

Heads or Tails (Success or Failure)? Using Logit Modeling To Predict Student Retention and Progression

Michael C. Budden, Southeastern Louisiana University
Yu Hsing, Southeastern Louisiana University
Connie B. Budden, Southeastern Louisiana University
Michelle Hall, Southeastern Louisiana University

Abstract

Using a sample of 2,137 university students and applying the logit model, we find that the probability for students to return in fall 2008 is higher with a higher cumulative GPA, a higher grade for SE 101, and a returning status in the previous semester. Several other explanatory variables are tested and have insignificant coefficients. A few variables such as the Board of Regent's core requirements (CORE) and high school graduating GPA (HSGPA) have the expected signs and z-statistics closer to one, suggesting that the correlation coefficient may rise if the sample size were larger. The findings suggest that the cumulative GPA is a dominant factor and that the large number of failures in SE 101 may need to be examined in order to fulfill its described purpose: "a course designed to ensure first-year student success."

Keywords: retention, graduation rates, persistence, progression, logit

Introduction

A major goal of many in higher education is to increase student retention. Increased retention would lead to increased progression. Increased progression would lead to increased graduation rates. The ability of U.S. higher education institutions to attract students is widely heralded. However, while increasing enrollment numbers make administrators giddy, the lack of student retention and progress is causing alarm in many quarters, especially those related to public higher education. Jacobs and Archie (2008) cite an average departure rate for first year students of 25% and mention that retention has been of concern for decades. This decades-long concern ostensibly has not produced the result expected by many. Indeed, state legislators are increasingly demanding accountability for their support of higher education and expect universities to increase their output of qualified, educated graduates.

As an example, Louisiana is among several states that have expectations that institutions of higher education will increase their graduation rates – rates that are a function of retention and progression (Marklein, 2009). While such intent has long been publicly desired, it has only been in recent years that Louisiana began a program to reward institutions financially for their successes. As Marklein mentions, the idea of rewarding educational institutions for attaining certain goals has been around for decades, the focus of providing significant rewards for attaining specific student outcomes is a more recent concept.

As a sign of the times, in Louisiana there are no less than two statewide commissions looking to increase accountability of higher education. One, the Postsecondary Education Review Commission created by the legislature is taking the lead in assuring higher education is operating efficiently and in a manner that serves the people of Louisiana. Among other issues the governor expects that Commission to look into is whether or not degree programs are aligned with the economic development needs of the state and to review the less than stellar relation between state expenditures and institutional performance. Indeed, while Louisiana ranked third in percentage increase in state appropriations for higher education during the past ten years, the state's graduation rate

of 37% places it last among Southern Regional Education Board (SREB) states (BayouBuzz.Com, 2009). However, ten years ago the majority of Louisiana institutions had open admissions – something that has changed. As a result, same institution graduation rates have increased from 32.1% of the 1995 cohort to 37.7% of the 2002 cohort (Board of Regents Graduation Rates Report). Louisiana may not be alone in failing to equate increased funding with increased performance. Sander and Krautman (1991) noted among their findings that spending per student did not increase ACT scores.

Louisiana's governor is not alone in expecting higher education institutions to contribute to economic development needs by producing relevant, prepared graduates. Alkandari (2008) believes institutions should focus on preparing students to provide communities with professionals for work needs and to participate in the development of different arenas. Further, Alkandari believes administrators have the responsibility of creating an environment that will motivate students to persist in their studies and progress.

Poor retention rates plague many institutions. And, there is concern by some that culture may play a role in the inability of some to progress. Harbrecht, Neidermeyer and Tuten (2006) discussed the specific needs of Hispanic students and recommend among other things that group work be used to increase retention and satisfaction. As Rivas et al (2007) mention, the theory of college persistence has validated models with links between student background, educational and institutional commitment, and academic and social interaction. Abrego, Morgan and Abrego (2009) mention research that found that improvements in instruction could contribute to increased student persistence and success and found through their own efforts that students (many of whom were Hispanic) attending an annual Teaching, Learning and Service Conference valued learning to work as a team.

