Examining the Impact of Pre-Collegiate Factors and Course Difficulty on Collegiate Success

Manchester University

Thomas Blake

Sreenath Majumder

Midwest Economic Association Meetings

Chicago, Illinois

March 22, 2014

Abstract

College administrators are constantly looking for ways to enhance the quality and image of their institutions. One of the best ways to do this is to admit students that will be successful and ultimately graduate from college. This study examines pre-collegiate academic performance, socioeconomic status, course difficulty, and demographic variables to evaluate the most significant predictors of student success (measured by college GPA). The population being examined in this study is 285 students that entered Manchester University as first-year students in the fall semester of the 2008-2009 school year. Key preliminary findings from this research suggest that high school GPA is the best predictor of first-semester college GPA. Standardized test scores also add some predictive power to first-term college GPA when included with high school GPA. High school GPA continues to be a significant predictor of second, third, and fourth-year college GPA, while standardized test scores lose almost all of their predictive power after a student's first year. Results also indicate that first-semester course selection may have a significant impact on both first and second year grades, which raises some interesting policy implications.

Introduction

Students entering their first semester of college are at varying levels of academic preparedness. Some students breeze through their first-semester of classes, while other students struggle to attain passing grades. Identifying the causes of student success and failure is of vital importance to college officials trying to admit students with the highest likelihood of succeeding. In attempting to predict student success (measured by college GPA) a variety of variables were examined for significance. These variables included the academic indicators of high school GPA, SAT/converted ACT scores, and high school quality ratings, as well as socioeconomic and demographic information such as family income, gender, ethnicity, and parent marital status. First-semester course difficulty was also included as a potential determinant of long-term student success.

Data was collected for 395 students who entered Manchester University as first-year students in the fall semester of 2008. Of the 395 students, complete pre-collegiate information was available for 285 of the students. A multiple regression analysis was performed on the data of the 285 students in an attempt to explain the impact of pre-collegiate academic indicators, demographics, socioeconomic information, and course difficulty on first-semester grade point average. Separate regressions were also run to evaluate the pervasiveness of high school GPA, standardized test scores, and course difficulty as predictors of college success across time. College officials can utilize the information presented in the body of this report to formulate more effective criteria for admissions and to target the needs of at-risk students.

Literature Review

Students that withdraw from a college or university present several potential problems for school officials. First, high levels of student failure tarnish the image that higher education institutions work very hard to maintain. Furthermore, a high withdrawal rate means lost revenues and additional costs to recruit students to replace the ones that have withdrawn. According to research conducted by Vincent Tinto, Distinguished University Professor at Syracuse, approximately seventy-five percent of students that withdraw from college leave within the first four semesters, with the greatest number leaving after the first and second semesters (Tinto, 1993). Additionally, first-semester college grade point average is one of the most significant determinants of whether or not a student decides to drop out (McGrath & Braunstein, 1997). With these two pieces of information in mind, it becomes extremely important to be able to predict students' first-term college GPA, so as to minimize the recruitment and admission of students likely to drop out because of poor academic performance.

In a recently published book titled *Crossing the Finish Line: Completing College at America's Public Universities*, William Bowen, Matthew Chingos and Michael McPherson examined the need for standardized testing and questioned its use as a tool in admitting students to college. Based on extensive research, the authors found that high school grades are a much better predictor of graduation rates than standardized tests like the SAT and ACT. They note that even when controlling for the quality of the high school attended, students who earn high grades in high school tend to have high graduation rates and students who earn low grades in high school tend to have low graduation rates. Furthermore, Bowen, Chingos and McPherson believe that the reason high school GPA is a better predictor of college graduation rates is because it

incorporates perseverance, study habits, motivation, and other student traits that cannot be captured by standardized testing (Bowen, Chingos & McPherson).

