Data Driven Student Portal for Improved Student Success

Andy T. Clark, Vice President for Enrollment, Marketing & Communications

Brian A. Haugabrook, Chief Information Officer

Barrie D. Fitzgerald, Senior Research Analyst

Valdosta State University

atclark@valdosta.edu bahaugab@valdosta.edu bdfitzgerald@valdosta.edu

Abstract — Retention and graduation rates have long been student success indicators, which occur at the end of a term or academic year. These indicators have become non-effective, in today's times, in providing necessary resources to help students achieve academic success. By changing the culture on one campus to focus on providing faculty with indicators early in the semester or even before students step foot in class dramatically helped to increase the students' success. Predictive analytics can assist faculty and academic support staff in helping students achieve academic success. Utilizing technology and predictive analytics to communicate and facilitate strategies not only increase retention, progression, and graduation, they provide opportunities for students' success.

Introduction

For the past seven years, a recent trend began to occur at Valdosta State University (VSU). The university's entering freshman class experienced an increase in the number of entering freshman students, while the one-year retention rate experienced a decline. The one-year retention rate steadily declined since Fall 2005's 73.6% retention rate. VSU's three-year retention average, Fall 2010 to Fall 2012, is approximately 67.5%, meaning almost a third of the students leave the university within one year (Strategic Research & Analysis, 2012). In terms of revenue lost to the institution, this equates to approximately $6.5 million per year when taking into account the multiplicative effect over a four year college career.

Prior to 2012, VSU students typically did not seek help from academic support centers until mid-term grades were posted, which at this point in time the chances of the students succeeding within the struggling courses decreased drastically. This contributed to VSU's declining retention because of students not succeeding in their academic studies. Moreover, any adjustments to any learning communities and other academic support offices and/or resources were deferred a year before retention and graduation reports were ready; by that time, the students who exhibited signs of struggles academically had already left the university.

During Spring 2012, the State of Georgia implemented its Complete College Georgia, which is an adapted version of the national Complete College America. The main focus of the plan is to increase the number of students who graduate from a post-secondary school. The plan states that in order "[t]o remain competitive, Georgia must not only maintain current graduation levels, but also produce an additional estimated 250,000 graduates in upcoming years" (University System of Georgia, 2013). For VSU, this means a shift needs to occur in its focus towards retaining students at the university, progressing them through their academic careers, and graduating them with a degree in a timely manner.

As a result of the Complete College Georgia initiative, VSU began exploring methods of developing predictive metrics of student success in college coursework based on indicators. The purpose of developing these predictive metrics centers around utilizing intervention strategies developed to assist high risk students to achieve academic success. Several indicators were analyzed for inclusion in the predictive metrics beginning with its admission requirements. A student is admitted to VSU based on a freshman index, where students are required to meet a minimum set of scores of 900 SAT composite with a minimum SAT critical reading subscore of 430 and SAT mathematics subscore of 400 or 19 ACT composite with a minimum ACT English subscore of 17 and ACT mathematics subscore of 17. Additionally, students are required to have completed the college preparatory high school curriculum (Valdosta State University, 2013).

Literature Review

Using predictive analytics and at-risk modeling, it may be possible to identify future students who are at-risk of not retaining, progressing, and graduating from VSU. This literature review is a conglomeration of studies with the salient focus of extrapolating all of the factors within attrition rates and poor college performance, and using that data to assuage current retention issues as well as predict future at-risk students. The result, in theory, would be an increase in future retention and graduation rates, as well as a model to identify at-risk students early in their careers and aid them in various academic and non-academic pathways.

Koppius and Shmueli (2011) constructed an article that concerns the importance of predictive analytics and the lack of such in information systems research. Predictive analytics includes statistical models with the intention of empirical predictions. This is different from other predictive analyses as it focuses on more than just prediction from theory. Methods for assessing the quality of those predictions in practice also fall under this terminology. Predictive analytics ability hinges upon six aspects: generating new theory, developing measures, comparing competing theories, improving existing models, assessing relevance, and assessing predictability (Dubin 1969; Kaplan 1964, as cited in Koppius & Shmueli, 2011).

