DOCUMENT RESUME

ED 397 726 HE 029 326

AUTHOR	Huesman, Ronald L., Jr.; And Others
TITLE	Identifying Students at Risk: Utilizing Traditional and Non-Traditional Data Sources. AIR 1996 Annual Forum Paper.
PUB DATE	May 96
NOTE	23p.; Paper presented at the Annual Forum of the Association for Institutional Research (36th, Albuquerque, NM, May 5-8, 1996).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Academic Persistence; College Entrance Examinations; *College Freshmen; College Sophomores; Dropout Characteristics; Dropout Prevention; Higher Education; *High Risk Students; Identification; Information Sources; Institutional Research; Models; *Predictor Variables; Student Financial Aid
IDENTIFIERS	*AIR Forum; University of Iowa

ABSTRACT

A study tracked 3,192 University of Iowa freshmen through their first year and into their second year on campus. Logistic regression analyses using multiple data sources (admissions and registrar files, a standardized entrance test (the American College Testing Program Assessment) student profile section, an entering freshman survey) were conducted to determine models of student persistence at two points: freshman year spring re-enrollment and sophomore year fall re-enrollment. Two relatively successful models for predicting sophomore persistence were derived, although both models predicted correctly rather small percentages of non-persisters. A followup study on the reasons given for withdrawal by false-positives in the model is recommended. For this sample, it was found that modeling persistence was related to college-level academic indicators; the only non-academic factor to enter the prediction equation was students' perceived need for financial aid. (Contains 17 references.) (MSE)

54

Identifying students at risk: Utilizing traditional and non-traditional data sources

Ronald L. Huesman, Jr.
Evaluator, Retention Services
The University of Iowa
300 Jefferson Building
Iowa City, IA 52242-1470
(319) 335-0356

Joyce E. Moore
Director, Evaluation & Examination Service
The University of Iowa
300 Jefferson Building
Iowa City, IA 52242-1470
(319) 335-0356

Chi-Yu Huang
Research Assistant
The University of Iowa
300 Jefferson Building
Iowa City, IA 52242-1470
(319) 335-0356

Shuqin Guo
Research Assistant
The University of Iowa
300 Jefferson Building
Iowa City, IA 52242-1470
(319) 335-0356

The Association for Institutional Research
36th Annual Forum
May 5 - 8, 1996
Albuquerque, New Mexico.

2

AIR

for Management Research, Policy Analysis, and Planning

This paper was presented at the Thirty-Sixth Annual Forum of the Association for Institutional Research held in Albuquerque, New Mexico, May 5-8, 1996. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum Papers.

Jean Endo
Editor
AIR Forum Publications

Abstract

The purpose of this research was to identify factors from a wide variety of traditional and non-traditional data sources that impact student persistence. Persistence models developed on the 1994 entering freshmen class from a large public midwestern university will be used to identify students at-risk in subsequent classes. Persistence was modeled at two points in time: re-enrollment in the spring term and subsequent re-enrollment in the fall of the sophomore year. Logistic regression procedures were used to identify students at-risk and cross validation procedures within the 1994 class provided an assessment of the accuracy and validity of the models.

Identifying students at-risk: Utilizing traditional and non-traditional data sources

This paper is one part of a larger research project on student retention being conducted at The University of Iowa. Over 3,200 1994 entering freshmen have been tracked through their first year and into their second year on campus. Logistic regression analyses utilizing multiple data sources were conducted to determine models of student persistence at two points in time; spring semester re-enrollment within the freshmen year and re-enrollment in the following fall term of the sophomore year. It is hoped that results from this study will aid in identifying students at-risk and developing appropriate intervention strategies.

Related literature

This study is in part a replication of the longitudinal study of factors affecting student persistence done by Gillespie and Noble (1992). They argued that many current studies of student persistence fail to include important non-traditional variables found to be related to persistence and that a more encompassing approach to retention should be explored. The models developed by Gillespie and Noble were based on readily available academic and demographic variables along with responses from surveys designed to assess information central to Tinto's model (goal commitment, institutional commitment, academic fit/integration, etc.). The models were both term specific as well as institutional specific as stressed by Tinto (1975) and Bean (1986). Gillespie and Noble found that no single variable or group of variables was present across institutional models however, those variable clusters that were present closely mirrored Tinto's model. The generalizability of the results was severely limited by loss of data due to a low response rate to the surveys, missing data, the large number of variables examined and a low dropout rate within the first semester of the freshmen year. This study attempts to control those factors found to adversely affect the work done by Gillespie and Noble. The focus of the study is on identifying institution specific variables related to persistence and using this information to not only identify students at-risk but to use the model as an advising tool.

