

# Predicting the Success of Students Enrolled In SMART Program

By Alana R. McGraw

## Abstract

The START program at The University of Massachusetts Dartmouth invests in students who were once ruled out by college admissions and provides them with an alternate route to college. The objective of this report is to determine what variables are good predictors of student success in their preliminary year before enrollment in their undergraduate programs. Using Mathematica, we determined what variables were good predictors of student success by generating linear regressions and assessing the accuracy of the model's predictions based on psychological, behavioral, personal, and academic variables collected from previous START students. We determined that the workshop variable was the individual variable that was the best predictor of success and students (81.1% accuracy), and more generally, that variables related to responsibility were more likely to be better predictors of success. Based on these results and the lower accuracy of the predictions made using the psychological variables on their own, we would recommend the University seek out other predictive factors or optimize the current tests to search for variables related to responsibility which are likely better predictors of student success.

## Introduction

The successful completion of a college degree is becoming increasingly important for success in the modern world. However, a troubling percentage of students drop out and never receive their degrees. At the University of Massachusetts Dartmouth, the graduation rate is only 59%, just above the 57.93% average. While some of these student are transferring, many of them drop out. Some students leave for reasons that are beyond the control of the University, but some losses are preventable. For this reason, academic institutions have an increasing interest in the factors that influence a students ability to succeed in college (1).

Students with lower traditional quantitative predictors like GPA and SAT scores are considered more at risk of not completing their degrees successfully. The START/College Now Program at The University of Massachusetts Dartmouth is intended to provide students who were once ruled out by college admissions with a non-traditional route into higher education. The program considers that students may not have excelled academically during highschool because of social and personal difficulties and provides them with the opportunity to earn their degree. The program invests a lot of money into students through academic counseling, instruction and individualized attention so it is important to find non-traditional predictors for student's potential in college. One START administrator was interested in what other variables are predictive of student success. In this report, we will determine what variables are good predictors of success for those who are completing a preliminary year before they are accepted into their undergraduate programs. We will then make a recommendation of which variables should and should not continue to be

collected.

## Methods

### *Data*

Data was collected from previous students enrolled in START at University of Massachusetts Dartmouth. Administrators from The College Now/START program collected a wide range variables related to academic performance, personal characteristics, psychological characteristics, and behavior, along with the success of the student by the end of the program. The data set named “college\_prelim\_year” was obtained from the MTH 332 website. Once the data was downloaded, preprocessing began in the excel file. Gender and ethnicity were converted into numerical variables: “M”=0, “F”=1, “Black/African American”=1, “Hispanic/Latino”=2, “Asian”=3, “White”=4, “Two or more races”=5, and “Not specified”=6. All of the blank data cells were changed to zeros.

The data set was then imported into Mathematica. All entries without numerical data were removed from the dataset along with the Fall, Spring and cumulative GPA. Data subsets were also created with the personal, psychological, academic, and behavioral variables. The personal data subset includes “Federal Ethnic Group” (FEG), “Gender” (Gen), “Resident/Commuter” (RvC), and “Athlete” (A). The behavioral data subset includes “Completed Summer Bridge” (SB), “Completed Campus Event Requirement” (CE), “Completed Community Service Requirement” (CS), “Number of Faculty Advisor Meetings Attended” (FA), “Number of Peer Mentor Meetings Attended” (PM), and “Number of Workshops Attended” (W). The academic subset includes “High School GPA” (GPA), “SAT Score” (SAT), and “Pell Grant Eligibility” (PG). The psychological variable included were “Dropout Proneness” (D), “Predicted Academic Difficulty” (AD), “Educational Stress” (ES), “Receptivity to Institutional Help” (IH), “Receptivity to Academic Assistance” (AA), “Receptivity to Personal Counseling” (PC), “Receptivity to Social Engagement” (SE), “Receptivity to Career Guidance” (CG), “Receptivity to Financial Guidance” (FG), and “Desire to Transfer” (DT). These subsets had their own subgroups where each variable was successively removed. If it’s removal did not decrease the accuracy of the prediction, then we kept it out of the subset and removed the next variable. If it’s removal did decrease the predictor accuracy then it remained in the subset while the subsequent variables were removed. This was repeated for all 4 data subsets in order to determine which factors were good predictors of student success. All the data subsets used were described in Table 2.

### *Logistic Regression*

We used the LogitModelFit function in order to construct a fitted binomial logistic regression for each of the subsets described in Table 2, where the dependent variable for each model is the likelihood of success in the START program. We then used that model to generate a list with the predicted probability of success generated from the model and whether the individual student passed their courses in the START program. Then we calculated the accuracy of the prediction; if the predicted probability of success was above 50% and the student was successful or the predicted success was below 50% then those were considered to be accurate and they were put into their own list. A ratio was made of the number of accurate predictions over the total number of predictions and multiplied by 100 in order to generate the percent

accuracy of the prediction. This was repeated for all data subsets described in Table 2. After successively removing variables, we singled out the variables that caused a decrease in the percent accuracy of the predictors and combined them to identify a grouping of variables that would lead to more accurate predictions of success and then repeated the logistic regression and measure of percent accuracy procedure with these predictors.