Roach (2008) notes 62 institutions of higher education have higher graduation rates for their black undergraduates than for their white undergraduates. At the same time, he provides a statement that gender may play a role in these figures, as black males appear not to persist (progress) at the same pace as black females.

DeWitz, Woolsey and Walsh (2009) investigated the relationship between Frankl's purpose in life construct with Bandura's self-efficacy theory in order to predict students who might be at risk of leaving school. DeWitz et al's findings supported creating interventions based on self-efficacy theory in order to influence sense of purpose in life in order to improve student retention rates. Still, the question remains: what can be done about relatively poor retention and graduation rates?

The Study

The large number of non-returning students is cause for concern. Out of a total of 2,534 beginning students at Southeastern Louisiana University in fall 2007, 912 students or 35.99% did not return in fall 2008. In view of the large number of non-returning students, student retention has become a major focus for university and college administrators partly because it has become a major uncertainty in the university budgeting process and partly because administrators would like to learn how to improve student academic advising and programs in order to reduce the number of non-returning students and increase student retention. There are a variety of reasons why students fail to persist and progress.

Students who return may seek to complete course requirements, receive college degrees, develop or continue social interactions and perhaps recognize the relationship between future career aspirations and success in college. One concern is that many may enter higher education unprepared. Remington et al (2000) noted that marketing students were perceived as lacking in communication skills; skills that Lehman and Dufrene (2002) mentioned do not occur automatically. Noser, Tanner and Budden (1992) found that economics students were perceived as having poor writing skills. And twenty years ago, Cudd, Tanner and Budden (1989) noted that while many approaches have been tried to improve writing skills among business students, the problem remained. Further, Budden and King (2001) discuss the need for schools to develop essential skills in students before they graduate. Still, others may not be sufficiently motivated, have insufficient resources or be incapable of juggling a variety of responsibilities, only one of which may be higher education efforts. As Budden and King (2001) proffer, employers are seeking graduates that are able to learn, adjust and grow in the work environment.

This research examines the determinants of returning or non-returning students at Southeastern Louisiana University. The Ordinary Least Squares (OLS) is not employed in empirical work because it is not an efficient estimator and because the predicted value for the dependent variable may be less than 0 or greater than 1. The logit model is applied since it will restrict the forecast value of the dependent variable to between 0 and 1 and provides valuable estimates such as the percent of correct and incorrect predictions.

The Logit Model

The logit or logistic model can be expressed as:

$$f(W) = \frac{1}{1 + e^{-W}} \quad (1)$$

where W is defined as

$$W = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k \quad (2)$$

f(W) = the probability of a particular outcome with values between 0 and 1,

X's = explanatory variables, and

α 's = regression coefficients to be estimated.

In empirical work, one estimates the following:

$$\text{Logit}(P_i) = \ln\left(\frac{P_i}{1 - P_i}\right) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k + u_i \quad (3)$$

where

P = the probability that the event occurs,

$\frac{P_i}{1 - P_i}$ = the odds ratio, and

u = the error term.

The partial derivative of the probability with respect to any of the explanatory variables is:

$$\frac{\partial P_i}{\partial X_i} = \frac{\alpha_i e^{-W}}{[1 + e^{-W}]^2} \quad (4)$$

Data and Empirical Results

The source of the data came from the Office of Institutional Research and Assessment at Southeastern Louisiana University. The dependent variable refers to returning or non-returning students in the fall semester of 2008. Originally, there were 2,534 students in the cohort. Due to lack of complete data for several variables, the final sample size is reduced to 2,137. Of this total, 1,478 students or 69.16% returned, and 659 students or 30.84% did not return. It was expected that grades for the required beginning college math and college English courses would be a variable in the analysis. Since more than 800 of the cohort did not take college math (Algebra - MATH 155 or 161) or college English (Composition - ENGL 101) in the first year, inclusions of these two variables would have reduced the sample size substantially. As a result, they were not included in the analysis. Variables are as follows:

