to identify students who are at risk of earning a low GPA using multiple regression and data available prior to enrollment. Results also illustrate that first-semester GPA is a critical factor in retention and graduation rates. Students who had less than a 2.0 first-semester GPA are less likely to make successful progress toward graduation. Therefore, institutions that are interested in increasing their retention and graduation rates may want to focus efforts on assisting students who are predicted prior to enrollment to earn less than a 2.0 during the first term and subsequently focus on students who did earn less than a 2.0 in their first semester.

Biographies

Ann M. Gansemer-Topf is an assistant professor of higher education at Iowa State University. Prior to becoming a faculty member she had worked in admissions, institutional research, and academic advising and residence life. She also has significant experience in program evaluation and assessment. Her research interests focus on examining the micro (student) and macro (institutional, state, federal) factors that impact student success. She has her doctoral and master's degrees from Iowa State University and her bachelor's degree from Loras College in Dubuque, Iowa.

Jonathan Compton is a senior research analyst in the Office of the Registrar at Iowa State University and is a member of the Enrollment Research Team. In this role, he is responsible for creation of reports, enrollment and course projections, and analysis of data in support of student success initiatives at the university. He has been in his current role for 7 years. He holds a PhD in educational leadership and policy studies from Iowa State University, a master's degree in teaching English as a second language from Iowa State University, and a bachelor's degree in English from Bryan College.

Darin Wohlgemuth is the director of assessment and enrollment research for the Division of Student Affairs at Iowa State University. He leads Iowa State's Enrollment Research Team (ERT) conducting research on a variety of area from strategic recruitment, tuition policy, and student success. Wohlgemuth, along with the ERT, have presented regularly at AACRAO's Strategic Enrollment Management conference. He has authored and coauthored more than 15 articles and book chapters. He earned his master's and doctoral degrees in economics from Iowa State University, where his research examined the demand for higher education at the aggregate and individual levels. His has a bachelor's degree in secondary math education from the University of Kansas and an associate's degree from Hesston College.

Greg Forbes is the research analyst for the Office of Student Financial Aid at Iowa State University. He provides data, assessment, and research support for financial aid and is a member of the Iowa State University Enrollment Research Team. Greg has seventeen years of experience in financial aid serving in a variety of functions including counseling, student employment, America Reads–America Counts, website design and development, loan reconciliation, IT support and training, verification, assessment, data analytics, and research. He has particular interest in the intersections of financial aid and student success, enrollment management, and student loan debt. Greg received a master's of public administration with a focus in higher education from Iowa State University and a bachelor's of science in forestry from the University of Illinois at Urbana-Champaign.

Ekaterina (Kate) Ralston is the research manager for Admissions at Iowa State University. Kate provides analytic support to the Office of Admissions as well as contributing to the projects conducted by the Enrollment Research Team. Kate has a doctoral degree in sociology, a master's in mass communications from Iowa State University, and a bachelor's in newspaper journalism from Moscow State University, Russia. Her areas of interest include multi-method approaches to data, focusing on quantitative techniques, such as structural equation modeling, longitudinal studies, and interaction analysis.

References