Model

Economic theory and previous academic research indicate that a variety of factors affect student performance. These include a variety of student characteristics (innate ability, work habits, etc.), demographic variables (ethnicity, gender, family income, etc.), and school characteristics (class size, class difficulty, etc.). The following model, used to predict first-semester college GPA, captures a number of these factors.

$$\widehat{TGPA\_08FA}_i = B_0 + B_1(\text{High School GPA})_i + B_2(\text{SAT/Converted ACT Score})_i + B_3(\text{High School Quality Rating})_i + B_4(\text{Course Difficulty})_i + B_5(\text{Parent Marital Status})_i + B_6(\text{Gender})_i + B_7(\text{Family Income})_i + B_8(\text{Ethnicity})_i$$

The variables included in the model are described in detail in the methodology section below.

Methodology

Participants

Participants in this study were 285 first-year students who enrolled at Manchester University during the Fall 2008 semester (2008 Cohort). The mean age of the 2008 Cohort on the first day of the Fall 2008 semester was 18.80, with a range between 17.50 and 20.67 years of age. Of the 285 students, 51.6% were females and 48.4% were males. Additionally, participants

had a wide array of standardized test scores and high school GPAs. SAT/converted ACT scores[1] ranged from 910 to 2300, with a mean score of 1518 and a standard deviation of 228. High school GPAs had a range between 1.90 and 4.50 with a mean GPA of 3.27 and standard deviation of 0.52. The large amount of variance in high school GPAs and standardized test scores shows that the students in the 2008 Cohort entered Manchester University with greatly varying levels of academic potential.

Variables

The first type of variables collected were those related to students' academic performance prior to entering Manchester University. These variables were high school GPA, high school quality rating, and SAT/converted ACT scores. Data for high school GPA was collected by the Manchester University Office of Admissions as students applied for college. High school GPAs were taken directly from student's transcripts and normalized using the equation[2] $x/y = z/4.0$. Due to the fact that high school GPAs were collected at different points in time and that some GPAs were weighted and others were not, the data is not perfectly comparable across students. However, this is the most relevant high school grade data that could be collected. High school quality rating is a continuous variable ranging from zero to one, with one being the highest quality a high school can achieve. The quality rating was calculated by an outside consulting firm used by Manchester University, and was determined using a variety of academic indicators from each high school. High school quality ratings in the 2008 Cohort ranged from a low of 0.1234 to a high of 0.9820 with an average rating of 0.6412.

[1] The maximum possible SAT/converted ACT score is 2400

[2] When the Manchester University Office of Admissions received a student's high school transcript they converted each GPA to a 4.0 scale. For example a 9.0/12.0 would be converted to a 3.0/4.0 using the formula $x/y = z/4.0$. If a 4.0 scale was already used by the high school, no conversions were made by the college.

SAT/converted ACT score was calculated as the highest SAT score or converted ACT score a student achieved. ACT scores were converted into SAT scores using ACT-SAT concordance charts developed by the College Board (College Board). Standardized tests like the SAT and ACT typically have a high correlation[3] with high school GPA. In this data set, the correlation between the two variables was 0.5578. The general rule of thumb is that when two independent variables have a correlation greater than 0.70, one of the variables should be omitted to prevent results from being misinterpreted due to high levels of multicollinearity[4]. Because the correlation of the two variables was close to 0.70, two separate regressions were run to examine the change in robust standard errors when the SAT/converted ACT score variable was included and when it was excluded. The change in the robust standard errors and t-scores was small when adding in the standardized test scores. Therefore, the SAT/converted ACT score variable was included in the model.

Socioeconomic status and personal/family demographics variables were also included in the model. These variables included total family income, gender, ethnicity, and parent marital status. Total family income is a continuous variable that ranges from -$67,928 to $521,972, with a mean of $73,454. It was calculated as the sum of parent's adjusted gross income plus the student's adjusted gross income. Gender is a binary variable where female = 0 and male = 1. Ethnicity is also a binary variable where 0 = white, non-Hispanic and 1 = other. As a predominately white, non-Hispanic college (90.53% of the sample was white, non-Hispanic), trying to compare the effects of ethnicity is difficult because there are so few observations for

[3] Correlation measures the relationship between two variables. Correlation values range from 0-1, with a correlation of 0 implying no relationship between two variables and a correlation of 1 implying a perfect relationship between two variables.