RETENTION = 1 if a student returns in the fall semester of 2008;

= 0 if a student fails to return in fall 2008;

SPR08 = 1 if a student returns in the spring semester of 2008;

= 0 if a student fails to return in spring 2008;

GPA = cumulative GPA;

- SE101 = grades for SE101 (3- course credit freshman success course designed to improve retention; this course replaced Orientation 101);
- ACTCOMP = composite ACT score;
- ACTENG = ACT score for English;
- ACTREAD = ACT score for reading;
- ACTSCI= ACT score for science;
- AGE = age of student;
- CORE = 1 if a student has taken Board of Regents core requirements;
= 0 otherwise;
- DEVENG = 1 if required to take developmental English (based on ACT score);
= 0 otherwise;
- DEVMAT = 1 if required to take developmental Math (based on ACT score);
= 0 otherwise;
- HSGPA = high school graduation GPA;
- HSRANK = student graduation ranking in high school;
- RACE = 1 if a student is a black,
= 0 otherwise;
- GENDER = 1 if a student is a male,
= 0 otherwise;
- TOPS = 1 if a student receives TOPS (Tuition Opportunity Program for Students)
= 0 otherwise.

TOPS (Tuition Opportunity Program for Students) is a scholarship program funded by the state. TOPS provides tuition to Louisiana’s high school graduates entering college who meet specific entrance requirements and progress according to Board of Regents’ expectations. As Budden and Hsing (2006) emphasized the importance of academic performance in determining a student’s eligibility for TOPS drives the need to assess overall performance.

Table 1 presents the estimated logit regression including potential explanatory variables. The log-likelihood method is used in estimation. The value of the log likelihood function is maximized at -966.89. McFadden R-squared is estimated to be 0.268, suggesting that 26.8% of the behavior of whether students would return or not can be explained by the right-hand-side variables. As shown, the coefficients of SPR08, GPA, and SE101 are significant at the 1% level and have the expected sign whereas the coefficients of other explanatory variables are insignificant at the 10% level.

Specifically, a higher GPA, a higher grade for SE101, and status as a returning student in the Spring semester of 2008 would increase the probability that a student is likely to return to in the Fall semester. Several variables such as CORE and HSGPA have the correct signs and z-statistics closer to one, but they are insignificant at the 10% level. Several variables such as ACTCOMP and HSRANK have the wrong signs probably due to a high degree of multicollinearity. According to the estimated logit regression, the prediction of whether a student would return is 95.20% correct, and the prediction of whether a student would not return is 50.08% correct. The overall prediction is 81.28% correct.

Table 1. Estimated Logit Regression

Dependent Variable: RETENTION
 Method: ML - Binary Logit (Quadratic hill climbing)
 Included observations: 2137
 Convergence achieved after 6 iterations
 Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-6.271965	1.762985	-3.557584	0.0004

SPR08	2.776855	0.269091	10.31938	0.0000
GPA	0.746504	0.090961	8.206862	0.0000
SE101	0.403857	0.066912	6.035643	0.0000
ACTCOMP	-0.028979	0.053397	-0.542703	0.5873
ACTENG	0.013235	0.031290	0.422987	0.6723
ACTREAD	0.020436	0.020081	1.017663	0.3088
ACTSCI	0.027382	0.024388	1.122804	0.2615
AGE	0.018912	0.080231	0.235725	0.8136
CORE	0.268758	0.231148	1.162709	0.2449
DEVENG	0.196639	0.257535	0.763541	0.4451
DEVMAT	0.132033	0.165305	0.798727	0.4244
HSGPA	0.280308	0.297454	0.942358	0.3460
HSRANK	-0.006675	0.005236	-1.274838	0.2024
RACE	0.159126	0.155552	1.022973	0.3063
GENDER	0.081988	0.127618	0.642445	0.5206
TOPS	-0.081235	0.153526	-0.529130	0.5967
Mean dependent var	0.691624	S.D. dependent var		0.461931
S.E. of regression	0.379970	Akaike info criterion		0.920814
Sum squared resid	306.0801	Schwarz criterion		0.965897
Log likelihood	-966.8897	Hannan-Quinn criter.		0.937312
Restr. log likelihood	-1320.229	Avg. log likelihood		-0.452452
LR statistic (16 df)	706.6776	McFadden R-squared		0.267635
Probability(LR stat)	0.000000			
Obs with Dep=0	659	Total obs		2137
Obs with Dep=1	1478			

