Predicted Retention and Graduation Rates as Performance Indicators: The Neglected Role of Uncertainty

**Predicted Retention and Graduation Rates as Performance Indicators:
The Neglected Role of Uncertainty**

ABSTRACT

The use of predicted retention and graduation rates from statistical models to evaluate universities and colleges has grown increasingly popular. More recently, several states have begun to link budgets to performance on these indicators. Proponents of predicted rates fail to recognize that these predictions contain substantial error, and that this error, or uncertainty about the estimates, must be taken into account when evaluating performance. Using data from a small state higher education system to estimate a system-wide one-year retention model, the analysis reveals that several so-called under-performing institutions are performing exactly as expected.

Predicted Retention and Graduation Rates as Performance Indicators: The Neglected Role of Uncertainty

Introduction

As state legislators demand increased accountability from public institutions of higher education, administrators and researchers have struggled to develop indicators that accurately reflect the state of their campuses as well as conform to legislators' expectations of simple and easy-to-understand measures. During this process retention and graduation rates have become popular indicators because they are viewed as direct measures of student outcomes. Yet many scholars believe raw retention and graduation rates are misleading at best, because they do not reflect differences in student characteristics between institutions (Astin, 1990, 1997; Kroc et al., 1995, 1997; Mortenson, 1997).

To derive meaningful retention and graduation rates these scholars have instead advocated using *predicted* rates from statistical equations controlling for student characteristics, rather than the simple *actual* rates. If an institution's actual rate is higher than its predicted rate, which takes into account characteristics of their student body presumably out of its control, then the institution must be doing something well to account for such over-performance (Astin, 1990). Using either institutional-level or individual student-level data, these indicators have become increasingly prominent. *U.S. News and World Report* (e.g. Smith, 1998) has used such predicted rates as one element in their controversial college rankings for several years. The state of Virginia is planning to use predicted rates as an element in performance funding for its university system (State Council of Higher Education of Virginia 1999), and the states of New York and North

Carolina are also considering similar uses in their university system budgets (Blose, 2000; Mayes et al., 2000; State University of New York, 1998).

From the perspective of state institutions, the recent shift in the use of these predicted rates from national college rankings to state performance funding is very important. Previously a poorly performing institution only suffered a slight drop in its ranking, because the predicted graduation rate is only one element among many in *U.S.News*'s ranking system. Now, under these performance-funding approaches, institutional budgets will be directly linked to the predicted values from these statistical models. With performance funding, if an institution does not perform, it does not receive additional funds.

Such linking is very problematic due to the nature of predicted retention and graduation rates. Proponents of this methodology implicitly assume that we can know with certainty the predicted rate for an institution. Yet our models of student behavior are always a gross simplification of a very complex reality, and thus will never perfectly reflect this reality. In statistical parlance, our models of student behavior are error-based, and thus the predicted rates from these models contain error. Only by taking into account this error or uncertainty can we say anything meaningful about an institution's predicted retention or graduation rate.

While policymakers have long understood that uncertainty about future outcomes must be taken into account when developing policy (Morrison and Mecca, 1989), proponents of these predicted rates have neglected this aspect of policy analysis. Such neglect can be perilous. As this paper will demonstrate, under these proposed performance funding measures some institutions will be unfairly penalized as so called under-performers, when in reality they are performing exactly as expected once uncertainty about predictions is taken into account. It is vital that administrators and policymakers understand the implications of using such measures in budgets.

Predicted rates: the bad, the good and the uncertain

Raw retention and graduation rates have long been used as measures of institutional performance, even to the extent that these data must now be submitted to the federal government. Many scholars and higher education administrators believe these numbers to be misleading. Retention and graduation rates are determined by the quality of entering students as well as what an institution does (Astin, 1990; Mortenson, 1997). An institution may have a low retention rate not because they have a low "retention capacity," but because of their student body.

These critics argue that a better measure of institutional performance would take into account these student characteristics. Predicted rates from a statistical equation controlling for student characteristics are increasingly used as performance measures. Although the independent variables and methodology may vary, these models all attempt to control for factors outside an institution's control. Any remaining differences found between institutions in predicted retention and graduation rates must therefore be due to differences in institutional policies and planning.