The first step in developing a predictive tool is goal definition, which is centralized around what specifically needs predicting. A common goal that exists in predictive modeling is attempting to accurately predict an outcome value for a new set of observations. This is known as prediction for numerical data, and classification for a categorical outcome. If the outcome is categorical, a different goal is used to attempt to rank a new set of observations according to their probability of belonging to a certain class, also known as ranking. The next progression would be data collection and study design. Ideally, the data used for modeling and for prediction consists of the same variables and are from the same population. Predictive analytics needs to have a larger sample size than explanatory modeling or regular experimental procedures because there is a higher uncertainty for predicting individuals behavior juxtapose to population-level parameters. Increasing sampling size also reduces both model bias and sampling variance (Koppius & Shmueli, 2011).

Data dimension is another consideration in this phase and begins with a large number of variables dependent upon domain knowledge and potential for new relationships. Data preparation follows from data collection, and involves missing values. Missing values are handled much the same as any data analysis, and uses proxy variables, dummy variables, and regression trees to counter the absent values. Exploratory Data Analysis (EDA) is used primarily in predictive analytics. It is used in a free-form fashion to uncover potential underlying constructs that may be less formulated. Principle components analysis (PCA) is a data reduction technique that is often used prior to the EDA to reduce sampling variance and perhaps increase predictive accuracy (Koppius & Shmueli, 2011).

Predictive models and analytics can lead to the discovery of new constructs or relationships, and provide evidence regarding unknown patterns. They are more data driven than explanatory statistical models in that they are derived from empirical information, and are a good source to assess practical relevance of theories. A potential problem with predictive analyses could be the extent to which they are willing to reduce sampling variance. The result could be an increase in the amount of method bias involved in the predictive analysis and may lead to artificial numbers. Predictive models and analytics can be used in the education setting as well in an attempt to determine several factors including graduation, persistence, and retention rates (Koppius & Shmueli, 2011).

While predicting retention rates is possible and plausible, another area of concern is identifying at-risk students and working with them which could raise retention and graduation rates. At-risk students are found to remain at-risk throughout their college career. Also, the degree to which the student was at-risk is predictive of whether the student subsequently re-enrolls elsewhere and the type of institution at which the re-enrollment occurs (Singell & Waddell, 2010). Burgher and Davis (2013) conducted a study to determine students' enrollment behavior by analyzing ethnicity, gender, high school grade point average, ACT scores, and college attended by family members. Building on those variables described by Burgher and Davis, additional variables affecting retention rates include institutional selectivity and academic preparation (Singell & Stater, 2006, as cited in Singell & Waddell, 2010). There exists uncertainty regarding the prediction of retention and the efficacy of treatment as to whether administrative action and associated resource expenditures would yield a net benefit. This problem is becoming more important with the rise of conscious higher education institutions. Burgher and Davis (2013) found that additional circumstances occurring in an individual's life make it impossible to predict what factors have the most influence on their reasoning for continuing their education.

There are two ways to view targeting at-risk students in correlation to retention rates. The first is a type I error in which you forgo treatment of all students in the class. Conversely, type II error can be defined as applying treatment to all students, regardless of whether or not they are at-risk. The results of the research indicate, as might be hypothesized, that a balance must be struck to avoid using too much capital. There must be a way to target those who are *a priori* at-risk and work with them to increase retention rates without attempting treatment on everyone in the class. As stated previously, at-risk status usually continues for students throughout their collegiate life, so collecting data to analyze if a student is at-risk after their first semester allows for treatment to a large majority of people who will remain at-risk. A possible correlation that exists between expected grade point averages (the average of all grade point averages for that class) and actual grade point averages for the student could assuage some of the difficulties in becoming at-risk early. It was discovered that students who take somewhat harder classes in their initial schedules maintain a higher grade point averages comparative to the expected grade point averages for that class. Previous literature suggests that what happens to students after they enter college is more influential in their persistence decisions than characteristics they bring to college (Pascarella and Terenzini, 2005, as cited in Singell & Waddell, 2010), but because we know that at-risk students stay at-risk throughout their studies, more comprehensive advising could be beneficial. Creating a model to observe retention rates of several cohorts would be a start to experimenting with treatments.