Retention studies in the past have utilized discriminant analysis regression procedures, Terenzini & Pascarella (1977), Pascarella & Terenzini (1980), Pascarella, Duby, Miller and Rasher (1981), Getzlaf, Sedlacek, Kearny & Blackwell (1984), and Delaney (1993). However, recent studies have shown logistic regression to be a useful tool for studying student persistence, Ott (1988), Gillespie and Noble (1992), Molnar (1993). Logistic regression procedures outperform discriminant analysis in terms of error rates and the types of errors made in classifying students as persisters or non-persisters (Huesman, Moore, Druva-Roush, Wang & Huang (1994). As a statistical method, logistic regression can be used to guide decisions regarding the potential risk of a student not persisting and provide a means for assessing the accuracy of those decisions.

Data

Predictor variables selected for this study came from several sources and were based on the traditional pre-enrollment variables and academic indicators emphasized by Pascarella, Duby, Miller & Rasher (1981) and the more encompassing selection reflected in Gillespie and Noble (1992). Traditional information was obtained from university admission and registrar files (basic demographics, ACT test scores, GPA, etc.). In addition, selected items from the ACT Assessment Student Profile Section (SPS) and a Entering Freshmen Survey (EFS) were included in the data set. The SPS provided more detailed background information on high school coursework, family income and extracurricular activities. The EFS was a customized version of an instrument used by Gillespie and Noble (1992) and is designed to measure factors from Tinto's (1975) model of student retention.

The following variables were identified as potential predictors of student retention:

I. Background information

a. Demographic characteristics (age, gender, race, college distance from home, size of home community & racial composition of high school)
b. Academic achievement indicators (ACT test scores, high school rank, high school GPA, subjects studied & years studied)
c. High school extracurricular activities (music, debate, clubs, athletics, etc.)
d. Financial (family income & residency status)
e. Academic and personal needs (expected need for help in writing, reading, study skills, math, personal, occupational and educational planning)

f. Family's attitude toward education (parents education level, parent's attitude regarding attending college in general and this institution in particular, financial support and perceived financial hardship)
g. College admitted to (Liberal Arts or Engineering)

II. Initial commitment to institution

a. Institutional choice (this institution a first, second, third choice, etc.)
b. Purposes/reasons for enrolling
c. Planned enrollment status (full-time/part-time)
d. Primary educational goal (no goal, transfer, Bachelor's degree, etc.)
e. Importance of institutional characteristics in attending (ratings of admission materials, social, academic reputation, physical characteristics, etc.)

III. Initial academic goal commitment

a. Expected degree & strength of certainty
b. Choice of career/major and certainty of choice
c. Expectations of academic life (expected grades, hours of study)
d. Concerns about the value of going to college

IV. Student/institution academic fit

a. Course enrollment/completion, grades
b. Expectations of relationships with faculty, staff and advisors

V. Student/Institution social fit

a. Concerns with discrimination by faculty & students
b. Expectations for making friends & peer support
c. Opportunities for active social life, extracurricular activities, etc.

VI. Student/institution financial fit

a. Concerns with having enough money to stay in school
b. Expected family support
c. Type of financial aid (loans, grants, scholarships, etc.)

The two criterion variables of student persistence defined were: 1) spring semester re-enrollment within the freshmen year and 2) re-enrollment in the following fall term of the sophomore year.

Method

Data collection

The sample under consideration consisted of first time freshmen entering the university in the fall of 1994. The cohort definition by default resulted from the administration of the EFS during the summer of 1994. Surveys were mailed to incoming freshmen and collected at orientation sessions throughout the summer. A total of 2,956 usable surveys were collected from the target population of 3,210 entering freshmen for a return rate of 92% (see Table 1).