## Results

We found that variables related to behavior resulted in the highest percent accuracy at 84.9% (Table 1). We also found that the number of workshops alone predicted student success correctly 81.1% of the time (Table 1). We also found that when the ethnicity variable was removed from the personal predictors, the percent accuracy of the predictions dropped from 66% to 37.7% (Table 1). When each variable was dropped from its distinct subsets, we saw that this resulted in decreases in prediction accuracy for some variables. This was true for the behavioral variables CE, W, FA, and PM, the psychological variables AD, AA, and SE, and the personal variables FEG and Gen (Table 2). These variables from the different categories were combined into their own subset and the variables were removed until the highest percent accuracy was generated at 89.6% accuracy. This predictor was a combination of behavioral and psychological factors. The full description of the results from all of the models can be found in Table 2.

Table 1. Summary of key findings from the predictive models. It includes the accuracy of predictions based on all variables in the same model, all the variables in the 4 subsets (personal, behavioral, academic, and psychological factors), the factors with a strong influence on the accuracy of the subset prediction and the most favorable combination of predictors.

Data Subset	% Accuracy
All Variable	100
<b>All Behavioral</b>	<b>84.9</b>
All Psychological	68.9
All Academic	68.9
All Personal	66.0
+W	81.1
+FEG+Gen+RvC+A	66.0
+Gen+RvC+A	37.7
<b>+AD+IH+AA+CE+FA+PM+W</b>	<b>89.6</b>

Table 2. Summary of all findings from the logistic binomial regressions generated with the behavioral, psychological, academic, and personal variables. It includes the accuracy of predictions of all variables and all of the different combinations of variables tested including the behavioral, psychological, academic, and personal variables.

Data Subset	% Accuracy
All Variables	<b>100</b>
Behavioral	
+SB+CE+CS+FA+PM+W	84.9
+CE+CS+FA+PM+W	85.8
+CS+FA+PM+W	84.0
+FA+PM+W	84.0
+FA+PM	<b>13.2</b>
+PM+W	83.0
+W	81.1
Psychological	
+D+AD+ES+IH+AA+PC+SE+CG+FG+DT	68.9
+AD+ES+IH+AA+PC+SE+CG+FG+DT	72.6
+ES+IH+AA+PC+SE+CG+FG+DT	69.8
+IH+AA+PC+SE+CG+FG+DT	70.8
+AA+PC+SE+CG+FG+DT	71.7
+PC+SE+CG+FG+DT	68.9
+SE+CG+FG+DT	69.8
+CG+FG+DT	68.9
+FG+DT	68.9
+DT	68.9
Academic	
+GPA+SAT+PE	68.9
+GPA+SAT	68.9
+PG	68.9
+SAT	68.9
Personal	
+FEG+Gen+RvC+A	66.0
+Gen+RvC+A	<b>37.7</b>
+FEG+RvC+A	66.0
+FEG+A	65.1
+FEG+Gen+A	66.0
+FEG+Gen	68.9
+FEG	68.9
+Gen	68.9
Combinations	
+FEG+Gen+AD+IH+AA+CE+FA+PM+W	87.7
+AD+IH+AA+CE+FA+PM+W	<b>89.6</b>

## Discussion

The objective of this report is to show the College Now/START program administrators what factors are good predictors of student success, and assist them in determining what variables they should continue to measure. We found that the behavioral variables has the highest percent accuracy (84.9%), with the number of workshops attended being the most accurate individual predictor of success (81.9%). We also found that ethnicity influenced the accuracy of the personal variables, decreasing the personal variables accuracy of predicting success from 66% to 37.7% when it was removed. While both of these percent accuracies are low, the change in accuracy does indicate that ethnicity plays some role in the prediction of student success. The strongest predictor of success (89.6%) was a combination of psychological and behavioral factors, including the student's predicted academic difficulty, their willingness to receive help from the institution, and their willingness to attend various meetings that could help improve their performance. Generally, these variables indicate that the students are responsible individuals, who are open to using the resources the university provides. Beyond the behavioral variables, the academic, personal, and psychological variables usually predict student success with about 70% accuracy, which is considerably less than the accuracy of predictions from the behavioral variables. Based on these results, we determined that the existing variables that measure students before they attend UMass, including the psychological factors, do not work very well as predictors. We suggest that it would be worth testing out new variables that have the potential to measure student responsibility before they join the START program.

There are limitations to the models that we generated in this report. The first is that we do not know very much about the context of the data, including when the data was taken and whether there are students from different years included in the data set; if the data comes from more years, it may apply to a broader population that will make our predictions more broadly applicable. Empty data entries were also treated as zeros which would have an impact on how well the affected variables predict student success. There is also a higher change of the prediction of success being inaccurate when the projected probability of success is around 50%; this influences the calculation of the accuracy of the prediction of success. Future experiments should be undertaken to determine whether there are other variables, including different psychological tests, that would be more effective at predicting student success by focusing more on variables that correlate to student responsibility. Further analysis should also be used to determine whether the differences in success between groups is significant and whether higher or lower values of the variables lead to student success.

In conclusion, we recommend that the START administrators investigate variables related to student responsibility to better predict the chance of student success before their enrollment into the program.

## References

1. Pritchard, M. E., & Wilson, G. S. (2003). Using emotional and social factors to predict student success. *Journal of college student development*, 44(1), 18-28.