Duan-Barnett and St. John (2012) were interested in testing whether or not there was a correlation between various high school elements and college continuation. The elements included difficulty of high school math curriculum and mandatory exit exams. There was a significant relationship between high schools requiring algebra or above and college continuation as well as the number of math courses required and college continuation. Mandatory exit exams were also found to share a positive correlation with college continuation. However, mandatory exit exams also exert a negative effect on high school completion rates, while a rigorous math curriculum does not.

To conclude, there is a compendium of factors that must be accounted for when considering how to improve retention. The literature suggests that a mix of academic and nonacademic factors must be integrated. Factors like high school grade point average and SAT scores, as to be expected, have a high correlation with success at a university, but the social atmosphere and context of the modern student tends to be underestimated when retention is examined. Using predicted analytics that consider all of these variables including financial aid to students with varying income, it may be possible to locate and aid future at-risk students. When these at-risk students are found, systems such as peer support through blocked registration, freshman learning communities, and mentoring programs must be put into place. Keeping in mind that at-risk students stay at-risk throughout the entirety of their collegiate life, aid or support for these students at any time would be beneficial.

Hypotheses and Research Design

Since all new freshman students admitted into the institution are based on a freshman index and curriculum requirements, VSU decided to use these components as a beginning for the hypotheses of the development of its predictive analytics. VSU's wanted to test three hypotheses:

1. At-risk general: VSU postulates that there is a relationship in the variables of standardized test scores, high school grade point average, and high school curriculum rigor and performance to predict college retention.
2. At-risk mathematics-based courses: VSU postulates that there is a relationship in the variables of standardized test mathematics sub-scores, high school grade point average, and high school curriculum rigor and performance to predict success in mathematics based courses.
3. At-risk reading-based courses: VSU postulates that there is a relationship in the variables of standardized test critical reading sub-scores, high school grade point average, and high school curriculum rigor and performance to predict success in reading based courses.

Research Design and Data Collection

Population

A total of five cohorts—totaling 11,167 first-time, full-time freshmen—were analyzed to develop the predictive analytics. Attributes such as high school attended, high school grade point average, and standardized test scores were collected. For the retention component, students' enrollment in the university for one-year later was also collected. For the pass rate, students' initial semester course data was collected to include the grades for the course and the subject of the course. The courses' subject was examined to determine whether courses were open courses, reading-based courses, or mathematics-based courses.

Standardized Test Scores

With the standardized test scores, the SAT scores were used in the analysis. The SAT composite scores only accounted for the critical reading and the mathematics sub-scores. The SAT composite scores are scored in ten point intervals with a maximum score of 1600. The sub-scores maximum score is 800 for the both critical reading and mathematics. Students with only ACT test scores had their scores converted to SAT scores using a conversion chart (StudyPoints, 2014). For example, an ACT composite score of 24 was converted to an 1120 SAT Composite score. Also for students who had both an SAT composite and an ACT composite scores, the ACT composite was converted to an SAT composite score and the higher of the two scores were chosen for the analysis. For example, if a student had an 1100 SAT composite score and a 25 ACT composite score, the ACT composite score's converted score of 1160 SAT composite was chosen for the analysis. With the sub-scores the same process of the conversion occurred with the addition of dividing the converted score by two to get a single sub-score to use in the analysis. The top quartile is given a value of one, while the bottom quartile is given a value of four. The standardized test scores are divided into four categories or quartiles that are reevaluated each year.

High School Grade Point Average

Students' high school grade point averages were converted to a four-point scale based on completion of the college preparatory curriculum. The highest grade point average conversion a student could have received, regardless if they had taken advanced placement courses in high school, was a four. (CollegeBoard, 2014). Additionally, some schools do not assign numerical grade point averages; instead, they simply use letter grades that may or may not include a plus or negative sign with the grade. The students' high school grade point averages were placed into four categories or quartiles. The top quartile is given a value of one, while the bottom quartile is given a value of four. These quartiles ranges are reevaluated each year.

