J. COLLEGE STUDENT RETENTION, Vol. 5(4) 409-420, 2003-2004

PREDICTING GRADUATION RATES: AN ANALYSIS OF STUDENT AND INSTITUTIONAL FACTORS AT DOCTORAL UNIVERSITIES

CULLEN F. GOENNER
University of North Dakota, Grand Forks

SEAN M. SNAITH
University of the Pacific, Stockton, California

ABSTRACT

The importance that is placed on graduation rates as a measure of the success of institutions of higher education warrant the ongoing research into understanding the determinants of these educational outcomes. This study examines the role of institutional factors in determining graduation rates at doctoral universities. While controlling for student characteristics, we find that institutional characteristics are an important determinant of four-, five-, and six-year graduation rates. Student-faculty ratios, percentage of faculty that are full time, total expenditures and tuition and fees all play a significant role in explaining graduation rates at the universities in our sample.

INTRODUCTION AND OVERVIEW

Graduation rates are often used by the public and government to measure the success of institutions within higher education. Based on this measure alone we see that there is striking variation in terms of success among institutions of higher education. Among doctorate granting universities the six-year graduation rate ranges from 9% at Texas Southern University to 97% at Harvard University. Given these institutions share similar missions and relatively similar institutional characteristics based on their Carnegie Classification, one may perceive Harvard to outperform Texas Southern. Despite their shared classification, Harvard and Texas Southern differ in terms of their institutional organization and resources as

well as the characteristics of their student body. One must account for these differences between institutions to properly compare and evaluate performance.

Alexander Astin's research (1991, 1993, 1997) has considered the effects of student's backgrounds on whether or not they graduate, and has found that individual characteristics are an important influence on an institution's graduation rate. This reflects the notion that institutions differ in terms of the quality of students with whom they work. Thus, one would expect that institutions, such as Harvard, with well prepared students should have higher graduation rates than Texas Southern. Astin's (1997) analysis uses linear regression to estimate the probability of a student graduating, while controlling for sex, high school GPA, ethnicity, and SAT scores of the student. Using these estimated coefficients, institutions are able to predict their expected graduation rate conditional on the characteristics of their entering class. This method provides institutions a benchmark for comparing their actual graduation rate to that predicted by Astin's model in order to draw conclusions on the effectiveness of the institution.

In addition to the background of students, the characteristics of the institution are also relevant to student outcomes. For example, an output of higher education, such as completions, is influenced not only by the quality and quantity of inputs (students) but also through the method (institutions) of production. Institutions differ in terms of their background and commitment of resources to education. Background characteristics include the type, geographic location, and mission of the institution. These factors reflect that institutions may not share similar goals or means of achieving their goals. Public schools may face a mandate of serving as many in-state students as possible which creates constraints in admissions and resources that private schools may not face. Geographic location, rural versus urban settings, provides different environments to students. Tinto (1987) found that urban schools generally provide more remedial courses to students and have weaker infrastructure than less urban settings. Urban institutions find that many students use their institution as preparation to transfer elsewhere, and thus the mission of these students is less concerned with completion.

An institution's commitment of resources to academics is also important to creating a positive environment for student achievement. One measure of this commitment is accreditation. Accreditation typically considers the quality of the academic environment by assessing the quality of the institution's facilities, faculty, and curriculum. Also contributing to the academic environment are individual programs designed to integrate students into the intellectual community of the institution. Orientation programs that create linkages between students and other students, faculty and staff, as well as faculty mentoring are just two examples that Tinto (1987) provides as generating rewarding interactions between students and the institution. Increasing these positive interactions increases the likelihood of positive student outcomes.