Such an approach appears quite attractive, but there are several problems with this approach. Porter (2000) identifies four problem areas with these models using the *U.S. News* methodology as an example. While the *U.S. News* methodology uses institutional data rather than student data, the same problems arise with individual level data:

- Changes in the sample can affect an institution’s predicted rate.
- Seemingly irrelevant changes in variable definitions can affect an institution’s predicted rate.
- Adding additional explanatory variables can radically change how well an institution performs, to the point of changing an institution from an over-performer to an under-performer.
- After taking into account the statistical uncertainty about the predicted rates, fully 95% of the institutions in the sample end up performing exactly as expected.

This paper focuses on the fourth problem area: taking into account the underlying uncertainty in these statistical models.

Usually error in predictions or forecasts is taken into account through a confidence interval for the predicted value. If the actual rate for an institution lies within this interval, we cannot say that there is a statistically significant difference between an institution's actual rate and its predicted rate. This confidence interval is a measurement of our uncertainty about the prediction due to sampling error and random behavior. Taking into account uncertainty can radically change the substantive conclusions we draw from comparisons of predicted to actual rates. Using confidence intervals for the institutions in *U.S.News*'s set of national universities, only 9 out 179 institutions had predicted graduation rates that were significantly above or below their actual rates (Porter, 2000). The important policy conclusion is that under this methodology many institutions may be penalized (or rewarded) when in reality we cannot draw any conclusions about their behavior other than they appear to be performing exactly as expected.

Unlike the *U.S. News* approach that uses institutional-level data, the performance funding approach uses individual-level student data to estimate their statistical models. They then take the predicted probabilities of retention (or graduation) from their model and average them to get a predicted retention rate controlling for student characteristics. Such predicted rates can be a legitimate tool, but it is important to realize that there is substantial uncertainty about these predicted values due to fact they come from error-based models. Such uncertainty comes about in two ways (King et al., 2000). First, what King terms *estimation uncertainty* occurs because our parameter estimates are not perfectly estimated, either because of our use of a sample, or because we have not included all relevant explanatory variables. Second, random events may influence whether or not a student is retained, but these events are not included as explanatory variables in our models. Such *fundamental uncertainty* will result in predictions that err due to the impact of these random events. Only by taking into account these forms of uncertainty can we begin to have informed discussions about an institution's predicted retention rates, or where it "should" be, and its actual rate.

Taking into account uncertainty

How can we take into account this uncertainty when dealing with the predicted values from our equations? One difficulty in using confidence intervals (also called prediction intervals) with student-level data arises from the conclusions we wish to draw. The confidence intervals calculated on student-level data are derived for individual students, yet in the end we require some measure of uncertainty for each *institution*, not each student.

The solution is to view the predicted values and confidence intervals for each individual student in a slightly different way than usual. Suppose a student, Jane, has a predicted probability of being retained of .85, and the confidence interval brackets this value from .80 to .90. The classical interpretation is that if we repeated this process on many different samples, using a 95% error level, we would expected the predicted value to fall between .80 and .90 approximately 95 times out of 100. Note that under this formulation *there is nothing special about the predicted value of .85*: it is simply the center point of the prediction interval. Any value between .80 and .90 is equally likely to occur as we estimate our model on repeated samples.

The values of .80 and .90 could be considered the best and worst case scenarios for Jane. We can be fairly certain that her probability of retention lies between .80 and .90, but that is about all we can say, due to possible exclusion of variables from our model that should have been included, or random events in Jane's life that my affect her retention outcome. We can then combine the best and worst case scenarios for each individual student to come up with an estimated best and worst case outcome for the institution. If every student's worst-case outcome occurred, that is, his or her probability of retention was the lowest part of their prediction interval, by averaging these we can come up with the worst possible predicted outcome for the institution, that is, the lowest possible predicted retention rate. We could also do the same with the best-case scenarios as well. By averaging the end points of the confidence intervals for each student, we can develop a prediction interval for the institution. Rather than rely on the predicted value for the institution, which is nothing more than the average of the individual predictions, we can instead develop an interval that takes into account our uncertainty about the individual student predictions.