High School Curriculum Rigor and Performance

Within the State of Georgia, the public high schools' performance on state mandated tests are published on the Governor's Office of Student Achievements' website. Three years of schools' graduation tests and performance tests were obtained and analyzed to determine which public school curriculum was more rigorous and produced successful students academically. Schools' pass rates were examined and assigned point values based on the percentage of students who passed and those who exceeded the expectations of the graduation and performance tests. The schools with higher percentage of students who exceeded the test expectation were given a higher point value. The total points were placed into four categories or quartiles. Students who came from a non-Georgia public high school, or a Georgia public high school with too few students were given a null category. The top quartile is given a value of one, while the bottom quartile is given a value of four. This component, like the standardized test scores and high school grade point averages, are on a rolling basis where each year the quartiles ranges are reevaluated to adjust to the changes in the academic success of students.

Explanation of the At-risk Coding

With the high school curriculum rigor and performance, the high school grade point average, and the standardized test scores tiers, the results yielded 80 different combinations. Tier value of 1 means the top, while tier value 4 means the bottom. A student's code would look like the following: 3-1-4. The first number would be for the high school curriculum rigor and performance, the second number would be for the high school grade point average, and the last number would be for standardized test scores.

Data Analysis and Findings

In order to test the significance of the categories, a Pearson product-moment correlation coefficient was conducted to assess the relationship between retention, high school curriculum rigor and performance, high school grade point average, and standardized test scores. The results are shown in Table 1. In regards to retention, there was no correlation with the high school curriculum rigor and performance variable (r=-.002, n=11,167, p=.798); however, there was a significant weak positive relationship with the high school grade point average tiers (r=.169, n=11,167, p<.001) and a significant very weak positive relationship with the standardized test scores (r=.059, n=11,167, p<.001). For high school curriculum rigor and performance, there was a significant very weak negative relationship with high school grade point average (r=-.073, n=11,167, p<.001), and a significant weak negative relationship with standardized test scores (r=-.125, n=11,167, p<.001). Also, there was a significant strong positive relationship between the high school grade point average tiers and the standardized test scores tiers (r=.446, n=11,167, p<.001).

Table 1: Pearson's Correlation Matrix of At-risk Factors

		1	2	3	4
1. Retained	Pearson's Correlation		-.002	.169	.059
	Sig. (2-tailed)		.798	.000	.000
	Number		11,167	11,167	11,167
2. High School Curriculum Rigor and Performance	Pearson's Correlation	-.002		-.073	-.125
	Sig. (2-tailed)	.798		.000	.000
	Number	11,167		11,167	11,167
3. High School Grade Point Average	Pearson's Correlation	.169	-.073		.446
	Sig. (2-tailed)	.000	.000		.000
	Number	11,167	11,167		11,167
4. Standardized Test Scores	Pearson's Correlation	.059	-.125	.446	
	Sig. (2-tailed)	.000	.000	.000	
	Number	11,167	11,167	11,167	

High School Curriculum Rigor and Performance's Effect

With the correlation indicating the high school curriculum rigor and performance not having a significant relationship with retention, an in-depth analysis was conducted to determine if there is any effect occurring on students' retention, high school grade point average and standardized test scores. Table 2 shows the retention rate of the students by the high school tier. Of the tiers, Tier 1, the toughest high school curriculum rigor, has the lowest retention rate. Further analysis of Tier 1 students showed that while these students are from the highest tier school, they transferred to another institution, such as University of Georgia and Georgia Institute of Technology. These institutions may have been the students' first choice in enrolling in an institution and were admission into the institution. Thus, they enrolled at VSU to gain credit hours and higher grade point average so they could be accepted into their initial choice institution. Additionally, VSU is located in Lowndes County which has two large public school systems, Lowndes County School System and Valdosta City School System contributing a large number of students into the university. The students who are from these two school systems have a high retention rate, which distorts the retention rate for their respective tier. One possible reason is the possibility that these students are more likely to live at home where their family could support the students and commute to VSU for classes. When Lowndes and Valdosta high schools are excluded, the retention rates are about the same throughout the tiers.

