

Measuring Predictors of Student Retention Rates

Allen L. Webster and Vince E. Showers

Department of Finance and Quantitative Methods, Bradley University, Peoria, IL, USA

Abstract: Problem statement: Colleges and universities place more emphasis on student retention rates than ever before. Educational institutions are intensifying efforts to discourage student departure and preserve their established student base. Economic pressures that bear heavily on academic administration make such efforts highly critical for institutional success. Research on this vital issue is especially important due to the vigorous competition among college and universities to recruit students and maintain enrollment levels. This pressure is heightened by the fact that it costs more to attract students than it does to retain them. **Approach:** This study uses institutional-based data to examine student attrition levels with the intent to identify their chief determinants and provide the foundation for post-secondary institutions to explore the viability of their own retention programs. Discriminant analysis is used to distinguish those schools that exhibit a higher degree of success in retaining student enrollments from those who suffer higher departure trends. **Results:** Tuition, student/teacher ratio and the amount of dollar aid offered the students all play substantial roles in encouraging persistence. The acceptance rate and enrollment levels were found to provide less discriminatory power. **Conclusion:** Students are influenced by the personal attention they receive and the manner in which their chosen institution of higher education caters to their individual needs, problems and concerns. Colleges and universities concerned with student retention would benefit from concentrated programs designed to provide individualized student services that address immediate student needs.

Key words: Student attrition, discriminant analysis, retention rates, enrollment levels

INTRODUCTION

Given the continued financial pressure on our nation's institutions of higher education resulting from budget shortfalls and in some cases, a declining level of high school enrollments, colleges and universities are facing a fiscal challenge on an epic scale. The U.S. Department of Education recently placed 114 private colleges on a watch list of schools that failed a fiscal-responsibility test because of insufficient equity, income and savings in 2008. Each suspect institution is subjected to a financial stress test that features composite scores based on three critical ratios. The important elements include the school's debt load, their expenses as compared to income and the overall resources of the institution. Some specialists worry that the financial conditions of such schools-particularly religious and liberal arts colleges with tiny endowments and budgets that depend heavily on tuition revenue-will deteriorate further as the recession lingers, raising the possibility that some could be forced to shut their doors. Colleges that fail the test are subject to extra monitoring on their use of federal student-aid funds.

In order to survive colleges and universities must become more responsive to rapidly changing circumstances and develop a willingness and ability to implement innovational policies leading to a continued growth in the productivity of the American educational system. Our ability to contend with an increasingly complex and competitive world depends directly on the worth of the educational attainment imbued to those who, in the future, will serve as our national leaders in all aspects of the business, legal and administrative areas of our national character. Our economic and political welfare depend on an effective didactic infrastructure that supports and fosters the highest attainable level of educational development.

According to a report by the Paris-based Organization for Economic Cooperation and Development (OECD), the 30-nation organization that develops yearly rankings comparing international educational achievement:

The United States is losing ground in education, as peers across the globe zoom by with bigger gains in student achievement and school graduations. Among adults age 25-34, the US is ninth among industrialized nations in

Corresponding Author: Allen L. Webster, Department of Finance and Quantitative Methods, Bradley University, Peoria, IL, USA

the share of its population that has at least a high school degree. In the same age group, the United States ranks seventh, with Belgium, in the share of people who hold a college degree. By both measures, the United States was first in the world as recently as 20 years ago...

The United States ranks 22nd among the 30 members of the Organization in terms of science scores (The 30 member countries of the OECD are: Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, South Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Turkey, United Kingdom and the United States.). The average score was 500. The U.S. reported a score of 489, one point above the Slovak Republic. The United States of America is the only OECD country where 25-34 year-olds are not better educated than 55-64 year-olds. It seems unequivocal that our nation's future is threatened by an impoverished labor base whose educational level has badly eroded over the past several years.

The report bases its conclusions about achievement mainly on international test scores released last December. They show that compared with their peers in Europe, Asia and elsewhere, 15-year-olds in the United States are below average in applying math skills to real-life tasks. Top performers included Finland, Korea, the Netherlands, Japan, Canada and Belgium.

To make matters worse, Facione (2009) states the great majority of institutions, public and private, are in dread of some of the deepest budget cuts in educational financing in more than a generation. Many anticipate reductions as much 36%. The most difficult situations menace those institutions that began the 2008-09 academic term with soft enrollments, depleted endowments, excessive discount rates, heavy debt service burdens, operating deficits and perhaps most ominous, the often-fatal tuition gap. Even a brief review of the literature is replete with worrisome accounts of pending financial stress placed on institutions of higher education. One universal response by these schools seems to be a greater emphasis placed on student retention. Since it costs less to retain a student than to recruit a new one, retention rates have become a prime concern of academic administrators charged with the fiscal care of their institutions.

This study examines the institutional characteristics of colleges and universities and how they relate to retention rates and overall student persistence. As fiscal pressures mount educational institutions find

themselves under an escalating strain to lessen their financial burdens. Any action school administrators might take or any programs they might employ to abate departure risk offers an almost irresistible appeal.