Table 1

Description of freshmen cohort administered the EFS

Status	Returned EFS	Did not return EFS	Total
Matriculator	2,924	247	3,171
Non-matriculator	60	117	177
Summer session admit	12	2	14
Stopouts	32	7	39
Other*	3	0	3
Total	3,031	373	3,404

* 1 deceased student, & 2 high school students

The final analysis group (n=3,192) excluded stopouts from first to second semester (n=18). Summer session admits were not included in the analysis group in order to follow more closely the definition of a freshmen cohort used by the Office of the Registrar. Information from the EFS was available for 92.2% of the analysis group; the ACT Assessment was taken by 95% of the analysis group; and information from the ACT Assessment SPS was available for 88.5% of this group.

Analysis

Several steps were taken to reduce the number of potential predictor variables under consideration. The first step involved creating factor scores from selected items of the EFS and the SPS surveys in order to stabilize the results and aid in the interpretation of the regression analyses (Noble & Gillespie, 1992). An SAS Principal Component Analysis of 114 items selected from the EFS was conducted. A six factor solution using varimax rotation provided the necessary variable reduction. The resulting solution accounted for 24.3% of the variance. Only items with factor loadings greater than or equal to .30 were selected for inclusion in the creation of the factor scales (Bryman and Cramer, 1990, Kim and Mueller, 1978). Also, items that loaded strongly on more than one factor were not included (Bryman and Cramer, 1990). The six EFS factors included items with the following common themes: campus support; personal concerns; financial need; academic concerns and goals; and university contacts and recruitment. Factor scores were calculated for each individual using the SAS factor score procedure . In order to increase the number of factor scores produced, missing values on the EFS were replaced with item means. A second SAS factor analysis of 79 items selected from the SPS was conducted. A five factor solution using varimax rotation provided the necessary data reduction. The resulting solution accounted for 18.1% of the variance. The five factors included involvement in: athletics; special interest groups, leadership, service, science; music activities; government, debate, speech and drama; and art activities. A collinearity diagnosis was conducted with the remaining variables and the newly created factor scores to detect the presence of collinear relationships among the data and the severity of such relationships (Belsley, Kuh, & Welsch, 1980). Two or more variables with high variance decomposition proportions (>=.5) associated with a high condition index greater than or equal to 30 were removed from the regression analysis. (p. 112, Belsley et al., 1980). The remaining variables were examined for redundancy, low response rates and timeliness of data in relation to intervention strategies for at-risk students. Table 2 contains a description of the reduced selection of predictor variables.

Table 2

Predictor variables examined

Variable code	Variable description
ACT_COMP	ACT Assessment composite test score
ACTFS1	High school music activities
ACTFS2	High school student government/debate/speech/drama activities
ACTFS3	High school athletic activities
ACTFS4	High school special interest groups/service/leadership/science activities
ACTFS5	High school art activities
ATH_CODE	Recruited athlete
COLLEGE	Undergraduate college admitted to
EFSF1	Campus support
EFSF2	Personal concerns
EFSF3	Institutional concerns
EFSF4	Financial need
EFSF5	Academic concerns and goals
EFSF6	University contacts and recruitment
GENDER	Male/Female
GPA_943*	Fall GPA earned
HS_RANK	High school rank
RACE_R	Racial categories (Asian, Minority, & White)
RATIO1*	Credit hours earned fall semester/Credit hours enrolled
RATIO2*	Credit hours earned spring semester/Credit hours enrolled
RESD	Resident status
YHMATH	Years of high school math studied (Geometry, Algebra, & Higher math)
YHSCIEN	Years of high school science studied (Chemistry & Physics)

* not included in regression analysis for spring semester re-enrollment

Forward step-wise logistic regression analyses using SPSS version 6.1.1 for the Macintosh were conducted using the remaining 23 predictor variables. Logistic regression is a method specifically developed to examine a dichotomous dependent variable. The logistic regression model is represented as:

$$Index = b_0 + b_1 x_1 + b_2 x_2 + \ldots. + b_p x_p$$

The index is created from a weighted combination of predictor variables (x_1, ...x_p), b_0 denotes the intercept, and b_1, ...b_p are the estimated raw score regression coefficients. Logistic regression provides an estimate of the probability of a case being in a particular group.

$$probability = \frac{1}{1+e^{-Index}}$$

Where e=2.718 represents the base of the natural logarithm. The Index is different than the predicted value resulting from an Ordinary Least Squares multiple regression. The Index represents the log odds of persistence, not the predicted value of the criterion. Therefore, the regression coefficients also differ in that they represent the degree of change in the log odds of persistence given a one unit change in x (Gillespie and Noble, 1992).