Table 2: Crosstabulation by High School Tier and Retention

All High Schools

High School Tier		Retained	Not Retained	Total
Tier 1	Number	1,554	787	2,341
	Percent	66.4%	33.6%	100.0%
Tier 2	Number	2,241	957	3,198
	Percent	70.1%	29.9%	100.0%
Tier 3	Number	1,551	754	2,305
	Percent	67.3%	32.7%	100.0%
Tier 4	Number	1,241	564	1,805
	Percent	68.8%	31.2%	100.0%
No Tier	Number	1,024	494	1,518
	Percent	67.5%	32.5%	100.0%
Total	Number	7,611	3,556	11,167
	Percent	68.2%	31.8%	100.0%

Lowndes and Valdosta High Schools Excluded

High School Tier		Retained	Not Retained	Total
Tier 1	Number	1,554	787	2,341
	Percent	66.4%	33.6%	100.0%
Tier 2	Number	1,708	808	2,516
	Percent	67.9%	32.1%	100.0%
Tier 3	Number	1,551	754	2,305
	Percent	67.3%	32.7%	100.0%
Tier 4	Number	1,066	480	1,546
	Percent	69.0%	31.0%	100.0%
No Tier	Number	1,024	494	1,518
	Percent	67.5%	32.5%	100.0%
Total	Number	6,903	3,323	10,226
	Percent	67.5%	32.5%	100.0%

A one-way analysis of variance (ANOVA) was conducted to evaluate the relationship between the high school curriculum rigor and performance and the high school grade point average in four tiers. The schools without a tier were not included in the analysis because it contained a vast variety of types of schools. The independent variable was the high school curriculum rigor and performance, which consisted four levels: (a) Tier 1, (b) Tier 2, (c) Tier 3, and (d) Tier 4. The dependent variable was high school grade point average of a student admitted to VSU. The ANOVA was significant $F(3, 9{,}609) = 100.894$, p<.001. The strength of the relationship, as assessed by η_p^2, was very weak with the high school curriculum rigor and performance accounting for 3.1% of the variance of the dependent variable. Follow-up tests were conducted to evaluate pairwise differences among means. The Tuskey HSD test indicated significant differences between the Tier 1 (*M*=2.95, *SD*=0.439) and Tier 2 (*M*=3.05, *SD*=0.458), Tier 1 and Tier 3 (*M*=3.12, *SD* 0.447), and Tier 1 and Tier 4 (*M*=3.17, *SD*=0.446). There was a difference between Tier 2 and Tier 3, and Tier 2 and Tier 4. Also there was a difference between Tier 3 and Tier 4. The 95% confidence interval for the mean, as well as the means and standard deviations are reported in Table 4. This means that students from a Tier 1 have a significantly lower high school grade point average when compared to students from Tier 4.

Table 4: Mean High School Grade Point Average by High School Tier

High School Tiers	Mean	SD	95% Confidence Interval for	
			Lower Bound	Upper Bound
Tier 1	2.95	.439	2.93	2.97
Tier 2	3.05	.458	3.03	3.06
Tier 3	3.12	.447	3.11	3.14
Tier 4	3.17	.446	3.15	3.19

Note: Schools without a tier were excluded from the table.

A one-way ANOVA was conducted to evaluate the relationship between the high school tiers and the standardized test scores in four tiers. The schools without a tier were not included in the analysis because it contains a vast variety of types of schools. The independent variable is the high school tiers, which consisted of four levels: (a) Tier 1, (b) Tier 2, (c) Tier 3, and (d) Tier 4. The dependent variable is the standardized test score of a student admitted to VSU. The ANOVA was significant $F(3, 9{,}536)=51.634$, p<.001. While the ANOVA indicates significance, the underlying assumptions of the test were violated based on the Levene Statistic, $F(3, 9{,}536)=8.573$, p<.001. However, two robust tests of equality of means were conducted to further support the significance found in the ANOVA test. The

Welch test was significant, $F(3, 4{,}976) = 52.728$, p<.001; additionally, the Brown-Forysthe was significant, $F(3, 9{,}106) = 52.625$, p<0.001. These two tests provide support that the significance found within the ANOVA test is still significant. The strength of the relationship, as assessed by η_p^2, was very weak with the high school curriculum rigor and performance accounting for 1.6% of the variance of the dependent variable.