Literature review: It has been well established that college costs are rising much faster than disposable incomes, general inflation rates and the overall costs of living (Kaufman and Woglom, 2008; Kienzl *et al.*, 2007). Yet it is also a well-known fact that public appropriations for educational purposes have simply not maintained pace with the rising cost of education (Crampton, 2009; Cheslock and Gianneschi, 2008; Astin and Oseguera, 2004). Jointly, the confluence of these repressive constraints has had a pronounced detrimental effect on the financial plight facing our academic institutions.

Much of the past research into this crucial matter of student retention has focused on the academic preparation of entering freshmen (David and Fike, 2008; Salinitri, 2005; Jamelske, 2009). These studies centered on the nature of the students' high school preparation including such factors as its qualitative integrity, GPAs and class standing upon graduation, as well as demographic features including parental history, race, family structure and other personal characteristics. Other studies have examined the psychological characteristics of university students and their ability to adjust to the university environment (Elias *et al.*, 2010). From their study of university students in Malaysia, Elias *et al.* (2010) suggests the need for counseling sessions for first year students.

Less emphasis has been placed on unique and innovative efforts institutions have adopted in the battle to reduce student departure risk. Inexperienced students who find themselves in an alien setting that taxes their adaptive skills to survive an unfamiliar setting often require special handling. As they embark on a new and often unsettling experience, these students confront a newfound independence offering an array of sudden emotions and unforeseen issues. These anomalies range from homesickness and time management to financial pressures and unaccustomed scholastic obligations (Chaves, 2006). Adaptation to these sociological and psychological changes often proves too daunting, causing the student to abandon his or her educational goals.

To cope with this dilemma, many schools are developing their own "persistence programs" (Jamelske, 2009; Lillibridge, 2008). These specialized programs, often denoted as First Year Experience (FYE) or Academic Exploration Program (AEP), are designed to encourage students to continue their educational

pursuits. While schemes vary in form among schools, they all have one common factor - to comfortably integrate students into the college life style. Most often students find the process of launching a college experience to be an unsettling endeavor. Disoriented, apprehensive and confused, many fail to successfully establish an adaptive life style conducive to continued scholarly achievement. Retention programs are intended to ease this transition from a relative sheltered existence before high school graduation to the less certain world students confront upon entering college life.

Some designs entail an established curriculum beginning students must follow that administrators deem likely to further students' likelihood of successful assimilation into the college scene. Specific courses are identified to give the beginning student a flexibility allowing programs tailored to the needs and interests of each student. It serves as a uniquely designed program that assists students in identifying their academic interests, skill sets and career options that reflect personal values and aspirations. AEP programs provide students with the support, guidance and tools to determine the academic direction most likely to lead to the successful completion of a college degree. Students receive the personal attention to which they have become accustomed in their earlier academic experiences. Commonly, this dynamic program provides a weekly seminar class, individualized academic advising, an academic curriculum exhibit hosted by faculty members and job shadowing opportunities. Recognizing that each student has special talents and goals, professional academic advisors guide students through the decision-making process of selecting an area of study to which they can relate and in which they can develop a continued interest.

To the extent these administrative models achieve their intent student retention rates are likely to diminish. Retention can be used as a fundamental method to measure an institution's success in meeting student needs (Wild and Ebbers, 2002).

Many studies of retention focus on a survey approach in which a certain cohort of students at a single institution is followed from semester to semester. Cabrera *et al.* (1993) caution that this approach suffers from certain weaknesses. Findings from such studies are not readily generalized among institutions due to the diversity in the character of retention rates, institution type, student characteristics and other measurable and immeasurable factors. The findings in this study are based on a cross-section of four-year institutions varying in size, location, religion-base/unaffiliated and public/private status. This provides a more universal perception to the problem of student departure.

Factors generally recognized as influences on student retention rates cross a broad array of individual and institutional characteristics. Perhaps the most commonly cited deontological foundations for the study of persistence patterns are offered by Tinto (1994) theory of 'integration' and Astin (1999) principle of 'involvement'. Their iconic contributions to the field of student persistence are still heralded as the foundation for research into this all-important issue.

The former drew a metaphor between suicidal behavior as developed by Durkheim (1966) and academic withdraw. Both, Tinto argued, were the result of one's inability to socially and intellectually integrate into their environmental setting. Such failure will more likely result in 'death'.

This argument, validated by many subsequent studies (Astin 1997; Pascarella and Terenzini, 1991; Highbee *et al.*, 2005), places extensive emphasis on assimilation into the mix of college life. Building a strong sense of an inclusive educational and social community on campus is essential. Tinto argued that the student must become involved in many campus activities. A sense of belonging is essential to academic persistence. Austin's theory of 'involvement' paralleled that of Tinto. His premise called for strong efforts to generate acceptance on both the part of the student as well as the institution. A cordial association with faculty, campus residence and membership in campus organizations of both a social and professional nature were all seen as instruments in preserving a student presence. A student's socioeconomic status plays a major role in persistence as the educational level of the parents.