The literature that empirically analyzes graduation and retention rates typically uses longitudinal studies (Astin, 1997; Dey & Astin, 1993; Kroc, Howard, &

Hull, 1995; Murtaugh, Burns, & Schuster, 1999; Smith, Edminster, & Sullivan, 2001), in which outcomes and controls are measured at the individual student level. Student outcomes are examined within and across institutions. In the former case, institutional characteristics are largely ignored as controls, though Murtaugh et al. (1999) do control for participation in a freshman orientation program in their analysis of student retention. In several studies that examine students across institutions, differences between institutions are ignored (Astin, 1997; Dey & Astin, 1993; Smith et al., 2001). Kroc et al. (1995) consider in their analysis of student outcomes across 53 research and land grant institutions the effects of both student and institutional characteristics. Their findings suggest that the inclusion of institutional controls for cost, size, quality, and budgetary expenses improves their prediction of graduation rates.

In the analysis below we estimate four-year, five-year, and six-year graduation rates for doctoral universities, when controlling for individual and institutional characteristics. Our method of analysis is similar to that of Porter (2000) in that we examine outcomes at the level of the institution. This allows us to model the heterogeneity of outcomes for universities of similar mission as they relate to the characteristics of students and the institution. The effects of institutional characteristics, while theoretically relevant to predicting graduation rates, have largely been ignored in past studies. The purpose of this analysis is to determine the relative importance of institutional characteristics on producing positive student outcomes and to allow better comparison of an institution's performance versus predicted values. Our findings suggest that it is important to control for institutional characteristics when predicting graduation rates, particularly when longer time periods are used to measure graduation rates.

DATA AND REGRESSION ANALYSIS

The graduation rate of a university is the result of a complex production process incorporating inputs from both the students and the institution. The theoretical importance of individual/cohort characteristics (age, race, grades, standardized test scores, etc.) and quality of the academic environment as determinants of both graduation and retention rates are discussed in multiple works (Astin, 1991; Pascarella & Terenzini, 1991; Tinto, 1987). Two students with identical individual characteristics who attend two different institutions will likely each have a differing probability of completing their degree. In order to avoid the comparison of graduation rates for institutions which have vastly different overall missions, we focus our analysis on data for a sample of 258 Carnegie I research universities. Even within this more homogenous grouping of universities, there remain significant institutional and student specific differences.

We study the determinants of aggregate graduation rates at the four-, five-, and six-year time frames using multivariate regression analysis. In our selection of independent variables we considered several measures of student preparation and

motivation as well as measures of the academic environment. Potential student descriptors, or in the words of Astin (1991, 1997), "input variables," we considered in our modeling process included their SAT scores, percentage of students who graduated in the top 10% of their school class, average age of the student body, and percentage of students from out-of-state. Institutional characteristics considered included class size distributions, percentage of full-time faculty, percentage of faculty holding a Ph.D., student-faculty ratio, institutional affiliation, degree of urbanization and total educational and general expenditures. From this list of potential variables, we chose a subset based on general to specific methodology of model selection, since theory is not unambiguous on what variables should be included. Kennedy (1992) describes this approach, "To begin, the initial specification is made more general than the researcher expects the specification ultimately chosen to be, and testing is undertaken to simplify this general specification . . ." (p. 75). Our testing process examined the significance of individual variables, the impact the removal of variables had on the adjusted coefficient of determination, as well as *F*-tests for omitted variables. Ultimately we arrive at a subset of variables that is common to all three models. This set of variables is listed, along with their summary statistics, in Table 1.

Data used in this analysis were gathered from several sources. These include; US News and World Report's online version of America's Best Colleges 2002, Integrated Postsecondary Education Data System (IPEDS) Fall 2000 Data File and IPEDS Finance Data File for Fiscal Year 1995-96.