For spring re-enrollment 20 of the 23 variables were selected and entered into a forward step-wise logistic regression using the entire analysis group (n=3,192), see table 2. The information provided by the three variables excluded from the spring semester re-enrollment analysis (GPA_943, RATIO1 & RATIO2) would not be available for use in an early intervention model. A cross-validation of the spring re-enrollment model was not possible because of the small group of non-persisters (n=107) and the number of variables under consideration. For sophomore re-enrollment a randomly selected sample (calibration group) was used to develop the models (n=1,596). The remaining students (validation group) were used to cross validate the selected models and compare the classification errors based on selected cutoffs (n=1,596). The cross-validation of a fitted model to a sample different from the one used to develop the model is an important check on external validity, since the method mathematically capitalizes on chance idiosyncrasies in the data and tends to be overly optimistic. The use of a validation group yields more realistic estimates of the classification results (Stevens, 1992, Hosmer and Lemeshow, 1989). For sophomore re-enrollment all 23 variables were included in the initial analysis. However, a second post hoc logistic analysis was conducted without RATIO2 to determine if its absence greatly affected the predictive ability of the model. This was done because RATIO2 would not be available to advisors until a student had nearly completed the second semester of the freshmen year thus delaying intervention for students at-risk.

A decision table like the one shown is used to illustrate the predictive accuracy and classification errors made using these models . A decision table is created by determining a critical point (i.e., cut score) at a probability value which results in a classification of students as persisters (those at or above the critical point) and non-persisters (those below the critical point).

Decision table

	Predicted Group	
Actual Group	Non-persister	Persister
Non-persister	A	B
Persister	C	D

Cell A: Correctly identified Non-persisters
Cell B: False positives: Non-persisters identified as Persisters
Cell C: False negatives: Persisters identified as Non-persisters
Cell D: Correctly identified Persisters

Observations in cells A & D are referred to as "hits" (i.e., correct decisions). Cells B & C represent "misses", (i.e., incorrect decisions). A + D represents the number of correct decisions made. When this sum is represented as a proportion it is referred to as the accuracy rate. The sum of the observations in cells A & C represent the identified at-risk group. The accuracy rate and the severity of the decision errors are the means by which cutoffs are selected.

Results

Nearly 3.4% (n=107) of this cohort failed to re-enroll in the spring semester of their freshmen year after completing the fall semester. Approximately 11% (n=349) failed to re-enroll for their sophomore year. Table 3 provides summary information of the metric variables selected for analysis.

Table 3

Summary statistics for Persisters/Non-persisters (spring re-enrollment and sophomore re-enrollment)

	Spring semester re-enrollment*		Sophomore re-enrollment**	
Variable code	Mean	Std. Deviation	Mean	Std. Deviation
ACT_COMP	24.5/24.6	3.6/3.7	24.6/24.0	3.6/3.5
ACTFS1	.01/-.05	.95/.93	.01/.04	.96/.92
ACTFS2	.00/-.03	.90/.90	.00/.00	.89/.91
ACTFS3	.00/-.15	.88/.75	.01/-.01	.88/.86
ACTFS4	.01/-.23	.93/1.0	.02/-.04	.92/.96
ACTFS5	-.01/.02	.90/.94	-.03/.10	.88/.99
EFSF1	.00/-.04	.95/.87	-.02/.15	.94/.97
EFSF2	-.01/.13	.92/.89	-.01/.00	.92/.96
EFSF3	.01/-.26	.90/.81	.02/-.05	.89/.96
EFSF4	-.02/.11	.91/.90	-.05/.24	.89/.95
EFSF5	.00/-.01	.87/.78	.02/-.06	.87/.84
EFSF6	.00/-.07	.87/.89	.02/-.08	.87/.87
GPA_943Δ	272/193	71/117	281/213	61/91
HS_RANK	74/68	18/20	75/69	18/18
RATIO1	.93/.70	.15/.36	.95/.82	.11/.25
RATIO2	.	.	.93/.72	.14/.31
YHMATH	3.8/3.7	.72/.73	3.9/3.7	.72/.70
YHSCIEN	1.7/1.6	.67/.64	1.7/1.6	.67/.65