Common to both theories, students must meld into the fabric of college life and assume a position of belonging. A home-like atmosphere must prevail giving students a sense of fit or feeling that dissuades them from becoming disassociated with campus life.

More students discontinue their education as a result of inadequate preparedness for the social rigors of college life than do those who post a deficient academic record after arriving on campus (Salinitri, 2005) However, logic would dictate that there prevails a strong connection between academic survival and the psychological pressures that beset incoming freshmen. Those who are able to cope with the transition from high school are more likely to report favorable scholastic achievement.

National data: Before examining the results of the present study a brief review of national persistence levels would provide valuable insight into the general nature of the retention rates among colleges and universities. Table 1 provides a thumbnail image of retention rates within various types of institutions

Table 1: Summary Table: National-first-to-second-year retention rates by institutional type

Degree Level/Control	N	Std Dev	Mean (%)
Two-year Public	824	12.4	53.7
Two-year Private	102	25.6	55.5
BA/BS Public	78	12.5	67.6
BA/BS Private	390	17.5	69.9
MA/1st Professional Public	216	11.4	69.8
MA/1st Professional Private	494	13.7	72.0
PhD Public	228	10.6	74.4
PhD Private	228	11.9	80.6
Total	2,560	N/A	65.9

Source: <http://www.act.org/research/policymakers/pfd/retain2009.pdf>

Table 2: The selectivity categories used by the American college testing program are based on these general descriptions

Selectivity level	ACT middle 50%	SAT middle 50%*	Definition
Highly selective	25-30	1710-2000	Majority admitted from top 10% of H.S. class
Selective	21-26	1470-1770	Majority admitted from top 25% of H.S. class
Traditional	18-24	1290-1650	Majority admitted from top 50% of H.S. class
Liberal	17-22	1230-1530	Majority admitted from bottom 50% of H.S. class
Open	16-21	1170-1480	Generally open to all with H.S. diploma or equivalent

Source: <http://www.act.org/research/policymakers/pfd/retain2009.pdf>; *: SAT score ranges are a point-to-point concordance with ACT scores based on a joint study by ACT and the College Board

Table 3: First- to second-year mean retention rates for four-year colleges and universities by admission selectivity

	Bachelor's Only		Bachelor's and Masters		Bachelor's, Master's and Ph.D.		All	
	Public	Private	Public	Private	Public	Private	Public	Private
Highly Selective	92	93.6	90	82.1	92	91.4	90.4	92.3
Selective	87	71.9	81.7	81.4	82.4	82.3	82.2	81.5
Traditional	71	66.6	71.1	70.6	73.2	72.8	71.5	69.7
Liberal	63.6	60.3	65.3	61.3	63	69.3	64.3	61.4
Open	59.1	65.5	63.5	65.9	74.4	72.9	65.1	66.6
All By Column	67.9	69.9	69.8	72	77.6	80.6	72.9	73

Source: <http://www.act.org/research/policymakers/pfd/retain2009.pdf>

(ACT, 2008) (Founded in 1959, the American College Testing Program is an independent, not-for-profit organization that provides a broad array of information, research and assessment in the area of education.). Perhaps that feature of Table 1 that stands out the most is that in all cases private institutions have higher retention rates than do public institutions. Although beyond the scope of this study, this detail offers fertile area for research as to just why this attribute prevails. The mean overall rate is shown to be 66%.

Of equal interest is the relationship between retention and institutional selectivity. Table 2 provides a description of the selectivity levels identified by the ACT. Table 3 displays the mean retention rates for both public and private schools based on this hierarchy of selectivity. Perhaps not surprisingly, schools with more stringent selection requirements exhibit higher retention rates.

MATERIALS AND METHODS

As noted earlier, much of the past research has centered on the social and demographic characteristics

of the student in identifying retention risk as a prime example see Wells (2008). Cabrera *et al.* (1993) cautioned that findings based on student characteristics cannot be reliably generalized across institutions because patterns of persistence may vary among institutional types and settings. They should be used only as a point of departure for retention studies for any particular institution.

However, many individual institutions do not have the resources to construct and administer a retention program based on their own unique character. Thus, it is of great benefit to isolate predictor measures that identify conditions that promote student retention. This approach addresses the question as to what colleges and universities can do to more easily integrate incoming students into the persistent cohort of resolute undergraduate students. By following the resulting guidelines schools would be able to undertake actions that would elevate retention rates and mitigate financial pressures brought on by student turnover.

The sample data set is from 2005-06 comprising of 201 four-year institutions offering Bachelor, Master

and/or Ph.D. level degrees. For operational purposes the retention rate is defined as first-year-fall to second-year-fall persistence. The median retention rate is 79.00% with a mean of 78.95%.