Follow-up tests were conducted to evaluate pairwise differences among means. Table 5 shows the means The Games-Howell test did not indicate any significant difference between Tier 1 (*M*=1010, *SD*=99.440) and Tier 2 (*M*=1010, *SD*=105.366). The follow-up test indicated significant differences between Tier 1 and Tier 3 (*M*=985, *SD*=95.903) and Tier 1 and Tier 4 (*M*=982, *SD*=96.611). Additionally, Tier 2 and Tier 3 indicated to be a significant difference, as well as Tier 2 and Tier 4. There was no significant difference between Tier 3 and Tier 4. The ANOVA test indicated the significance that is found within the interquartile range of the standardized test scores by high school tiers.

Table 5: Mean Standardized Test Scores by High School Tier

High School Tiers	Mean	SD	95% Confidence Interval for	
			Lower Bound	Upper Bound
Tier 1	1010	99.440	1006	1014
Tier 2	1010	105.366	1006	1013
Tier 3	985	95.903	981	989
Tier 4	982	96.611	978	987

Note: Schools without a tier were excluded from the table.

While the high school curriculum rigor and performance did not have a significant effect on the students' retention rates, it did have a significant effect on the students' high school grade point average and standardized test scores. The high school curriculum rigor and performance variable was kept as an added factor due to its effect on the other two variables.

At-risk Predictive Analytics

In order to develop the three predictive models, the types of outcomes are examined. The outcomes would be the following: (a) for retention, a student retains or not retains, (b) for mathematics-based course, a student passes or does not pass, and (c) for reading-based courses, a student passes or does not pass. Since linear or ordinary least squares would violate the normality and constant variance assumption, logistic regression was performed. Logistic regression accounts for a dichotomous dependent variable outcome. Using the following equation, one could predict a student's likelihood of retaining to VSU or passing a VSU reading-based and mathematics-based course:

$$P = \frac{e^{[\beta_0+\beta_1(X_1)+\beta_2(X_2)+\beta_3(X_3)]}}{1+e^{[\beta_0+\beta_1(X_1)+\beta_2(X_2)+\beta_3(X_3)]}}$$

At-risk General Retention

With the high school tiers having five categories (five categories including the null category) and the high school grade point average and standardized composite test scores having four categories, this results in a total of 80 different combinations to predict the retention. A test of the full model against a constant only model was statistically significant, indicating that the predictors as a reliable set to determine the probability of students retaining to the university (χ^2=372.326, *df*=3, *p*<.001). The constant is significant (β_0=1.675, df=1, p<.001), the high school curriculum rigor and performance was not significant (β_1=-0.001, df=1, p=.865), the high school grade point average was significant (β_2=-0.362, df=1, p<.001), and the standardized test scores was not significant (β_3=0.011, df=1, p=.583). While the high school curriculum rigor and performance and standardized test scores were not significant, the overall model was significant. Using the following equation from above, students with a 1-1-1 have a retention probability of 79.0%, while 4-3-1 students would have a retention probability of 64.4%. Within the actual and predicted retention rates, any combination that has an actual or predicted rate of 65.0% or lower was flagged as at-risk for retaining to the university.

At-risk Mathematics-based Courses

Like the at-risk general, the at-risk mathematics-based courses have a total of 80 different combinations. A test of the full model against a constant only model was statistically significant, indicating that the predictors as a reliable set to determine the probability of students passing a mathematics-based courses at the university (χ^2=1641.501, *df*=3, *p*<.001). The constant is significant (β_0=2.794, df=1, p<.001), the high school curriculum rigor and performance was not significant (β_1=-0.007, df=1, p<.359), the high school grade point average was significant (β_2=-0.608, df=1, p<.001), and the standardized test scores was significant (β_3=-0.270, df=1, p<.001). While the high school curriculum rigor and performance was not significant, the overall model was significant. Using the following equation from above, students with a 1-2-3 would have a mathematics-based course pass rate probability of 68.2%, while students with 4-4-4 have a 32.2% pass rate. Within the actual and predicted pass rates, any combination that has an actual or predicted rate of 65.0% or lower was flagged as at-risk for failure in a mathematics-based course at the university.