[4] Multicollinearity occurs when two or more variables are highly correlated. Coefficient estimates may change drastically with small changes in the model or data when multicollinearity is present. Multicollinearity does not affect the predictive power of the model as a whole, but may significantly alter the coefficients of individual predictors.

each ethnicity other than white, non-Hispanic. Therefore, the ethnicity variable was set up to measure the difference between being a minority or majority student in terms of ethnicity. Parent marital status is also a binary variable where 1 = parents are married and 0 = parents are not married. 77.5% of the students in the 2008 Cohort came from families with married parents.

Selecting courses is a very important decision for all college students and it can greatly impact the grades a student receives. Therefore, an index was created to evaluate the difficulty of the various courses at Manchester University. This index was used to estimate the impact of course difficulty on current and future semester grades. A course difficulty value was calculated for each course offered in the Fall 2008 semester by using the following formula[5]:

$$\frac{\textit{Average GPA of students in course X}}{\textit{Average GPA of students in course X in all classes except for course X}}$$

This equation examines how students perform in a given class compared to how they perform in the rest of their classes. If the index value for a course is greater than 1.0 it means that students do better in that course than their average in their other courses. Conversely, if the index value of a course is below 1.0, it signifies that students do worse in that course than their average in their other courses. For example, if students in course X earn an average GPA of 3.0, but earn an average GPA of 2.0 in all of their classes except for course X, then the index value for course X is 1.50, signifying an easier than average class. It is important to note that this equation considers the difficulty of a course only by the academic performance of students, while no weight is given

[5] This equation was devised by Professor Matthew Hendryx of Manchester University. This measure of course difficulty controls for academic differences in the students in a class. This prevents a course from being considered easy just because the course had a large number of very academically strong students.

to the content or workload of the course. The average course difficulty index value for students in the 2008 Cohort was 1.02, with a minimum of 0.84 and a maximum of 1.40.

Technique

The data used in the regressions was collected from three primary sources: Office of the Registrar, Student Financial Services, and Office of Admissions. After combining all of the given data into a usable form, standard multiple regression techniques were used to estimate a model to predict first-term college grade point average. The results were corrected for heteroskedasticity[6] by running a robust standard regression.

Results

After running a robust, standard multiple regression, the results indicate that both SAT/converted ACT scores and high school GPA are significant at the 99% confidence level at predicting first-semester college GPA. Both variables have a positive coefficient, meaning that increases in SAT/converted ACT scores and high school GPA both lead to predicted increases in first-semester college GPA. The elasticity of high school GPA is 1.331 meaning that a one percent increase in high school GPA will lead to an estimated 1.331% increase in first-semester college GPA. The elasticity of SAT/converted ACT scores is 0.396 which means that a one percent increase in SAT/converted ACT scores will lead to a 0.396% increase in predicted first-semester college GPA. These elasticities show that high school GPA is a better estimator of how a student will do in his or her first semester than standardized test scores. High school GPA has a correlation of 0.631 with first-semester GPA, while the correlation between SAT/converted ACT scores and first-semester GPA is 0.461. These correlation values show that both high school

[6] Heteroskedasticity occurs when there is not constant variance in the error terms as the value of an independent variable changes. If a model is heteroskedastic and the regression does not take this into consideration, standard errors will be too small and confidence intervals will be too large, creating unreliable results.

GPA and standardized test scores are highly associated with how well a student performs in college. The variables of high school quality rating, parent marital status, gender, and ethnicity were not statistically significant in predicting first-term college GPA. The regressors used in this model explain 49.20% of the variance in first-semester GPA. Table 1 visually represents the results of the regression, including the estimated coefficients, robust standard errors, and the R-squared value.

Table 1: Dependent Variable – Fall 2008 College GPA

Independent Variable	Model 1	Model 2	Model 3
High School GPA	0.947***	0.813***	0.938***
	(0.070)	(0.087)	(0.107)
SAT/Converted ACT Score		0.001***	0.001***
		(0.000)	(0.000)
High School Quality Rating			0.500**
			(0.223)
Course Difficulty			2.512***
			(0.419)
Parent Marital Status			0.098
			(0.087)
Gender			-0.032
			(0.067)
Family Income			-0.000
			(0.000)
Ethnicity			0.112
			(0.126)
Total R^2	0.400	0.415	0.492
Observations	285	285	285

