FORECASTING FIRST-TERM COLLEGIATE SUCCESS FROM PRE-ENROLLMENT INFORMATION

Donald I. Price[1] and Gregory B. Marsh[2]

Abstract

Admissions officials are asked to make decisions about individual students based upon the information that can be known prior to enrollment. In the present study we show that first-term success of previous students from a given high school can be a significant predictor of first-term success of individual students after accounting for other variables; such as standardized test scores and class standing; that are also known prior to initial enrollment.

Introduction

In an era of tight budgets, when legislatures are using retention and graduation rates to make funding decisions, it is critically important for institutions of higher education (IHEs) to identify first-year students who are well matched to the institution and, therefore, likely to succeed. However, admissions officers at IHEs are forced to make decisions when they possess incomplete information. Student class percentiles, secondary school grades, and standardized test scores are known prior to enrollment but many institutions still find that a substantial proportion of their first-time students do not remain at the institution beyond the first year. Many admissions officers contend that the percentile standing of a student is more meaningful when they know something about quality of the student's high school. It is, therefore, a potentially a useful exercise to combine information from the past first-time-in-college performance of various high schools with the individual student's percentile in high school. When doing this for a single IHE it is important to remember that the IHE's calculation of student success by high school may not reflect the overall quality of the high school. An IHE such as the study institution may receive the better students from one high school and not be seriously considered by the better students of another high school. What is important from the standpoint of retention and eventual graduation is the quality of students who are actually choosing the institution.

Literature in education suggests that one measure, first grade point average, is the singular most important predictor of student retention. Admissions decisions, of course, must be made without knowledge of first grade point average. We propose to combine an aggregate of past

---

[1]   Professor of Economics, Department of Economics and Finance, Lamar University, Box 10045, Beaumont, TX 77710.

[2]   Director, Office of Institutional Research and Reporting, Lamar University, Box 10073, Beaumont, TX 77710.

first-term success of students from each high school with the class percentile ranking of the student to better evaluate potential first-term success after we first correct for the effects of other variables that can be known prior to initial enrollment at the IHE. We approach problem in two ways.

First, we use the standard rank percentile of the student and past first time GPA by high school as separate independent variables. Second, at the suggestion of an anonymous referee, we create a variable called success-adjusted student percentile (SASP) which adjusts student percentiles for past student success at the IHE. These values are measured, as are the other independent variables used in the study, using data that are available prior to the student's initial enrollment. Because our objective is to identify students, prior to admission, who are likely to be successful measureable post-enrollment variables are not considered as part of the models.

The study is limited to the first-time-in-college students at a single institution, Lamar University. The confinement of the study to a single institution's market has the advantage of limiting the variations in supply and demand among college markets that are characterized by a variety of instructional missions. The use of single institution limits any generalizations from the results but it is useful in that it provides a methodology that can be applied by other IHEs to their specific situations.

The paper begins with examination of retention literature. We will then present a model to explain student first year success measured by first-term GPA and analyze the results. Finally, we will summarize and discuss the results and suggest areas of future study.

Literature Review

The quality of the product of our public school systems is known to vary considerably from one school to next. Measurement of output quality has focused on student test scores and earnings. In the present study, we look at one aspect of quality, the first-term success of past students from each high school, as a predictor of individual student success at a single IHE. The measurement links high school quality with what education literature finds to be a key factor in student retention and graduation.

Retention studies in education suggest that initial grade point average (GPA) is the single most important factor whether students return for their second year in college. If pre-enrollment factors can be identified which explain first GPA of individual students then it may be possible to better predict initial success at institutions of higher learning.

Support for the importance of first-GPA as measure of success can be found frequently in the literature. McGrath and Braunstein (1997) discovered that of more than 20 academic, demographic, and financial variables studied, the single most important variable in predicting retention was first semester GPA. Allen (1999) observed that among both minorities and non-minorities, freshman GPA was the strongest predictor of retention behavior. Desjardins, Kim and Rzonca (2003) confirmed the importance of first semester GPA, finding that as the GPA increases, the chances of student attrition decreased. Ishitani and DesJardins (2002) concurred, as did Kiser and Price (2007) who found that a student's chance of leaving decreased as their first-year GPA increased. Murtaugh, et. al. (1999) further confirmed the importance of college GPA, finding that freshmen with a first quarter GPA between 0.0 and 2.0 had a probability of returning of 57.2%, while those at the highest GPA range between 3.3 and 4.0 had a 90.7% probability of being retained.