At-risk Reading-based Courses

With the 80 different combinations, a test of the full model against a constant only model was statistically significant, indicating that the predictors as a reliable set to determine the probability of students passing a reading-based course at the university (χ^2=1641.501, *df*=3, *p*<.001). The constant is significant (β_0=3.083, df=1, p<.001), the high school curriculum rigor and performance was significant (β_1=-0.012, df=1, p=.030), the high school grade point average was significant (β_2=-0.555, df=1, p<.001), and the standardized test scores was significant (β_3=-0.111, df=1, p<.001). Using the following equation from above, students with a 1-2-1, they would have reading-based pass probability of 86.4%, while 4-4-4 students have a 59.2% pass rate. Within the actual and predicted pass rates, any combination that has an actual or predicted rate of 65.0% or lower was flagged as at-risk for failure in a reading-based course at the university.

Implementation and Results

With the predictive analytics developed to provide information about the entering students to faculty, a portal, called Valdosta State University Faculty portal, was launched in August 2012. The faculty portal provides faculty with information on students who are enrolled in their courses and it also provides information on faculty advisees. This information includes the following, a picture of the student, student contact information, the at-risk variables, integration with degree works, and a link to a faculty reporting form for attendance and academic progress. This allows the institution to provide critical student success metrics to faculty rapidly, and this information is available to faculty as soon as registration opens for a given term. When a faculty member reports a student for attendance or academic progress the technology running behind these forms sends automated communications to staff members who are charged with providing interventions to help the students succeed. The timely reporting of the students and deployment of the interventions should lead to enhance student success with regard to pass rates in their courses.

When the portal with the predictive analytics was first implemented at the university, there were three expected or desired results that would stem from the implementation:

1. Faculty who used the portal to inform students about their progress would have higher pass rates than those who did not use the portal.
2. Students who were flagged at-risk academically, either through predictive analytics or by the faculty would have improved pass rates.
3. As a result of improved pass rates, the cohort retention would increase.

Over the course of the academic year 2012-2013, the data collected by the portal was analyzed, especially focusing on the faculty who had a high number of at-risk students enrolled in their courses. Largely, this was the Department of Mathematics and Computer Sciences. Table 12 shows the crosstabulation of the pass rates by faculty views. The threshold was set at least 100 views for improvement to occur. Pass rates of faculty who had 100 views or more had a 6.3% higher pass rate than those who had less than 100 views. In order to determine if the increased pass rates were statistically significant, a chi-square test for independence was conducted. The relation was significantly different, $\chi^2(1, N=7{,}475)=28.097$, $p<.001$. The size effect, Cramer's V, is a weak relationship, .061. This means that students who had a faculty who had at least 100 views in the portal are more likely to have higher pass rates than students who had a faculty who had less than 100 views.

Table 12: Crosstabulation of Pass Rates by Faculty Page Views

Views		DFW	Pass	Total
Less than 100 views	Number	1,795	3,444	5,239
	Percent	34.3%	65.7%	100.0%
100 views or more	Number	626	1,610	2,236
	Percent	28.0%	72.0%	100.0%
Total	Number	2,421	5,054	7,475
	Percent	32.4%	67.6%	100.0%

Additionally, the flag set by a faculty member would potentially show faculty's intentions of helping a student to succeed in the course. The threshold was set at a minimum of five flags. Table 13 shows the crosstabulation of pass rates by the faculty who set at least a minimum of five flags. Of the faculty who set at least five flags, the pass rate is 10.2% higher than the pass rates of the faculty who set fewer than five flags. In order to determine if the increase in pass rates was statistically significant, a chi-square test for independence was conducted. The relation was significantly different, $\chi^2(1, N=7{,}475)=50.078$, $p<.001$. The size effect, Cramer's V, is a weak relationship, .082. This means that faculty who set at least five flags in the portal are more likely to have higher pass rates than the faculty who had set less than five flags.