*Spring semester re-enrollment: sample sizes varied between n = 2,758 - 3,085 for persisters and n = 93 - 107 for non-persisters

**Sophomore re-enrollment: sample sizes varied between n = 2,343 - 2,625 for persisters and n = 311 - 349 for non-persisters

Δ GPA calculated without decimal

The results of the logistic regression analyses are shown in Tables 4, 5 & 6. For spring re-enrollment only two of the 20 potential predictors entered the regression: HS_RANK & ACTFS3 (see Table 4). Four variables entered the sophomore re-enrollment model of student persistence: EFSF4, GPA_943, RATIO1& RATIO2 (4-variable model, see Table 5). Table 6 contains the results of the forced entered logistic regression of EFSF4, GPA_943 & RATIO1 on sophomore persistence (3-variable model).

Table 4

Logistic regression model for predicting spring semester re-enrollment

Variable	B	S. E.	Wald	df	Sig	R	Exp(B)
ACTFS3	0.3611	0.1620	4.9705	1	.0258	0.0665	1.4350
HS_RANK	0.0205	0.0061	11.1279	1	.0009	0.1165	1.0207
Constant	2.0652	0.4330	22.7442	1	.0000		

Table 5

Logistic regression model for predicting sophomore re-enrollment (with RATIO2)

Variable	B	S. E.	Wald	df	Sig	R	Exp(B)
EFSF4	-0.3352	0.1029	10.6089	1	.0011	-0.0940	0.7152
GPA_943	0.0114	0.0018	38.9832	1	.0000	0.1948	1.0114
RATIO1	1.8958	0.6813	7.7432	1	.0054	0.0768	6.6578
RATIO2	2.7068	0.4352	38.6934	1	.0000	0.1940	14.9817
Constant	-4.7589	0.6003	62.8390	1	.0000		

Table 6

Logistic regression model for predicting sophomore re-enrollment (no RATIO2)

Variable	B	S. E.	Wald	df	Sig	R	Exp(B)
EFSF4	-0.3867	0.1001	14.9239	1	.0001	-0.1146	0.6793
GPA_943	0.0148	0.0017	73.6905	1	.0000	0.2700	1.0149
RATIO1	2.2305	0.6480	11.8496	1	.0006	0.1001	9.3044
Constant	-3.5947	0.5431	43.8136	1	.0000		

Tables 7,8 & 9 provide a comparison of actual versus predicted outcomes for the three models of student persistence (i.e., decision tables, default critical value p=.5). The accuracy rate of the spring re-enrollment model of student persistence was 97%, but the model failed to classify any of the non-persisters correctly. The 4-variable model of sophomore persistence yielded a 90.2% accuracy rate versus 89.6% for the 3-variable model.

Table 7

Decision table for predicting spring re-enrollment using critical value p = .5 (accuracy rate = 97%)

	Predicted Group		
Actual Group	Non-persister	Persister	Total
Non-persister	0	75	75
Persister	0	2408	2408
Total	0	2483	2483

0% of the spring semester non-persisters were correctly identified
0% of the predicted at-risk group were actual spring semester non-persisters
97% of the predicted spring semester persisters were actual persisters
100% of the spring semester persisters were correctly identified

Table 8

Decision table for predicting second year re-enrollment using 4-variable model with critical value p = .5 (accuracy rate = 90.2%)

	Predicted Group		
Actual Group	Non-persister	Persister	Total
Non-persister	36	123	159
Persister	12	1210	1222
Total	48	1333	1381

22.6% of the 2nd year non-persisters were correctly identified
75% of the predicted at-risk group were actual 2nd year non-persisters
90.8% of the predicted 2nd year persisters were actual 2nd year persisters
99% of the 2nd year persisters were correctly identified