A discriminate model is developed to classify schools into those that suffer low retention rates and those that are more successful in retaining student commitments. The response variable, retention rate, is metric and not categorical in its measurement. Therefore, in accord with an artificial dichotomy is created based on the median retention rate for all schools included in the sample. Schools with a retention rate below the median are identified as 0 while a class designation for those with high rates are placed in group 1. This process as a “polar approach” in which two or more extremes are identified and then compared via discriminate means. Such a methodology may detect patterns and distinctions not revealed using standard regression techniques. Such dichotomies will expose group differences not accessible by other statistical techniques. This is particularly true if the choice differences are not overly prominent when analyzed under assumptions upon which standard regressions models are built. Gambo and Yusuf (2010) provide an excellent summary of assumptions and advantages of discriminate analysis in their examination of the placement of students in courses of study in Nigeria’s National Diploma program.

Discriminate analysis was also chosen as the preferred methodology in order to foster the primary intent of the analysis to provide a classification system into which select schools could be categorized and thus compartmentalize institutions on the basis of retention characteristics. Standard regression procedures do not, of course, permit such balkanization.

The discriminate argument is the appropriate statistical tool to develop a linear combination of predictor variables that will discriminate best between a prior defined groups. Based on Fisher’s linear discriminant (Fisher, 1936), discriminant analysis is a classification method that maximizes the distance between the means of the two classes while minimizing the variance within each class. It is designed to maximize between-group variance relative to within-group variance. The probability that accurate groupings are achieved is enhanced by classifying data into groups so that this difference is maximized. If the variance between groups is large relative to that within groups then each observation is more likely to be classified into its respective group. It is more probable that an observation belongs in that group whose mean it is closer to than to a group with a more distant mean. More accurate classification is attained by grouping observations based on their relative position to group

means. By doing so, the within-group variation is minimized relative to the variation between observations in different groups, i.e., between-group variation.

Discriminant analysis is selected in preference to logit due to the desire to isolate those institutional features most associated with high retention rates. While logit techniques will clearly provide probabilities as to inclusion into one of two or more defined groups, discriminant analysis is more aptly designed to measure what characteristics explain the classification. Since the intent of this study is to assist schools in a restructuring more compatible with heightened student persistence it is imperative that such features be accurately identified. The discriminant function is expressed as:

$$Z = \beta_0 + \sum_{i=1}^k \beta_i X_i \tag{1}$$

Where:

β_0 = The constant term

$\beta_i X_i$ = The parameterized vector of predictor variables 1 through k

Given the configuration specified in (1), each case receives a discriminant score, Z.

Using Bayesian principles, discriminant analysis maximizes the between-group variance relative to the within-group variance. It classifies each case based on its discriminant score into that group in which it has the higher probability of belonging according to the standard conditional probability expression as shown by Eq. 2:

$$P(G_i|D) = \frac{[P(D|G_i) * P(G_i)]}{\sum [P(D|G_i) * P(G_i)]} \tag{2}$$

Where:

$P(G_i|D)$ = The posterior probability a case is in Group i given it has a specific discriminant score

$P(D|G_i)$ = The conditional probability a case has a discriminant score of D given it’s in group G_i

$P(G_i)$ = The prior probability the case is in G_i which is defined as:

$$n_i/N$$

Where:

n_i = The number of cases in group i

N = The total number of observations

Canonical discriminant function coefficients: After each observation is given a discriminant score using Eq. 1, these scores are averaged for each group. This mean discriminant score for a group is called its centroid. The average of the two centroids is the cutoff score used to

Table 4: Descriptive Statistics of Selected University Characteristics

	Minimum	Maximum	Mean	Std. Deviation
Retention rate	54.0	98.0	78.95	10.54
Acceptance Rate (AccRt)	13.0	99.0	65.49	20.44
Mean Dollar Amount of Aid to Students (DolAid)	1063.0	27292.0	13947.92	7169.70
Enrollment (Enroll)	516.0	40000.0	7400.81	7212.24
Percent of Financial Need Met by Aid (PctNeed)	18.0	100.0	78.09	16.77
Room and Board (RandB)	2910.0	11297.0	6696.19	1631.70
Faculty Salary* (FacSal)	45.5	134.2	78.71	18.75
Student/Teacher ratio (S/T)	7.0	24.0	15.25	3.90
Tuition (Tu)	2784.0	41975.0	14704.98	10887.93

*: Salary is for tenured full professors in thousands of dollars

classify each case into one of the groups. If both groups are of the same size, $n_1 = n_2$ and the variance/covariance matrices are equal, the cutoff score will be zero because both groups will have the same mean centroid except for sign. That is $Z_2 \bar{Z}_2 = -\bar{Z}_1$ where \bar{Z}_i is the mean discriminant score, the centroid, for group i . Those observations with lower Z scores will take on a negative value and be classified into one group while those with high scores will carry a positive sign and be classified into the other group.