An example using a state university system

The effect of uncertainty on how we judge institutions is best illustrated with data from a state higher education system. The data used here are taken from official datasets submitted to a small state university system. Data for all first-time, degree-seeking first-year students who enrolled in Fall 1996, including their SAT scores, age, gender and one-year retention outcomes are used to estimate a dichotomous logistic regression model, and the uncertainty about the predicted rates from these equations is quantified.

Table 1 presents summary information for the seven institutions in the state system. As can be seen, the schools in this system vary widely from a small, historically black institution (school A) to two second-tier research universities (schools F and G).

The results of the logistic regression model are presented in Table 2. The direction of the individual coefficients make sense, with high SAT students and female students having higher probabilities of being retained, while older students have on average a smaller probability of being retained. The impact of a unit change in an independent variable on the probability of being retained is given in the far right column in the upper part of the table. Overall the equation predicts greater than average, being correct for 58% of the cases.

The bottom half of the table lists the actual and predicted retention rates for each school. The predicted retention rates are calculated by taking the individual predicted probabilities for the students in each school and averaging them. The difference between the two, or how well each school is performing, is listed in the column on the far right. As can be seen, school G is an over-performer, with its actual retention rate over two percentage points higher than its predicted rate, or where its retention rate should be after taking into account characteristics of its students. Several other schools are serious under-performers, with differences up to almost five percentage points.

Prediction intervals were also calculated for each school by taking the average of the endpoints of the confidence intervals for each individual predicted probability of retention. These intervals, along with the predicted rates and actual rates, are displayed in Figure 1.

An X marks the actual rates, while a solid circle marks the predicted rate for each institution. The bars represent the prediction interval for each institution. As can be seen, for three out of the seven schools (F,E and D) the prediction interval brackets the actual retention rate. For these schools, we cannot say that their predicted retention rate significantly differs from their actual rate.

For the other institutions, the difference between one endpoint of the prediction interval and the actual rate is much smaller than the difference between the predicted rate and the actual rate. For school C, proponents of the approach used by some performance funding plans would state that the school is under performing by almost four percentage points, because the actual rate is 3.7 percentage points lower than the predicted rate. However, we can only say that the predicted rate will fall somewhere in the bracket in the figure. Thus after taking into account our uncertainty about the interval, the largest difference we can assume is 1.5 percentage points, the difference between the lower part of the prediction interval and the actual value. The difference may be bigger, but since we can not be sure where our predicted retention rate really is (except that 95 times out of a 100 it will lie within the bracket), we can only conclude that school C is under-performing by 1.5 percentage points.

School performance with and without uncertainty is summarized in Table 3. The section in the left portion of the table shows how the schools in this system would be rated under the current approaches in use. The section in the right-hand side of the table indicates that the performance differences are much smaller once we take into account our uncertainty about our estimates. Indeed, the differences are now so small as to call into question the whole enterprise of developing predicted rates for these schools.

Conclusion

Most scholars and higher education administrators would agree that we need to take into account uncertainty when trying to predict how well an institution should be performing. The results presented here indicate that the differences between actual and predicted retention rates become much smaller once we take into account uncertainty about estimates.

Given these smaller differences, one possible approach would rely not on the percentage point difference between actual and predicted rates, but rather on a simple three-outcome choice: over-performance, expected performance or under-performance. This formulation seems more sensible, in that it should be more robust to small changes in methodology than using the percentage point differences.

Also, scholars need to think carefully about equity issues when promoting these approaches to measuring performance. Does it really make sense to include small, historically black colleges with research universities when estimating these models? Or would it make more sense to rate these institutions with comparable institutions across the country? In addition, the relative size of the schools in the system may be driving some of the results. Referring back to Table 1, we can see that school G makes up almost half of the sample, and coincidentally is one of the two school rated as an over performer. Shouldn't the students in smaller institutions be weighted more, so that each school has equal weight in the analysis? Only by answering these questions can we begin to move forward with this research agenda.

References

Astin, A.W. 1990 Retention-Rate Data Mislead Student Consumers. *Chronicle of Higher Education*, Nov. 21, 1990, p. B2.

Astin, A. W. 1997. How ‘Good’ is Your Institution’s Retention Rate? *Research in Higher Education*, 38, 647-658.

Blose, G. Modeled Retention and Graduation Rates: Calculating Expected Retention and Graduation Rates for Multi-Campus Universities. In *The New Three R's: Recruitment, Remediation and Retention*, Gerald Gaither, ed. Jossey-Bass.