Table 13: Crosstabulation of Pass Rates by Faculty Flag Sets

Flags		DFW	Pass	Total
Less Than Five Flags	Number	2,114	4,080	6,194
	Percent	34.1%	65.9%	100.0%
Five Flags or More	Number	307	974	1,281
	Percent	24.0%	76.0%	100.0%
Total	Number	2,421	5,054	7,475
	Percent	32.4%	67.6%	100.0%

Due to the success in the increase of passing grades, the retention rate within the Fall 2012 cohort increased if they had a faculty member who used the portal. Table 14 shows the crosstabulation of the retention rates of the cohort students. Students who had a faculty who utilized the portal had a 4.9% higher retention rate than those who did not have a faculty that utilized the portal. A chi-square test for independence was conducted to determine the significance of the relationship. The relationship was found to be significantly difference, $\chi^2(1, N=1{,}880)=4.776$, $p=0.029$. This means that students who had a faculty utilizing the portal are more likely to retain at a higher rate than those who did not have a faculty who utilized the portal.

Table 14: Retention of Fall 2012 Cohort by Portal Users

Portal Usage		Not Retained	Retained	Total
Non-Portal Users	Number	216	416	632
	Percent	34.2%	65.8%	100.0%
Portal Users	Number	365	883	1,248
	Percent	29.2%	70.8%	100.0%
Total	Number	581	1,299	1,880
	Percent	30.9%	69.1%	100.0%

Conclusion

Overall the data gathered from developing predictive analytics and implementing a portal system that was utilized by VSU's faculty shows that these interventions have helped students succeed in their academic careers at VSU. Moreover as a result of the success of the portal, the information was distributed to advisors and teaching faculty so that they become more aware of the abilities and possible struggles of the students they advise and teach in their courses.

Essentially the predictive analytics and the portal take a more proactive role in student success and intervention strategies. When faculty members log into the portal and select courses for which they are teaching, they will see their student roster with pictures of their students and indicators if the student is at-risk in any of the three areas. This portal system has been integrated with our advising software, DegreeWorks, so faculty advisors will have a better understanding of their advisees and their likelihood to struggle with certain coursework. Faculty also have the ability to flag a student who is struggling within their course, regardless of whether the student is at-risk or not, and by flagging the student a series of automated communications to academic and student support services will be generated so that a proactive approach to tutoring in student success can be made by the institution.

As a result of the improvements made from predictive analytics, VSU has begun to research advanced math at-risk where it identifies a student who will be more likely to struggle with advanced mathematics courses and new at-risk indicators for when students successfully reach 30 credit hours.

References

Burgher, E. K., Davis, M. (2013). Predictive analytics for student retention: group vs. individual behavior. *College & University*, 88(4), 63-72.

CollegeBoard. (2014). "How to convert your GPA to a 4.0 scale." How to convert your GPA to a 4.0 scale. Retrieved April 9, 2014, from http://www.collegeboard.com/html/academicTracker-howtoconvert.html.

Duan-Barnett, N., St. John, P. E. (2012). Constrained curriculum in high schools: the changing math standards and student achievement, high school graduation and college continuation. *Education policy analysis archives*, 20(5), 1-25.

Koppius, R. O., Shmueli, G. (2011). Predictive analytics in information systems research. *MIS Quarterly*, 35(3), 553-572.

Singell, D. L., Waddel, R. G. (2010). Modeling retention at a large public university: can at-risk students be identified early enough to treat? *Springer Science+Business Media 51*, 546-572. doi: 10.1007/s11162-010-9170-7

Strategic Research & Analysis. (2012). Office of Institutional Research. Retrieved from Fact Books: http://www.valdosta.edu/administration/emc/institutional-research/data-and-reports/fact-books.php

StudyPoints. (2014). "How to interpret the SAT to ACT conversion chart." SAT to ACT conversion chart: convert your SAT scores to ACT scores. Retrieved April 9, 2014, from http://www.studypoint.com/ed/sat-to-act-conversion/.

University System of Georgia. (2013, December 9). Educational access and success. Retrieved from Complete College Georgia: http://www.usg.edu/educational_access/complete_college_georgia

Valdosta State University. (2013). Freshman. Retrieved April 7, 2014, from Undergraduate Admissions: http://www.valdosta.edu/admissions/undergraduate/student-information/freshman-admission-requirements.php