In this paper, we show that a variable measuring the prior first-term performance of a student's high school is one of several significant pre-enrollment variables explaining the entering student's first-term GPA.   Specifically, we calculate an aggregate GPA for each individual high school in the study based upon the past first-year performances of their students in a single market, that of the study institution and use it as one predictor of the expected initial student success of students from each of those high schools. We also use the first-time GPA of each student's high school to adjust the individual student percentile in a measure we describe as success-adjusted student percentile (SASP). SASP reflects both the student's standing in his or her high school and that high school's past success at the study institution.

Materials and Methods
The model used is an OLS estimate of the factors thought to influence first semester GPA of full-time, first-time-in-college students. The explanatory variables measure data values that are known prior to the student's entry to the study institution. The model employs variables specific to each student, specific to the high school from which the student graduated, and a measure of economic conditions in the areas from which the students come. Data were collected for Fall FTIC students who first enrolled at the study institution between the fall of 2004 and the fall of 2009. The model is specified below.

$GPA100 = f$ (STUPER, GPAbyHS, test score, LOGMILES, HHINC, CLASSSIZE)

$GPA100 = f$ (SASP, GPAbyHS , test score, LOGMILES, HHINC, CLASSSIZE)

TABLE 1: VARIABLES	
Variable	Variable Description
GPA100	Individual Student First Semester GPA times 100
SASP	Student Percentile multiplied by the ratio of a high schools overall 1st GPA to overall GPA of most successful high school
GPAbyHS	Combined FTIC GPA for previous students from a particular High school times 100
STUPER	Individual Student Percentile in Graduating Class
SAT	Individual Student SAT Verbal plus SAT Quantitative
ACTCOMP	Individual Student ACT Comprehensive Score
LOGMILES	Natural log of distance in miles from student's high school to study institution
HHINC	Median Household Income in zip code where student's High School in Located (in thousands of $)
CLASSSIZE	Number of Graduates in Student's Senior Class

The study institution prefers SAT scores in the admissions process but will also accept ACT scores. The model was estimated using data sets that include either SAT or ACT scores. The data set for students submitting SAT scores is several times the size of the one for students submitting ACT scores. Because some students submit both, some students are included in both data sets. Variables used in the study are described in Table 1.

Each of the independent variables measure characteristics of a student's background that can be known at the point of the student's initial enrollment.

Student percentiles, class size and test scores are collected as part of the admissions process and were obtained from the study institution's records. Distances between high schools and the study institution were calculated as distances between the zip code addresses. The income level data are measured by the median household income for the high school zip code and were collected from ERSI's *Community Sourcebook America.*

The data to calculate SASP, the student-adjusted student percentile, were collected from the records of the study institution. First-term GPAs by high school of the entering students were calculated for each of the high schools in the study. The study was limited to high schools that contributed the largest number of semester credit hours by FTIC students. The ratio of a high school's GPA to the GPA of the most successful high school was used as an index to adjust student percentiles based on past success of student from that school. The variable SASP was calculated as:

$$SASP = (GPA_{HS}/GPA_{THS}) \times STUPER$$

where GPA_{HS} is a the average first-term GPA for students from a particular high school in the past, GPA_{THS} is the average first-term GPA of students from the high school with the highest average first-term GPA in the past, and STUPER is the student's percentile at his or her high school.

Values of SASP may vary from 100, for the valedictorian of the high school which has the highest first-term GPA, to values approaching zero when students who have low percentiles and/or are from poorer performing high schools. The adjustments create a measure that is specific to the IHE and should be understood as such. SASP is not a measure of the overall quality of any high school. Instead we are solely interested in measuring the quality of product directed to a single market; i.e., the study institution. For example, the SASP measure could be significantly different for two similar schools if one typically sends its better students to the study institution and the other does not. It is a measure of student success for those entering the particular IHS and cannot be generalized beyond that. It can, nonetheless, be a valuable piece of information for the study institution and is a measurement that can be replicated and used by other institutions in evaluating the likely success of their first-time students.

It is generally expected that FTIC students from high schools that have had previous success at the study institution will be more successful than those who enter from high schools whose previous first-time students have had less success. Standardized test scores are expected to be positively associated with first-term success and the alternative measures student percentiles are both expected to be positively associated with student first-term GPA.