Robust standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

Separate regressions were also run to examine the pervasiveness of high school GPA and standardized test scores as predictors of college GPA across time. The results show that high school GPA consistently outperforms standardized tests scores at predicting college grades and that over time SAT and ACT scores tend to lose the power to significantly predict how a student will perform. The most probable reason for this is that high school GPA captures intrinsic characteristics such as study habits and motivation that standardized tests like the ACT and SAT are simply unable to capture. This agrees with the assertions made by Bowen, Chingos, and McPherson in their book titled, *Crossing the Finish Line: Completing College at America's Public Universities*. Results from the regressions of semester GPAs on standardized test scores and high school GPA are shown in Table 2 below.

Table 2: Dependent Variables – GPA by Semester

Independent Variable	Fall 2008	Spring 2009	Fall 2009	Spring 2010	Fall 2010	Spring 2011	Fall 2011	Spring 2012
High School GPA	0.744***	0.552***	0.731***	0.644**	0.463***	0.228	0.638***	0.380**
	(0.138)	(0.123)	(0.174)	(0.309)	(0.166)	(0.189)	(0.151)	(0.174)
SAT/Converted ACT Score	0.001***	0.001***	0.000	0.000	0.001*	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
High School Quality Rating	0.539**	0.178	0.082	-0.294	-0.309	0.132	0.854**	0.206
	(0.223)	(0.274)	(0.288)	(0.341)	(0.330)	(0.385)	(0.402)	(0.340)
Course Difficulty	1.695***	0.213	-0.473	-1.401**	1.022	-0.023	-0.998*	-0.388
	(0.448)	(0.548)	(0.633)	(0.678)	(0.892)	(0.805)	(0.589)	(0.654)
Parent Marital Status	0.074	0.062	0.144	0.174	0.188	-0.047	-0.014	0.050
	(0.089)	(0.090)	(0.097)	(0.147)	(0.167)	(0.162)	(0.112)	(0.116)
Gender	-0.065	-0.120	-0.022	0.064	-0.028	0.181	-0.083	-0.248**
	(0.073)	(0.080)	(0.099)	(0.100)	(0.119)	(0.115)	(0.094)	(0.104)
Family Income	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ethnicity	-0.145	-0.140	-0.240*	-0.368	-0.289	-0.926***	-0.133	-0.172
	(0.113)	(0.160)	(0.144)	(0.329)	(0.216)	(0.333)	(0.169)	(0.192)

Total R^2	0.525	0.387	0.283	0.118	0.173	0.130	0.266	0.248
Observations	163	163	163	163	163	163	163	163

Robust standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

As can be seen in Table 2, high school GPA slowly loses its ability to estimate college GPA over the course of time since the coefficient estimates diminish in magnitude. However, it still has significant explanatory power even during a student's final years of college. Standardized test scores, on the other hand, tend to lose most of their explanatory power after just two semesters of college. These results suggest that college administrators should consider giving more weight to high school GPA than standardized tests when admitting new students.

In addition to examining the explanatory power of standardized tests and high school GPA over time, course difficulty of first semester classes was also examined to determine its impact on grades each semester. The purpose of this study was to establish whether or not a student's first-semester course selection impacts his or her success in subsequent semesters. In order to do this, each student's grade point average (Fall 2008 through Spring 2012) was regressed on his or her average Fall 2008 course difficulty, while controlling for pre-collegiate academics, socio-economic status and demographics. The findings from the regression show that taking easy courses during your first semester of college leads to short-term success, but may harm academic performance in the long-run. Conversely, taking hard courses during your first semester may lead to poorer performance at first, but success in the future. The results of these regressions can be seen in Table 2.

The trends in significance and the coefficients for course difficulty indicate that taking easy courses during your first semester of college will boost your first-semester grades. However, the effects of taking easy classes fade during the second semester and turn negative

during the third and fourth semesters of college. Then, the effect seems to almost completely disappear by the end of the third year. These results indicate that students taking easy courses during their first semester may not be adequately prepared when it comes time to take more difficult, upper level courses. Faculty advisors may want to consider the difficulty levels of the classes taken by their advisees during their first semester at college in order to minimize the possibility of their students being underprepared for courses in the future.