Table 9

Decision table for predicting second year re-enrollment using 3-variable model with critical value p = .5 (accuracy rate = 89.6%)

	Predicted Group		
Actual Group	Non-persister	Persister	Total
Non-persister	27	132	159
Persister	12	1210	1222
Total	39	1342	1381

17% of the 2nd year non-persisters were correctly identified
69.2% of the predicted at-risk group were actual 2nd year non-persisters
90.2% of the predicted 2nd year persisters were actual 2nd year persisters
99% of the 2nd year persisters were correctly identified

Both models of sophomore re-enrollment incorrectly classified less than 1% of the persisters as non-persisters (i.e., false negatives) For the 4-variable model 22.6% of the non-persisters were correctly identified as non-persisters and for the 3-variable model 17% of the non-persisters were categorized correctly. Adjustments to the critical values can be made that maintain a given accuracy rate while increasing the number of identified non-persisters however, this is accompanied by an increase in the number of persisters incorrectly categorized as non-persisters (i.e., false negatives). Figures 1-3 illustrate this situation. Figure 1 illustrates the accuracy rate for both models of sophomore re-enrollment at selected critical values. The overall accuracy rate is quite high and consistent for both models of sophomore re-enrollment till a critical value of p=.70 is reached. Figure 2 shows that the percentage of non-persisters correctly identified for the two models increases as the cutoff is adjusted higher.