A positive relationship is also expected between first-term GPA and the measure of economic background, median household income. Individual student data were not available for household income so the median incomes for zip codes where their high schools are located were

used. The distance from home variable, LOGMILES, was used to capture the impact of 'freshman homesickness' and, as such, was expected to have a negative relationship with GPA. To capture effects of high school size on first term success, we added the size of the student's high school class. No particular relationship was hypothesized. Positive, negative and even bell-shaped relationships were considered possibilities.

Descriptive statistics for the data sets appear in Tables 2 and 3. GPA100 is the dependent variable in all equations.

TABLE 2: SAT DESCRIPTIVES					
Variable	n	Minimum	Maximum	Mean	Std. Dev.
GPA100*	3568	0	400	261.43	99.04
SASP	3658	1.9513	100	56.92	19.41
GPAbyHS	3658	141.99	371.60	258.30	40.03
STUPER	3568	2	100	69.22	20.49
SAT	3568	480	1540	947.17	161.84
LOGMILES	3568	1.6094	5.79	3.42	0.94
HHINC	3568	20.93	124.68	52.71	15.85
CLASSSIZE	3568	12	2885	356.44	210.89

*Dependent Variable

TABLE 3: ACT DESCRIPTIVES					
Variable	n	Minimum	Maximum	Mean	Std. Dev.
GPA100*	936	0	400	252.14	103.99
SASP	936	0.8922	99	53.92	19.15
GPAbyHS	936	134.76	371.60	234.66	36.32
STUPER	936	1	100	68.63	21.05
ACTCOMP	936	8	33	18.96	4.32
LOGMILES	936	1.6094	5.79	3.81	0.93
HHINC	936	20.93	124.68	50.79	17.96
CLASSSIZE	936	16	1135	361.34	233.28

*Dependent Variable

Results and Conclusions

Four versions of the model were tested, two each with the SAT and ACT data bases. Each of the data bases was analyzed using alternative ways of evaluating previous success of students by high school. Each was first analyzed using standard student percentiles and the GPAbyHS variables. Each data base was then analyzed using the SASP measure along with the GPAbyHS. Since GPAbyHS is used in the calculation of SASP there was some concern that this could lead to multicollinearity. However, tests for multicollinearity did not indicate that it was a problem.

The results of the regression estimates appear in Table 4. The percentile measures, standardized test scores, high school class sizes and the first-time-in-college GPAs by high schools are positively related to the individual students' first-time GPAs across all four models. Household incomes exhibited a positive relationship in the SAT models but not the ACT models and the homesickness variable was significant in only one of the equations and then with a positive rather than the predicted negative relationship. The results clearly support the positive influence of student percentiles, whether adjusted for past success or not, on first-time term student performance. The overall explanatory value of the models differed very little according to which of the percentile measure was used. It is also clear the past first-term success of students from his or her high school is a significant predictor of an individual student's first-term GPA.

The results support earlier findings indicating that standardized test scores are good predictors of initial success in college. Both SAT and ACT show strong positive relationships. The classsize variable also clearly supports the notion that students from larger high schools are more successful initially.

TABLE 4: REGRESSION RESULTS				
Models				
	SAT1	SAT2	ACT1	ACT2
SASP		2.170		2.600
		(.092)**		(.191)**
GPAbyHS	.793	.308	.813	.221
	(.045)**	(.046)**	(.095)**	(.098)*
STUPER	1.873		2.200	
	(.077)**		(.154)**	
SAT	0.078	0.077		
	(.010)**	(.011)**		
ACTCOMP			2.685	2.652
			(.811)**	(.825)**
LOGMILES	3.595	2.636	1.704	1.358
	(1.654)*	(1.660)	(3.325)	(3.354)
HHINC	0.492	0.447	0.332	0.270
	(.111)**	(.111)**	(.197)	(.198)
CLASSSIZE	0.018	0.017	0.031	0.030
	(.008)*	(.008)*	(.015)*	(.015)*
CONSTANT	-191.741	-52.954	-175.029	-19.944
	(15.204)**	(14.5)**	(30.17)**	(28.428)
R2	0.324	0.318	0.330	0.318
n	3568	3568	936	936

Standard errors in parentheses
**Significant at the .01 level, *Significant at the .05 level.

It is not clear why the household income variable was significant for only the SAT models. The variable was the median household of the zip code containing the student's high school. It may be that the smaller ACT data base contained too few high schools to exhibit the same variation that the SAT data base did but we are not sure why the data bases show a different result.