Looking at the first three rows of Table 1, it is evident that there exists a wide range of graduation rates for the universities in our sample. The four-, five-, and six-year graduation rates have spreads between the highest and lowest graduation rates of 90%, 94%, and 88%, respectively. The wide variation of educational outcomes for the universities in our sample group can largely be explained by the variation in the student and institutional variables that were subsequently employed in the regression analysis. The remaining rows of Table 1 characterize those variables. Not surprisingly, these variables have a wide variation across our sample group both at the institutional and the student level. Our price variable, the weighted average of tuition and fees for in-state and out-of-state students, has a range of over $24,000 in our sample. The SAT score that represents the 25th percentile of the incoming class for each institution has a maximum value in our sample of 1,450 and a minimum of 680. The other variables detailed in Table 1 display similar heterogeneity across universities. The question that still remains is how well do these variables predict graduation rates? It is on this question that we now focus.

Table 2 displays the results of our multiple regression analysis. The same model was specified and estimated for each of the three graduation rates. The first row lists the graduation rate that is the dependent variable in the regression that included the explanatory variables that are listed in the first column of the Table 2.

Table 1. Descriptive Statistics

Variable name	Number of observations (*n*)	Mean	Standard deviation	Maximum	Minimum	Median
4-Year graduation rate	225	35.71	22.67	90.00	0.00	31.00
5-Year graduation rate	226	53.73	20.04	96.00	2.00	52.00
6-Year graduation rate	248	57.16	18.58	97.00	9.00	55.5
Percentage of students in the top 10% of high school class	220	38.05	25.66	99.00	6.00	29.50
25th Percentile of student SAT scores	243	1029.85	140.14	1450.00	680.00	1010.00
Percentage of out-of-state students	235	27.58	25.21	99.00	1.00	19.00
Average age	220	21.24	2.08	39.00	19.00	21.00
Percentage of full-time faculty	236	87.64	9.69	100.00	28.00	90.00
Educational and general expenditures (millions of dollars)	258	310.03	279.60	1334.91	3.28	200.81
Student-faculty ratio	247	15.06	4.04	25.00	3.00	15.00
Weighted tuition/fees (dollars)	235	10287.58	7996.11	26746.00	1977.48	5646.90

Each cell contains the estimated coefficient, the top number, the standard error of the coefficient in parentheses, and the *p*-value below.

The variables that reflect the student body's characteristics and abilities are all significant in explaining the variation of graduation rates for the universities in our sample. These results at the aggregate level are by and large consistent with the findings of other researchers who have done analysis at the individual level. See, for examples, Astin (1997), Dey and Astin (1993), Kroc et al. (1995), Murtaugh et al. (1999), and Smith et al. (2001).

The first two student level variables are measures of students' backgrounds and preparation for post-secondary study. The results of standardized tests and secondary achievement have been found by other authors as important for the prediction of retention and graduation. The variables we employ to

Table 2. Multiple Regression Results

Variable name	4-Year graduation rate[a]	5-Year graduation rate[a]	6-Year graduation rate[a]
Constant	–12.1513	1.4733	15.3264
	(24.2376)	(21.2836)	(18.4426)
	0.617	0.945	0.407
Percentage of students in the top 10% of high school class	0.2196	0.1478	0.1600
	(0.0611)	(0.0535)	(0.0470)
	0.000	0.006	0.001
25th Percentile of student SAT scores	0.0405	0.0566	0.0507
	(0.0145)	(0.0127)	(0.01105)
	0.006	0.000	0.000
Percentage of out-of-state students	0.1635	0.0594	0.0868
	(0.0555)	(0.0487)	(0.0423)
	0.004	0.223	0.041
Average age	–1.2100	–2.0719	–2.0890
	(0.5927)	(0.5212)	(0.4338)
	0.043	0.000	0.000
Percentage of full-time faculty	0.0257	0.1423	0.1429
	(0.1001)	(0.0880)	(0.0756)
	0.798	0.108	0.060
Educational and general expenditures (millions of dollars)	0.0005	0.0035	0.0054
	(0.0033)	(90.0029)	(0.0025)
	0.868	0.225	0.032
Student-faculty ratio	0.4397	0.6774	0.5666
	(0.3091)	(0.2709)	(0.2304)
	0.157	0.013	0.015
Weighted tuition/fees	0.0010	0.0007	0.0003
	(0.0002)	(0.0002)	(0.0002)
	0.000	0.001	0.039
N	182	183	195
Adjusted R^2	0.79	0.77	0.78
Standard error of the estimate	10.4405	9.1893	8.1628
Regression *F*-statistic	83.8170*	77.4368*	89.5295*