The cutoff point is the simple unweighted arithmetic mean of the two centroids (Sharma, 1996). However, if n_1 and n_2 are not equal the cut off is the simple unweighted mean of the centroids:

$$(\bar{Z}_1 + \bar{Z}_2) / 2 \tag{3}$$

and it is not equal to zero. Although the two groups in the present study are classified on the basis of the median retention rates they are not of the same size due to missing data for some of the variables. The cutoff score must be determined on the basis Eq. 3. The centroids are -1.078 for the low-retention group and 0.821 for the high retention group. The cut off score is therefore -0.1285. Those cases with a discriminant score less than -0.1285 are classified into the low-retention group since that group had the lower mean centroid of -1.078. Cases with a discriminant score greater than -0.1285 are classified in the high-retention group since that group had the higher mean centroid (0.821).

Table 4 provides descriptive statistics for the variables used in the study. The variables are selected as a descriptors of the institutions and do not characterize students as do so in many earlier studies. The individual acceptance rates of the colleges and universities report a mean of 65.5%. It should be expected that schools with higher acceptance rates would also experience lower retention rates since an increasing acceptance rate generally would correspond

to a lowering of academic standards for incoming students. Two measures of financial aid given to attending students reflect the institution's readiness and capacity to assist students with the growing cost of education. They reflect on the schools' proclivity to provide aid and thus focus on institutional conduct and not that of individual students. An increase in either type of aid would be expected to help retain students who face financial challenges after entering college. In line with the research purpose to capture institutional effects and avoid any impact on retention of the demographic features of the students, the remaining variables for tuition and room and board along with salaries, student/teacher ratios and overall enrollment also reflect the characteristics of the schools and not their students. Both enrollment and the student to teacher ratio are expected to be inversely related to retention rates since larger classes would provide less one-on-one academic interaction between student and instructor and an increase in overall enrollment could potentially lead to a student feeling "lost in the crowd" in both academic and social settings. In a related manner, faculty salaries are expected to be positively related to retention rates assuming salaries are reflective of teaching performance. The "cost" variables, tuition and room and board, are more ambiguous in their directional impact. Generally, one would expect higher tuition to reflect higher academic standards; however, regional and public versus private institutional differences may dampen this impact.

Table 5 shows some of the results of the discriminate analysis. The discriminant function is reported using the unstandardized coefficients and is:

$$Z = -9.545 + 0.016AccRt + 0.027DolAid + 0.020Enroll - 0.006PctNeed + 0.012RandB + 0.045FacSal + 0.186S/T + 0.050Tu$$

While the unstandardized coefficient estimates are used in estimating Z scores, the standardized coefficients shown in Table 5 serve in much the same

Table 5: Canonical discriminant function coefficients

	Unstandardized Coefficients	Standardized Coefficients	Loadings
AccRt	0.016	0.298	-0.320
DolAid	0.027	0.804	0.703
Enroll	0.020	0.338	-0.059
PctNeed	-0.006	-0.083	0.326
RandB	0.012	0.225	0.461
FacSal	0.045	0.698	0.675
S/T	0.186	0.693	-0.494
Tu	0.050	0.806	0.704
(Constant)	-9.545		

capacity as standardized coefficient values in a multiple regression model. Sometimes referred to as ‘weights’, they measure the relative importance in terms of their contributions to the power of the model to properly discriminate among or between groups.

Those predictor variables with the higher absolute values provide greater discriminatory power. Tuition, student/teacher ratio and the amount of dollar aid offered the students all play substantial roles in encouraging persistence. While these factors come as no surprise, the significant discriminatory power of faculty salaries is subject to speculation. This problematical outcome may stem from the correlation between salaries and other predictor variables as evidence of multicollinearity. A problem discussed shortly. The acceptance rate and enrollment levels seem to provide less discriminatory power. Finally, the negative relationship shown by PctNeed is somewhat paradoxical.

However, standardized coefficients present problems of interpretations common to regression standardized coefficients. Small weighting factors may suggest the associated variable’s contribution to the grouping power is minimal or it may be interpreted to imply that it is fractionalized by some degree of multicollinearity.

Absolute levels of the predictor variables can also affect the weight measures. Had the metric of any of the variables been much higher or much lower the weights would have been different. Discriminatory weights are also subject to excessive instability. Therefore, some caution must be exercised in their interpretation. This final issue of instability is addressed below in the evaluation of the model. Nevertheless, they are reported here as is the custom in many discriminant studies.

Discriminant loadings, also called structural coefficients, are being used with increasing frequency due to the aforementioned deficiencies inherent in weights. Loadings measure the pooled with-in group correlation between each independent variable and the discriminant function. These loadings are displayed in Table 5. Results based on loadings will often prove dissimilar to those derived from weights as seen in the table. This incongruity is a consequence of the fact that loadings measure only with-in group variations. While

loadings are considered to be considerably more reliable in interpreting the relative discriminatory power of the individual independent variables, they are still subject to instability and caution should be exercised in their application. Interestingly, however, the estimated loadings were consistent with expectations concerning directional impacts while the signs of the discriminant weights for several explanatory variables (acceptance rate, aid to students, enrollment, percent of need met and student/teacher ratio) were not consistent. With higher absolute values of loadings representing a higher correlation between the explanatory variable and the discriminant score, the results indicate a relatively strong positive correlation between tuition, mean dollar amount of aid to students and faculty salaries.