The most surprising result was from the LOGMILES variable, the 'homesickness factor.' We had expected a negative relationship but the variable was only significant in the one model and was positively related to first-term GPA in that case. This result may be associated with a characteristic of the study institution. The study institution was historically a commuter school which had only about 15% of the student population living on campus during the study period. It may be that in a school where so few students are from out of the area that there simply is little measurable homesickness effect.

Overall, the models were able to consistently explain about one-third of the variation in first-term GPA. The analysis consciously excluded variables that could not be measured until after initial enrollment. The objective of the study was to identify pre-enrollment variables that would allow the IHE to make better admissions decisions. A more complete explanation of first-year success would include factors that reveal themselves only after the student is enrolled. Obviously, some portion of the explanation of success is eliminated when measurement is limited to pre-enrollment factors. There is room for future analysis using post-enrollment factors that can provide a more complete model first term GPA.

Certain useful data that were unavailable at the study institution may be available at other IHEs. For example, while the study institution does use individual secondary school GPAs in the admissions decision, those data are not entered into student records and were, therefore, not available to the authors. Where these data are available they may provide a stronger explanation even when restricting the study to pre-enrollment data.

Another potential data limitation is our measure of household income. Data for household income of individual students were not available. We used the median household income for the zip code of the students' high schools to measure income level. Individual data are available for students who apply for financial aid but not for other students. It would be an interesting study to determine whether the financial aid group would produce results similar to those of the larger group studied here and to observe whether greater variation associated with individual students could produce higher coefficients of determination for the equations used in the model.

References

Allen, D. 1999. "Desire to Finish College: An Empirical Link between Motivation and Persistence." *Research in Higher Education*, 40, 461-485.

Akin, John S., and Irwin Garfinkel. 1997. "School Expenditures and the Economic Returns to Schooling." *Journal of Human Resources,* 12(4), 460-81.

Community Sourcebook America, 2007 Edition, ESRI.

DesJardins, S. L., D. Kim and, C. S. Rzonca. 2003. "A Nested Analysis of Factors Affecting Bachelor's Degree Completion," *Journal of College Student Retention,* 4, 407-435.

Heckman, James J. 1995. "Lessons from the Bell Curve." *Journal of Political Economy,* 103(5), 1091-1120.

Ishitani, T. T. and S. L. Desjardins. 2002. "A Longitudinal Investigation of Dropout from College in the United States," *Journal of College Student Retention,* 4, 173-201

Johnson, George and Frank P. Stafford. 1973. "Social Returns to Quantity and Quality of Schooling," *Journal of Human Resources,* 8(2), 139-55.

Kiser, A. I. T. and L. Price. 2007. "The Persistence of College Students from Their Freshman to Sophomore Year," *Journal of College Student Retention,*9, 421-436.

Lotkowski, V. A., S. B. Robbins, and R. J. North. "The Role of Academic and Non-academic Factors in Improving College Retention. Washington, DC: ACT Policy Report (ERIC Document Reproduction Service No. ED485476).

McGrath, M. and A. Braunstein. 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.

Minor, Jacob. 1970. "The Distribution of Labor Incomes: A Survey with Special Reference to Human Capital Approach," *Journal of Economic Literature,* 8(1),1-26.

Murnane, Frank, , John B. Willett, and Frank Levy. 1995. "The Growing Importance of Cognitive   Skills in Wage Determination," *Review of Economics and Statistics,* 77(2), 251-266.

Murtaugh, P. A., L. D. Burns, and J. Schuster. 1999. "Predicting the Retention of University Students," *Research in Higher Education,* 40, 355-371.

Ribich, Thomas I. and James L. Murphy. 1975. "The Economic Returns to Increased Educational Spending," *Journal of Human Resources, 10(1), 56-77.*

Rosen, Sherwin. 1977. "Human Capital: A Survey of Empirical Research," in *Research in Labor Economics.* Vol. I. Ed.: Ronald G. Ehrenbert. Greenwich, Ct: JAI Press, 3-39.

Saito, Yoshie and Christopher S. McIntosh. 2003. "Monitoring Inefficiency in Public Education,"   *Journal of Agricultural and Applied Economics 35(3),* 611-623.

Watchel, Paul. 1976. "The Effect on Earnings of School and College Investment Expenditures," *Review of Economics and Statistics,* 58(3), 326-31.