[a]Standard Errors in parentheses, *p*-values listed below standard errors.
*Statistical significance at the 1% level.

capture these individual aspects are the percentage of students who graduated in the top 10% of their high school class and the 25th percentile of the SAT scores for all students. The former measure reflects the previous academic success of the students and the latter measure captures the level of aptitude of the lower quartile of students— those who may be the most at risk of not completing their degrees.

We find that the high school class ranking measure is positively and significantly related to graduation rates for all three horizons. This variable has the largest effect in the four-year graduation rate model, where, on average with other factors equal, an increase of 4.55% in the percentage of students in the top 10% of their graduating class will cause a 1% increase in the four-year graduation rate. The effect of this measure is slightly less on the five- and six-year graduation rates. As such, the finding that this variable is positively and significantly related to graduation rates at all horizons is not surprising. It is also consistent with the findings of related literature.

The results for the SAT scores indicate a positive and significant relationship with graduation rates across all horizons. The magnitude of this relationship is that a 100 point increase in the SAT scores of the first quartile, other things equal, would be expected to increase by 4.4%, 5.7%, and 5.7%, the four-, five-, and six-year graduation rates. Standardized test scores appear to have a strong impact on the graduation rate of these universities. This result is also consistent with the literature that emphasizes the importance of student inputs into the educational process.

The next student level characteristic we considered in our analysis is the percentage of students from out of state. In the aggregate this may affect graduation rates, as it could be a reflection of the motivation of the student about their studies. It may be a latent indicator of the student's motivation for an individual student to be willing to accept the additional costs, both social and financial, of moving out of state to attend a university. Finally, it may also be a signal of the quality of the institution in that students are willing to come from out of state to attend. The results for the percentage of students from out of state are not consistently significant. In all models the coefficient on this variable is positive and it is significant for the four- and six-year graduation rate but not the five-year rate.

The final student level variable included in our models is the average age of the students. The results for this variable are unambiguous. In all models this variable is significant and negatively related to graduation rates. Stated more directly, as the average age of the students increases, the graduation rates of the institution fall at all horizons. For two universities with otherwise identical student and institutional characteristics a one-year increase in the average age of the student body would result in a decrease of the five- and six-year graduation rates by slightly more than 2%. The effect on the four-year rate is slightly lower. The explanation behind this result is likely multidimensional and reflect both

social and academic causal factors. An older student is further removed from the material learned in secondary school and may also have additional burdens (family, work, etc.) that traditional students do not have to bear. This result justifies the need and existence of support programs at universities that target non-traditional learners.

The literature conducting empirical analysis of graduation rates pays less attention to the effects of institutional differences than it does to student characteristics. In addition to controlling for the characteristics of the student body, we also include characteristics of the institution that contribute to the quality of the academic environment. The percentage of full-time faculty and student-faculty ratio are included to measure the quality of the faculty and student interaction with the faculty. Also included are variables that may reflect the availability of resources, specifically the total level of educational and general expenditures and the weighted average of tuition and fees. Our results suggest that controlling for institutional factors is relevant and even more so at longer horizons. All four of the institutional variables are significant for the six-year graduation rate, but this is not true for the four- and five-year rates. This result corroborates Astin's (1997) suggestion that "the reason why some students take more than four-years [to graduate] may have as much to do with the *institution* as with the student" (p. 652).