King, G., M. Tomz and J Wittenberg. 2000. Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science*, 44 (2), 341-355.

Kroc, R., D. Woodard, R. Howard and P. Hull. 1995. Predicting Graduation Rates: a Study of Land Grant, Research I and AAU Universities. Paper presented at the Association for Institutional Research meeting, Boston,.

Kroc, R., D. Woodard, R. Howard, P. Hull and D. Woodward. Graduation Rates: Do Students’ Academic Program Choices Make a Difference? 1997. Paper presented at the Association for Institutional Research meeting, Orlando.

Mayes, L., L. Williford and R. Ussery. 2000. Policy Implications of Using Prediction Models of Student Retention and Graduation Rates. Paper presented at the Association for Institutional Research meeting, Cincinnati.

Morrison, J., and T. Mecca. Managing Uncertainty: Environmental Analysis / Forecasting in Academic Planning, pp. 334-382. In J. Smart, ed., *Higher Education: Handbook of Theory and Research V*, New York: Agathon.

Mortenson, T. 1997. Actual versus Predicted Institutional Graduation Rates for 1,100 Colleges and Universities, *Postsecondary Education Opportunity*, April, 58.

Porter, S. 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), 10-30.

Smith, S. 1998. *America's Best Colleges, 1999*. Washington, D.C.: U.S. News and World Report.

State Council of Higher Education of Virginia. 1999. *Virginia Higher Education Performance Funding Model*, June 14, 1999.

State University of New York. 1998. *Task Force Report on Performance Indicators and Merit-Based Performance Funding*, July 6, 1998.

Table 1. Institutional Characteristics of System Schools

Institution	Freshman class size	One-year retention rate	Mean SAT score	Mean age	Female (%)	Nonwhite[a] (%)
A	382	72.8%	886	19.6	70.2%	98.2%
B	357	73.4%	922	18.6	53.2%	95.0%
C	865	76.9%	994	18.3	51.3%	16.8%
D	886	84.8%	1110	18.4	58.1%	7.7%
E	1,136	84.2%	1177	18.1	53.3%	42.1%
F	1,808	83.4%	1094	18.3	59.7%	18.1%
G	3,960	88.2%	1199	18.2	47.4%	38.1%
All schools	9,394	84.2%	1126	18.3	53.0%	34.5%

[a]Includes international students.

Table 2. One-Year Retention Logistic Regression Results

	Coefficient	Standard error	Change in p(retained)[a]
Intercept	0.4697	0.3739	-
SAT	0.0020**	0.0002	0.02
Age	-0.0642**	0.0146	-0.10
Female	0.3879**	0.0581	0.05
African-American	0.0078	0.0788	0.00
Asian-American	0.2914*	0.1181	0.03
Hispanic	-0.3068*	0.1567	-0.05
Native-American	-0.1633	0.4611	-0.02
Foreign	-0.4276*	0.1981	-0.07
Other/unknown	0.3036	0.2250	0.04
N	9,394		
-2LL	7994.7		
% correctly predicted:	overall	58.0	
	retained	58.3	
	not retained	56.5	
Retention rates	Actual	Predicted	Performance
A	72.8	76.4	-3.6
B	73.4	78.0	-4.6
C	76.9	80.6	-3.7
D	84.8	84.0	0.8
E	84.2	86.1	-1.9
F	83.4	83.7	-0.3
G	88.2	86.1	2.1
All schools	84.2	84.2	0.0

[a] Given an unit change in independent variable, calculated at full sample means. Unit change is 100 for SAT, 10 for age and 1 for dummy variables.

Table 3. Institutional Performance and the Effect of Uncertainty

	Without uncertainty		With uncertainty	
Institution	Performance	Amount	Performance	Amount
A	Under	-3.6	Under	0.3%
B	Under	-4.6	Under	1.8%
C	Under	-3.7	Under	1.5%
D	Over	0.8	As expected	0.0%
E	Under	-1.9	As expected	0.0%
F	Under	-0.3	As expected	0.0%
G	Over	2.1	Over	0.4%

Figure 1. Actual and Predicted Retention Rates with Confidence Intervals