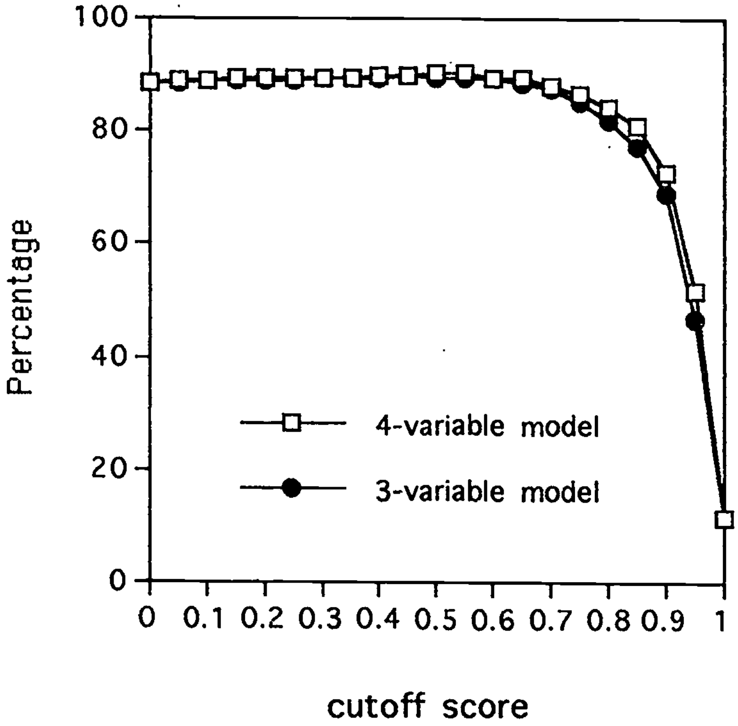

Figure 1. Predictive accuracy: Second year persistence model

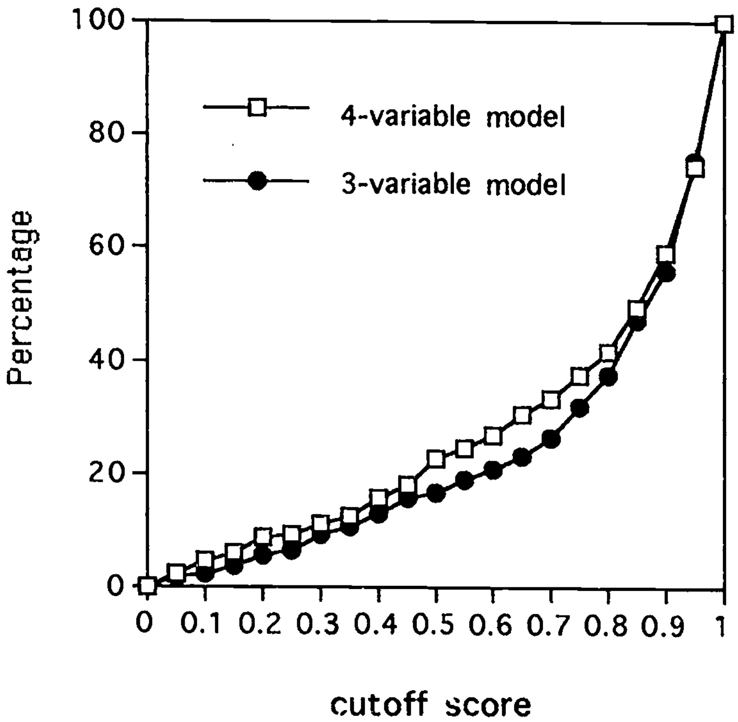

Figure 2. Percentage of 2nd year non-persisters correctly identified

Figure 3 illustrates that as the number of correctly identified non-persisters increases so does the number of persisters identified in the at-risk group for both models. As a consequence, the proportion of actual non-persisters in the identified at-risk group decreases as the cutoff is raised.

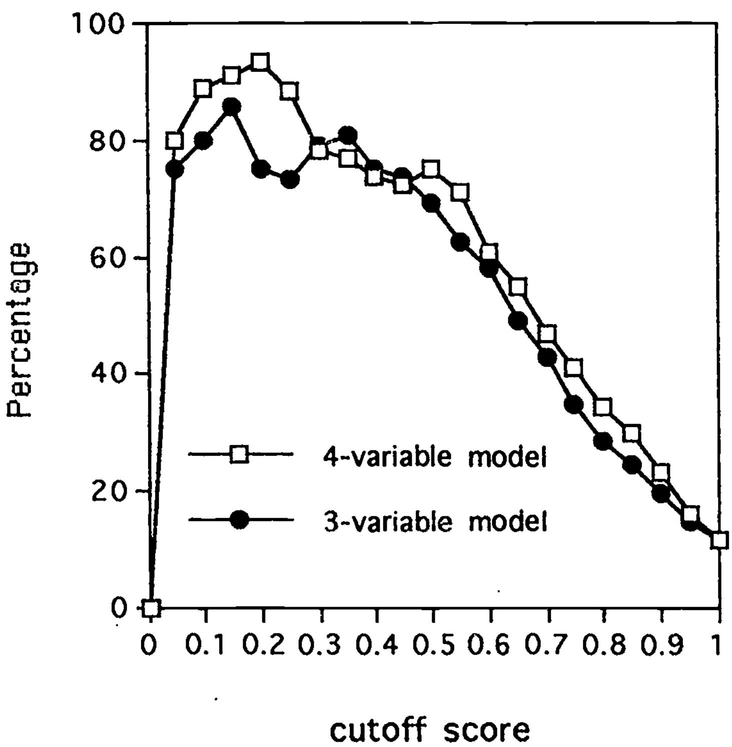

Figure 3. Percentage of 2nd year non-persisters in predicted at-risk group

For example, if the critical value is set at p=.65 versus p=.50 for the 4-variable model the percentage of non-persisters correctly identified increases 8.2% (n=36 vs. n=49) but the percentage of persisters in the identified at-risk group increases by 20% (n=40 vs. n=12). Table 10 contains the decision table for the 4-variable model using p=.65 as the critical value. The decision table for the 3-variable model using p=.65 (see Table 11) shows a similar pattern to the 4-variable model. The percentage of non-persisters correctly identified increases 6.3%

(n=27 vs. n=37) however, the percentage of persisters in the at-risk group increases by 20% (n=38 vs. n=12).