Conclusion

After analyzing the results obtained for the 2008 Cohort from Manchester University and comparing the results to previous studies, it is reasonable to conclude that high school GPA is by far the best pre-collegiate predictor of first-semester college GPA. This is evidenced in the Manchester University data by the high correlation coefficient between high school GPA and first-semester college GPA and an elasticity of 1.331. SAT/converted ACT score is also significant at the 99% confidence level in predicting first-term GPA although its elasticity is much lower at 0.396. Family income, high school quality rating, ethnicity, and parent marital status are not significant factors in determining first-term college GPA per the data for the 2008 Cohort. Furthermore, it appears that high school GPA is able to predict the success of a student in college even after a student is several years into his or her studies, whereas standardized test scores cannot. The pervasiveness of high school GPA as a predictor of college success indicates that high school GPA captures student characteristics that cannot be evaluated by standardized tests. Possible traits captured by high school GPA include work ethic and determination to succeed.

Course selection also plays a role in how well a student performs. Advisors should be aware of this and try to help their advisees select a mix of classes that will challenge, but not

overwhelm them. It would be a great disservice to let first-year students believe that they are adjusting well to college life and then have them learn that they are not at all prepared for their second year of college because their courses did not challenge or prepare them enough during their first year.

From the results obtained in this regression and prior studies, there are several policies that colleges should consider adopting or revising to more fairly and effectively enroll students capable of succeeding at the college level. First, colleges should put the heaviest weighting on high school GPAs when looking to admit new students. Standardized test scores should also factor into the admissions process because they appear to have an impact on first-term college GPA, although not nearly as much as high school GPA. The use of standardized test scores is more important at schools with a large number of applicants and a very competitive application process. This is the case because many applicants have high grade point averages and the college must differentiate between the students by some means other than high school GPA. However, at colleges with fewer applicants and a wide array of high school GPAs, standardized test scores should not be the primary factor in admitting students.

It is advantageous that the high school GPAs and standardized test scores of the 2008 Cohort had such a wide variance. Having high variance in an independent variable leads to a better prediction of its effect on the dependent variable. It would be much more difficult to estimate the effect of high school GPA or standardized test scores on college GPA at an Ivy League school because almost every student admitted would have very high GPAs and standardized test scores. While my model of the determinants of a first-semester college students' grade point average explains 49.20% of the variance observed in first-semester college grades, other factors not included in the model may also be important determinants of academic

performance. Other possible determinants may include peer effects through friends and classmates, classroom and teacher effects not captured in my measure of course difficulty, and time constraints (work and social activities) that limit attendance and study time. Given these limitations, my results indicate that course difficulty (a factor often overlooked in the studies of the determinants of academic performance) should be considered in future empirical studies when addressing the issue of academic performance. In addition, my findings raise possible issues of concern (or potential areas of opportunity) for admissions and academic advising.

References

"ACT - SAT Concordance Tables." *College Board*. n.d. Web. 20 Nov. 2011. <http://collegeboard.com>.

Bowen, William G., McPherson S. Michael, and Matthew M. Chingos. *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton, NJ: Princeton University Press, 2009. Print.

Farrar, Donald E., and Robert R. Glauber. "Multicollinearity in Regression Analysis: The Problem Revisited." *JSTOR*. The MIT Press. Web. 20 Mar. 2012. <http://www.jstor.org/discover/10.2307/1937887?uid=3739664>.

Kobrin, Jennifer L., et al. "Validity of the SAT for predicting first-year college grade point average." *College Board* 2008.5 (2008): 10. *College Board*. Web. 23 Nov. 2011.

McGrath, M. & Braunstein, A. (1997). *The Prediction of Freshmen Attrition: An Examination of the Importance of Certain Demographic, Academic, Financial, and Social Factors*. College Student Journal, 31, 396-408.

Tinto, V. (1993) *Leaving College; Rethinking the Causes and Cures of Student Attrition* (2nd ed.). Chicago: University of Chicago Press.

White, Halbert. "A Heteroskedasticity - Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *JSTOR*. The Econometric Society. Web. 20 Mar. 2012. <http://www.jstor.org/discover/10.2307/1912934?uid=3739664>.