RESULTS AND DISCUSSION

Table 6 presents the classification results and the ‘hit ratio’. The latter measures the percentage of cases in the sample set that were correctly classified as belonging to a specific group. It reflects the function’s ability to discriminate between the actual group membership and the predicted group membership. Of the 85 low retention institutions the model predicted 73, or 85.9% of them would fall in that group. Of the 113 high retention schools, the model classified 102 of them, or 90.3%, to fall in that group. This represents an 88.4% ‘hit rate’ for the model as a whole. The portion of Table 6 dealing with cross validation is discussed later.

Evaluation: Comparing the hit ratio to chance models is necessary at this point to determine if the model carries any discriminatory power. Chance models contrast the function’s hit ratio with what could be obtained just by mere chance. That is, does the model perform any better than pure happenstance? The Maximum Chance Criterion is one method of comparison open to models with unequal group sizes. It is most often used when the objective of discriminate analysis is to maximize the percentage of cases that are properly classified. It is applied by merely assigning all observations to the largest, or larger as in this case, group. Since 113 out of the 201 cases fell in the high retention group the Maximum Chance Criterion is 57.1%. The hit rate of the present model of 88.4% is well above chance and strongly suggests that it carries useful discriminatory power.

The Proportional Chance Criterion is also a revealing method of evaluating a model. It is found as $C_p = P^2 + (1-P)^2$ where P is the proportion of cases in either group. Given the proportion of those institutions in the high retention group is 57.1%, $C_p = 45.1\%$. Again, this chance occurrence is significantly less than the hit rate of 88.4% provided by the model.

Table 6: Classification results by high and low retention rates^{b,c}

		Predicted groupings		
		Low	High	Total
Original	Actual groupings			
	Low	73	12	85
	High	11	102	113
	Ungrouped	3	0	3
	Percentages			
	Low	85.90%	14.10%	100.00%
cross validation ^a	High	9.70%	90.30%	100.00%
	Ungrouped	100.00%	0%	100.00%
	Low	69	16	85
	High	16	97	113
	Percentages			
	Low	81.20%	18.80%	100.00%
	High	14.20%	85.80%	100.00%

^a: Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case; ^b: 88.4% of original grouped cases correctly classified; ^c: 83.8% of cross-validated grouped cases correctly classified

This significance of the model can be tested using PRESS'S Q Statistic. An acronym for prediction sum of squares, the PRESS'S Q statistic is:

$$Q = \frac{[N - (n)(g)]^2}{N - (g - 1)} \quad (4)$$

Where:

- N = The size of the entire sample
- n = The number of correct classifications
- G = The number of groups

In the present case Q = 117.3. The test is based on a χ^2 statistic with one degree of freedom. Thus, $\chi^2(1, 0.005) = 7.897 < 117.3$. The null hypothesis that the model is no better than chance is rejected with a p-value of 0.00.

Testing the global model also uses a χ^2 test based on the Wilk's Lambda statistic. A Wilk's Lambda, 7, of 0.523 is reported which translates into a χ^2 of 44.015 with k = 8 degrees of freedom where k is the number of predictor variables. The level of significance is again reported to be 0.00.

Complete confirmation of the model requires a process of cross validation. Only in this manner can it be determined how accurate the model is in classifying the cases. Cross validations involves using a different data set to test the classification accuracy. This can be done dividing the data set into two parts. One part, referred to as the analysis sample is used to estimate the model. The second part, called the hold-out sample, is used to test the accuracy of the model. This approach is suggested if the initial data set is quite large and dividing it in half still provides ample data to derive the discriminant function.

Using the same data to form the model and then using that model to classify those data, a process referred to as post hoc prediction, will result in an upward bias (Frank *et al.*, 1965). That action is analogous to shooting the arrow into a wall and then drawing the target around the arrow. The archer is certain to hit the bull's eye every time.

An equally popular method of cross validation made easier by advancements in computer technology is the Mahalanobis procedure that relies on the generalized Euclidian distance that adjusts for unequal variances. Mahalanobis distance is the statistical distance between two points that takes into account the covariance and correlation among variables. For a k-variable case, the Mahalanobis distance between two points is:

$$M_{ip} = (X_i - X_p)'S^{-1}(X_i - X_p) \quad (5)$$

Where:

- i and p = Two observations
- X = A k×1 vector and
- S = A k×k covariance matrix.

For uncorrelated variables, that is, if $r \approx 0$, S is a matrix with variances on the main diagonal and for uncorrelated standardized variables S is an identity matrix. The large variance/covariance matrices used in the present study are not presented here in the interest of brevity.

In order to use the Mahalanobis distance to classify a test point as belonging to one of i classes, it is necessary to first estimates the covariance matrix of each class based on samples known to belong to each class. Then, given a test sample of one or more observations, it remains to compute the Mahalanobis distance to the average or center mass of each class for

that set of observations. The test point is then identified as belonging to that class for which the Mahalanobis distance is minimal. This is equivalent to selecting the class with the maximum likelihood.