Table 2 re-estimates the logit model by including only those variables with significant coefficients. As shown, the coefficients of these three explanatory variables have the positive sign and are significant at the 1% level. McFadden R-squared is estimated to be 0.264, indicating that deleting insignificant variables would not affect the explanatory power much.

Table 2. Estimated Logit Regression with Significant Coefficients

Dependent Variable: RETENTION

Method: ML - Binary Logit (Quadratic hill climbing)

Included observations: 2137

Convergence achieved after 5 iterations

Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-4.398655	0.304348	-14.45271	0.0000
SPR08	2.786563	0.267586	10.41370	0.0000

GPA	0.732613	0.083565	8.766930	0.0000
SE101	0.398027	0.065012	6.122379	0.0000
Mean dependent var	0.691624	S.D. dependent var		0.461931
S.E. of regression	0.379608	Akaike info criterion		0.912678
Sum squared resid	307.3699	Schwarz criterion		0.923286
Log likelihood	-971.1963	Hannan-Quinn criter.		0.916560
Restr. log likelihood	-1320.229	Avg. log likelihood		-0.454467
LR statistic (3 df)	698.0645	McFadden R-squared		0.264373
Probability(LR stat)	0.000000			
Obs with Dep=0	659	Total obs		2137
Obs with Dep=1	1478			

According to the estimated logit regression, the forecast whether a student would return is 94.79% correct, the prediction whether a student would not return is 50.53% correct, and total prediction for returning and non-returning students is 81.14% correct. Based on the equation for the partial derivative, if a student's cumulative GPA increases 1.0, the probability that a student would return will increase 0.147. If a student's academic performance in SE101 improves by a letter grade, the probability that this student would return will increase by 0.08. If a student continues in spring 2008, the probability that he/she would return in fall 2008 will increase by 0.599.

Several other variables may explain the behavior of whether or not a student would return. Some students may lack adequate financial resources to support the continuation of college education. A few may quit school due to personal reasons such as marriage or child bearing. Students may not return due to job requirements or transfers. Unfortunately, lack of information on such personal matters precludes further analysis.

Summary and Conclusions

This paper has examined the determinants of returning and non-returning students based on a sample of 2,137 students. The logit model is employed in empirical work in order to yield better estimates. Sixteen explanatory variables have been considered. They include high school performance, ACT scores, age, gender, remedial measures, academic performance at the university, credit hours taken, whether a student would be a TOPS recipient, whether a student has completed Board of Regents' core requirements, etc.

Two versions of the logit regression were estimated. The first version includes all the potential variables available for the analysis. The second version deletes explanatory variables with insignificant coefficients and retains only those explanatory variables with significant coefficients. The results show that a higher cumulative GPA, a better grade for SE 101, or a returning student in the previous semester (spring 2008) would increase the probability that a student is likely to return in the fall semester of 2008.

There are several policy implications. The cumulative GPA seems to be the best indicator for student retention mainly because it is the composite average of all the courses taken. Hence, to improve student retention, special attention and more academic advising may be provided for those students with relatively low GPAs. If a student returned in spring 2008, there is a probability of 0.599 that the student will return in fall 2008. Thus, whether a student would return in spring 2008 may be used as a predictor for whether that student would return in fall 2008.

Further analysis is needed to examine the incidence of students failing SE 101. For instance, in the whole sample of 2,534 students, 395 students failed in SE 101, and 88 students withdrew from SE 101. Thus, better than 19% of students fail to complete or pass the course in their first semester – a stated requirement of the curriculum.