The percentage of the faculty that is full-time is included in the models to capture both a quantity and quality measure of one type of institutional input into the production process. Full-time faculty may be more available to their students and have more focus on the job of education and assisting students. Part-time faculty are likely to have other competing demands on their time, have less incentive to get involved with students, and may be less accessible to students. There is likely to be a difference in the quality of instruction of part-time faculty as a result. Without the incentive to work to improve their teaching, such as merit-based raises, the goal of achieving tenure, and taking into account the difference in pay received by part time instructors it is not unthinkable that the quality of instruction may suffer. Therefore, we anticipate that the percentage of faculty that is full-time should be positively related to graduation rates. Our anticipated outcome is true and statistically significant for the model of six-year graduation rates, but not significant for the four- and five-year outcomes measures. In all cases, the estimated coefficients have a positive sign.

We see this pattern again when examining the results for the next institutional measure, the level of total educational and general expenditures. This measure is included to capture in a very general fashion the dollar resources being expended by the university. Other things equal, an increase in a university's expenditures should lead to an increase in the university's graduation rate. These expenditures may themselves directly affect educational outcomes, such as expenditures for support staff, for programs for helping at-risk students, for any form of academic support. The salaries paid to professors would have

an effect on the quality of the faculty at a particular institution. Expenditures may affect outcomes indirectly for better classrooms, technology, library resources, etc. While our measure of expenditures is too broad and our current approach too general to draw any conclusions regarding the channels through which these expenditures might affect graduation rates, we do make some general conclusions from our analysis.

The total expenditures variable demonstrates the same pattern of significance that was observed for the percentage of faculty that is full-time variable. At the six-year horizon, the expenditures variable is statistically significant and positively related to the graduation rate. At other horizons the sign of this coefficient is positive, but not statistically significant. This once again affirms Astin's assertion of the increasing importance of institutional characteristics as the graduation horizon is extended.

The next institutional factor included in our models is the student-faculty ratio. Astin (1993) comments, "The student-faculty ratio is one of the most discussed policy issues in higher education" (p. 328). He finds that this ratio is important in determining student perceptions and satisfaction of the institution. He also notes a weak negative effect on attainment of the Bachelor's degree. We find the opposite result in our five-year and six-year models. A higher student-faculty ratio is positively related to graduation rates. This variable is insignificant for the four-year model. The magnitude of these coefficients is striking as well. An increase in this ratio by an amount equal to the sample standard deviation of this variable (approximately four), correlates to an increase of 2.3% in the six-year graduation rate. The causal relationship behind this result, if any, is not entirely apparent. Models that also included the enrollment of the institution did not qualitatively change this result and therefore it is not justified to explain this by appealing to economies of scale. The most plausible explanation in our view is that this variable is positively correlated with some other institutional variable that has not been accounted for in our models. For example, an institution with a high student-faculty ratio may be more likely to have in place other academic support systems such as advisement, tutoring, and honors programs that more than offsets any negative effects of a high student-faculty ratio. Finally, it may be the case that this variable is negatively related to the quality of the education received but not to the actual attainment of the degree.

The final institutional variable employed in our models is the weighted average tuition and fees, in dollars, for each of the institutions. The weighting scheme is applied due to the difference in tuition paid by instate students versus out-of-state students at many universities and is calculated using weights that reflect the percentage of students from out of state attending the institution. This price variable is an important determinant of graduation rates at all horizons. There are several potential reasons why this result is obtained and each will be addressed in turn.

The most direct channel through which the causal relationship may operate is simply the direct cost of not graduating in a given time frame is that the student must then pay tuition to continue on at the university. The higher the tuition rate, the greater the incentive to meet all requirements for graduation and avoid having to pay this penalty. This direct channel also explains why the effect is greater in the four-year model, which represents the shortest time in which most students can finish their degree. Examination of the raw data finds that of all the schools that have six-year graduation rates of 40% or below, none have a weighted annual tuition greater than $6,000. Another plausible explanation appeals to the notion of price as a signal of the quality of the education provided by a given university. Those schools with higher tuition may be providing a superior product and graduates from these institutions would be expected to have higher earnings potential once they have attained their degree and therefore a greater opportunity cost of not graduating. Both the direct price effect of tuition and the indirect effect of price as a signal of quality effect support our findings.