Still another approach of estimating the misclassification rate is the leave-one-out method. This is the method used here. The discriminant function is estimated using all but one of the observations. That function is then used to classify the observation that was excluded. This is done once for each observation in the sample. All observations are classified using a discriminant model unaffected by the observation being classified thus avoiding the 'archer's conundrum' mentioned above. In addition, it offers the advantage of using virtually all of the data that have been collected except for the single case left out and doesn't require dividing the data set into smaller less efficient units.

The results are shown in Table 6. Of the 85 schools in the low-retention group, 69, or 81.2%, were correctly classified. Further, 97 of the 113, or 85.8% of high-retention schools were placed in the proper group. Overall, 83.8% of the cross-validated cases were correctly classified. This is just under the 88.4% hit rate reported by the original model. This minimal upward bias testifies to the model's ability to reliably classify cases into their proper groups.

A comparison of between group variance and within-group variance often starts with an eigenvalue, λ . In matrix algebra, an eigenvalue is a constant, which if subtracted from the diagonal elements of a matrix, results in a new matrix whose determinant equals zero. If $\lambda = 0.00$, the model has no discriminatory power. The larger the value of λ , the greater the discriminatory power of the model. Using the eigenvalue it is possible to calculate the canonical correlation coefficient, R_c , for the model. The square of this coefficient, R_c^2 , also called the canonical root, measures the percentage of the variation in the dependent variable that is accounted for by the predictor variables. It measures the correlation of the discriminant function with the discriminant scores. As it approaches one, that is, unity, it can be reasoned that most of the variance in the discriminant scores can be attributed to group differences. The closer to one, the more discriminant power the model displays and the better the model's performance in classifying observations between or among groups.

It is calculated as:

$$R_c = \sqrt{\frac{\lambda}{1 + \lambda}} \quad (6)$$

The eigenvalue of 0.91 for the present model resulted in a correlation of 0.69 and a canonical root of 47.6%. Thus, 47.6% of the variation between the two groups is accounted for by the discriminator variables.

Summary: The importance of maintaining a high retention rate among our nation's colleges and universities cannot be overstated. The financial squeeze in which these institutions find themselves requires that cost-cutting measures of all forms must be considered as well all sources of revenues. Eight variables have been selected to discriminate between those schools able to maintain an elevated retention level and those who suffer pronounced drop-out rates. Each variable isolates some aspect of the schools' environment as encountered by the students. While other variables, such as public-private affiliation, are important and should be examined, this study concentrates on variables that can be, to some extent, controlled or changed by the university administration either in the short or medium runs. Examining both standardized weights and loadings, the discriminant model used in this study indicates tuition, dollar aid to students and faculty salaries are highly correlated with freshman retention.

Both the Proportional Chance Criterion and the Maximum Chance Criterion were used to determine if the model predicted any better than pure chance. In both cases the probabilities of accurate groupings were much higher using the model than merely relying on pure chance. The Proportional Chance Criterion proved to be 45.1%. This means that if grouping is left merely to chance, only 45.1% would be properly classified. The Maximum Chance Criterion was only 57.1%. The model's hit ratio of 88.4% for the original grouping and a cross-validation rate of 83.8% compares quite favorably to chance groupings and offers support for the reliability of the model. Further support for the model is found in the PRESS'S Q statistic of 117.3 which yields a Chi-square value of 7.897 and a p-value of 0.00.

Furthermore, using the Wilk's Lambda statistic, 7, to test the model as a whole supports its accuracy. The reported 7 of 0.523 translates into a χ^2 of 44.015 with $k = 8$ degrees of freedom where k is the number of predictor variables. The level of significance is again reported to be 0.00. The null hypothesis that the none of the variables in the model offer discriminatory power is firmly rejected. Finally, a canonical root of 47.6% suggests that the model does indeed offer insight into the discriminatory power of the institutional variables used in the study.

CONCLUSION

The results of this analysis suggest that certain institutional characteristics that promote student retention can be accurately identified. By placing the emphasis on the character of the schools rather than the demographic features of the student, it becomes possible to transfer the analysis from one college or university to another. Schools are able to compare their unique set of conditions to those identified here as conducive to student persistence.

As measured by the discriminant loadings, both the variables "Tuition" and "Mean Dollar Amount of Aid to Students" showed a positive relationship to retention levels. Certainly, financial aid by the institution should help retain students facing limited financial resources; however, one might expect higher student costs (e.g., tuition) to counter this impact. The positive relationship between tuition and retention rate is likely reflective of higher tuition being highly correlated with higher entrance standards. A positive relationship was also found with faculty salaries, perhaps reflecting the importance of quality faculty in retaining students. Related to this, was the negative impact between the "Student/Teacher Ratio" and retention indicating the importance of smaller class size in student retention. The variables "Acceptance Rate", "Percent of Financial Need Met by Aid" and "Room and Board" also showed loadings above .3 in absolute terms. Both percent of financial need met by aid and room and board costs were positively related to retention rate reflecting similar motivations for retention as that of amount of financial aid provided and tuition. Not surprisingly, acceptance rate showed a negative loading indicating universities and colleges with higher acceptance rates of incoming freshmen can expect greater difficulty in retaining those students.