The simple correlation coefficient between whether a student would return and the grade for SE 101 is 0.447. Southeastern's catalogue describes SE 101 as a course that "promotes student success." It would be desirable to investigate why 1 of 5 students fail to complete the course. Data on students' non-academic reasons for not returning would be a plus in future research. Pooled data for multiple years will increase the sample size and may improve statistical outcomes. In addition, changes to the SE 101 course have been made and initial data indicates these changes have led to more success in the course. It will be of interest to see if these changes improve both course success and retention.

References

1. Abrego, M., Morgan, B. M. and Abrego, C. (2009, 2nd Quarter). Creating win-win partnerships and adding relevance to educator preparation. *Contemporary Issues in Education Research*, Vol. 2, No. 2, pp. 51-57.
2. Alkandari, N. (2008, June). Factors affecting students' retention and Kuwait University. *College Student Journal*, Vol. 42, No. 2, pp. 483-492.
3. *BayouBuzz. Com* (2009). Governor Jindal Addresses Higher Education Reform. Retrieved 8/26/09, www.bayoubuzz.com/news/Louisiana/Politics/Gov._Jindal_addresses_Louisiana
4. Budden, M. C. and King, J. O. (2001, October). Ensuring the success of graduates: Completing the picture. *Proceedings of the International Business and Economics Research Conference*, pp. 33-36.
5. Budden, M. C. and Hsing, Y. Impacts of accountability indicators and socioeconomic factors on test scores and policy implications. *Journal of College Teaching & Learning*, Vol. 3, No. 4, pp. 21-25.
6. Cudd, M., Tanner, J. R. and Budden, M. C. (1989). The issue of student preparedness: Perceptions of finance professors. *Journal of Financial Education*, pp. 60-64.
7. DeWitz, S. J., Woolsey, M. L., and Walsh, W. B. (2009, January/February). College student retention: An exploration of the relationship between self-efficacy beliefs and purpose in life among college students. *Journal of College Student Development*, Vol. 50, No. 1, pp. 19+.
8. Harbrecht, A., Neidermeyer, P. E., and Tuten, T. L. (2006, October). Changes in higher education: How to address the learning needs of the Latino population. *Journal of College Teaching & Learning*, Vol. 3, No. 10, pp. 63-69.
9. Jacobs, J. and Archie, T. (2008). Investigating sense of community in first-year college students. *Journal of Experiential Education*, Vol. 30, No. 3, pp. 282-285.
10. Lehman, C. M. and Dufrene, D. D. (2002). *Business Communication*, 13th edition, Southwestern Publishing / Thompson Learning, United States.
11. Marklein, M. B. (2009). States consider basing college funding on graduation rates instead of enrollment. *USA Today*, Retrieved 8/26/2009, www.usatoday.com/news/education/2009-05-26-college-funding_N.htm.
12. Noser, T. C., Tanner, J. R. and Budden, M. C. (1992, Fall). The issue of student preparedness: Perceptions of economics professors. *Journal of Contemporary Business Issues*, Vol. 1, No. 1, pp. 13-26.
13. Remington, S., Guidry, J. A., Budden, M. C. and Tanner, J. R. (2000, Fall). When were the good old days? Revisiting perceptions of marketing students' prior preparation. *Journal of Marketing Education*, Vol. 22, No. 3, pp. 188-199.
14. Rivas, R. M., Sauer, P. L., Glynn, J. G. and Miller, T. E. (2007, 3rd Quarter). Persist/dropout differences in pre-matriculation attitudes of freshman towards college attrition: A longitudinal multiple group structural equations model. *College Teaching Methods & Styles Journal*, Vol. 2, No. 3, pp. 55-68.
15. Roach, R. (2008, July). College officials urged to monitor the graduation rate gap. *Diverse: Issues in Higher Education*, Vol. 25, No. 11, pp. 16-17.
16. Sander, W. and Krautmann, A. C. (1991). Local taxes, schooling, and jobs in Illinois. *Economics of Education Review*, Vol. 10, pp. 111-121.