The notion that higher tuition results in a budgetary constraint for the student and could result in an unfavorable outcome is more relevant for looking at an individual institution over time to see the effects on retention of tuition hikes. The importance of tuition levels on the affordability of attending an institution is reflected by the student's decision to matriculate at a particular institution and less important in determining graduation rates is discussed in Tinto (1987).

CONCLUSIONS

Student characteristics are irrefutably an important determinant of graduation rates at Carnegie I universities. At the aggregate level, high school class rankings, standardized test results, and percentage of out-of-state students are all positively related to graduation rates. The first two are significant for all three horizons examined in this article while the third is significant at the four- and six-year horizons. Average age of the student body is also an important factor and is negatively related to graduation rates at all horizons.

Student factors are not the only inputs into this complex production process. A comprehensive examination of these outcomes requires examination of the institutional input into this process as well. We find that in addition to the student characteristics, institutional factors are important to fully understand these educational outcomes. Student-faculty ratios, the percentage of faculty that are full-time, total expenditures, and tuition and fees all have an impact on graduation rates. These factors become more relevant as the graduation horizon is extended from four to six years.

Given the growing importance placed on educational assessment and outcomes by legislators and administrators alike, it is clear that full consideration of both student and institutional factors is necessitated. Our analysis suggests that an institution seeking to improve its graduation rates has several options at its disposal. Clearly changing admissions policy so that incoming classes are comprised of students who are better prepared and have a higher potential for success in completing their degree programs is one route an institution might take. Raising the percentage of students who graduated in the top of their high school class and raising the lowest quartile of SAT scores will unquestionably raise graduation rates. However, institutions may also raise their graduation rate by improving the quality of their faculty and increasing resources devoted to education. Raising the percentage of full-time faculty, tuition and total expenditures each improves the graduation rate.

Further research is needed to clarify these interrelationships. Potential avenues for this research include disaggregated analysis of expenditures made by universities. Particular attention should be paid to the types of expenditures to determine which are most likely to impact the graduation rate. It may be the case that expenditures on certain types of programs and activities, such as academic advisement and support, have a differential effect on graduation rates than do expenditures on technology and infrastructure. Knowledge of any differential effects will allow institutions to better target spending on outcome-improving programs.

REFERENCES

Astin, A.W. (1991). *Assessment for excellence.* New York: Macmillan.

Astin, A. W. (1993). *What matters in college?* San Francisco: Jossey-Bass.

Astin, A. W. (1997). How "good" is your institution's retention Rate? *Research in Higher Education, 38*(6), 647-658.

Dey, E. L., & Astin, A. W. (1993). Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression. *Research in Higher Education, 34*(5), 569-581.

Kennedy, P. (1992). *A guide to econometrics* (3d ed.). Cambridge, MA: The MIT Press.

Kroc, R. D., Howard, R. H., & Hull, P. (1995). *Predicting graduation rates: A study of land grant, research I, and AAU universities.* Paper presented at the Association for Institutional Research meeting, Boston, 1995.

Murtaugh, P. A., Burns, L. D., & Schuster, J. (1999). Predicting the retention of university students. *Research in Higher Education, 40*(3), 355-371.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students.* San Francisco: Jossey-Bass.

Porter, S. R. (2000). The robustness of the graduation rate performance indicator used in the U.S. News & World Report college rankings. *International Journal of Educational Advancement, 1*(2), 145-163.

Smith, W. R., Edminster, J. H., & Sullivan, K. M. (2001). Factors influencing graduation rates at Mississippi's public universities. *College and University, 76*(3), 11-16.

Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition.* Chicago: University of Chicago Press.

Direct reprint requests to:

Sean M. Snaith
Director, Business Forecasting Center
Associate Professor of Business Economics
Eberhardt School of Business
University of the Pacific
3601 Pacific Ave.
Stockton, CA 95211
e-mail: ssnaith@pacific.edu