REFERENCES

- ACT, 2008. National collegiate retention and persistence to degree rates. ACT, Inc. http://www.act.org/research/policymakers/pdf/retain_2006.pdf
- Astin, A.W., 1999. Student involvement: A developmental theory for higher education. *J. Coll. Stu. Pers.*, 40: 518-529.
- Astin, A.W. and L. Oseguera, 2004. The declining "equity" of American higher education. *Rev. Higher Edu.*, 27: 321-341. DOI: 10.1353/rhe.2004.0001
- Astin, A.W., 1997. *What Matters in College: Four Critical Years Revisited*. 1st Edn., Jossey-Bass, Pennsylvania, ISBN-10: 078790838X, pp: 512.
- Cabrera, A.F., A. Nora and M. Castaneda, 1993. College persistence: Structural equations modeling test of an integrated model of student retention. *J. Higher Edu.*, 64: 123-39.
- Chaves, C., 2006. Involvement, development, and retention theoretical foundations and potential extensions for adult community college students. *Commun. Coll. Rev.*, 34: 139-152. DOI: 10.1177/0091552106293414
- Cheslock, J.J. and M. Gianneschi, 2008. Replacing state appropriations with alternative revenue sources: The case of voluntary support. *J. Higher Edu.*, 79: 208-229.
- Crampton, F.E., 2009. Spending on school infrastructure: Does money matter? *J. Edu. Admin.*, 47: 305-322. DOI: 10.1108/09578230910955755
- David, S.F. and R. Fike, 2008. Predictors of first-year student retention in the community college. *Commun. Coll. Rev.*, 36: 68-88. DOI: 10.1177/0091552108320222
- Durkheim, E., 1966. *Suicide: A Study in Sociology*. 1st Edn., The Free Press, New York, ISBN-10: 0029086604, pp: 405.
- Facione, P.A., 2009. Adaptive budgeting: Thirty-four ideas for raising revenues, cutting costs, retaining students, and saving jobs in hard times. *Liberal Edu.*, 95: 24-31.
- Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. *Annals Eugenics*, 7: 179-188.
- Frank, R.E., W.F. Massey and D.G. Morrison. 1965. Bias in multiple discriminant analysis. *J. Market. Res.*, 2: 250-258.
- Gambo, A.I. and M.W. Yusuf, 2010. An application of multivariate analysis in modeling students placement in Nigerian higher institutions. *J. Math. Stat.*, 6: 350-356. DOI: 10.3844/jmssp.2010.350.356
- Elias, H., N. Noordin and R.H. Mahyuddin, 2010. Achievement motivation and self-efficacy in relation to adjustment among university students. *J. Soc. Sci.*, 6: 333-339. DOI: 10.3844/jssp.2010.333.339
- Highbee, J.L., D.R. Arendale and D.B. Lundell, 2005. Using theory and research to improve access and retention in developmental education. *New Direc. Commu. Coll.*, 129: 5-15. DOI: 10.1002/cc.181
- Jamelske, E., 2009. Measuring the impact of a university first-year experience program on student GPA and retention. *Higher Educ.*, 57: 374-391. DOI: 10.1007/s10734-008-9161-1

- Kaufman, R.T. and G. Woglom. 2008. Managing private college finances in an environment in which spending and revenues grow at different rates. *J. Educ. Finance*, 34: 196-211.
- Kienzl, S.G., A., Mariana and M. Tatiana, 2007. The effect of local labor market conditions in the 1990s on the likelihood of community college students' persistence and attainment. *Res. Higher Edu.*, 48: 751-774.
- Lillibridge, F., 2008. Retention tracking using institutional data. (Fall) *New Direc. Commu. Coll.*, 143: 19-30. DOI: 10.1002/cc.332
- Pascarella, E.T. and P.T. Terenzini, 1991. *How college Affects Students: Findings and Insights From Twenty Years of Research*. 1st Edn., Jossey-Bass Press, USA., ISBN-10: 1555423043, pp: 894.
- Salinitri, G., 2005. The effects of formal mentoring on the retention rates for first-year, low achieving students. *Canadian J. Educ.*, 28: 853-873.
- Sharma, S., 1996. *Applied Mutivariate Techniques*. 1st Edn., John Wiley and Sons, New York, ISBN-10: 0471310646, pp: 493.
- Tinto, V., 1994. *Leaving College: Rethinking the Causes and Cures of Student Attrition*. 1st Edn., University of Chicago Press, Chicago, ISBN-10: 0226804496, pp: 312.
- Wells, R., 2008. The effects of social and cultural capital on student persistence. *Commun. College Rev.*, 36: 25-46.
- Wild, L. and L. Ebbers, 2002. Rethinking student retention in community colleges. *Commun. Coll. J. Res. Practice*, 26: 503-519.