Table 10

Decision table for predicting second year re-enrollment using 4-variable model with critical value p = .65 (accuracy rate = 89%)

	Predicted Group		
Actual Group	Non-persister	Persister	Total
Non-persister	49	110	159
Persister	40	1182	1222
Total	89	1292	1381

30.8% of the 2nd year non-persisters were correctly identified
55.1% of the predicted at-risk group were actual 2nd year non-persisters
91.5% of the predicted 2nd year persisters were actual 2nd year persisters
96.7% of the 2nd year persisters were correctly identified

Table 11

Decision table for predicting second year re-enrollment using 3-variable model with critical value p = .65 (accuracy rate = 88%)

	Predicted Group		
Actual Group	Non-persister	Persister	Total
Non-persister	37	122	159
Persister	38	1184	1222
Total	75	1306	1381

23.3% of the 2nd year non-persisters were correctly identified
49.3% of the predicted at-risk group were actual 2nd year non-persisters
90.7% of the predicted 2nd year persisters were actual 2nd year persisters
96.9% of the 2nd year persisters were correctly identified

Discussion

Both of the models of sophomore re-enrollment do a very good job of identifying those students who will persist to their sophomore year. However, the percentages of non-persisters correctly identified is rather small for both models. This may be due to the academic nature of the predictors that entered the regression as well as the nature of withdrawal, whether a student withdrew voluntarily or not. Tinto (1987) and Ott (1988) agree that this distinction should be made since the outcome results from different patterns of interaction between the student and the institution. For the 4-variable model 77.8% of the non-persisters identified were not permitted to register for their sophomore year due to academic probation (two consecutive semesters with a cumulative GPA less than 1.70) and this percentage increases under the 3-variable model to 88.9%. For the non-persisters that were not identified (false positives n=123) only 20% were on academic probation at least once (only 11 students were not permitted to register due to academic probation) under the 4-variable model. For the 3-variable model 25.8% of the false positives (n=132) were on academic probation at least once (20 students were not permitted to register due to academic probation). The predictive accuracy and the types of errors made are of major importance but it is equally important to ask what happens to misclassified students beyond the time period that was modeled?

Currently a follow-up study of reasons given for withdrawal from the false positives is being conducted and should shed some light on why this group of students is failing to persist. An Enrolled Student Survey (ESS) administered to the freshmen cohort during the spring semester of their first year provided some additional information. In general the ESS responders from the non-persister group tended to be less satisfied with their experiences on campus than those responders who persisted, but this statement needs to be tempered by the fact the ESS had a return rate of 37% for this cohort and may not be an accurate reflection of this group. Following the false negatives through their first semester and re-enrollment in the spring semester of their second year demonstrates that these students are still at-risk. For the 4-variable model; 58% have not persisted within their second year, 92% were on academic probation at least once (50% more than once). For the 3-variable model using the same critical value; 25% have not persisted within their second year, 100% were on academic probation at least once (42% more than once).

The results of this study show that for this sample, modeling persistence is related to college level academic indicators such as GPA and course completion ratios. The only non-academic variable to enter into the prediction equation was the EFS financial need factor score. The influence of high school academic indicators (high school rank, courses studied, extracurricular activities), and affective measures from the EFS and ACT SPS did not materialize in this sample. The lack of success with pre-enrollment predictors, in particular the spring re-enrollment model may be due in part to the small proportion of students (3.4%) who did not re-enroll for the spring semester. Gillespie and Noble (1992) encountered a similar problem in their multi-institutional study. The question that needs to be answered at this point is whether or not it is important to pursue aggregating information across classes to establish an early model of at-risk students (i.e., within the freshmen year) or, is it more important to focus on the sophomore re-enrollment model to intervene with those students who persist through their freshmen year but fail to re-enroll.

References Cited

Bean, J. P. (March, 1986). Assessing and reducing attrition. In D. Hossler (Ed.). Managing College Enrollments: Number 53. New Directions for Higher Education (pp. 47-61). San Francisco: Jossey-Bass, Inc.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: John Wiley & Sons.

Bryman, A. and Kramer, D. (1990). Quantitative Data Analysis for Social Scientists. New York: Routledge, Inc.

Delany, A. M. (1993). Methodological Issues in Studying Minority Student Retention. A paper presented at the Annual Forum of the Association for Institutional Research in Chicago, Illinois.

Getzlaf, S. B., Sedlacek, G. M., Kearney, K. A., & Blackwell, J. M. (1984). Two types of voluntary undergraduate attrition: Application of Tinto's model. Research in Higher Education, 20(3), 257-68.

Gillespie, M. & Noble, J. (1992). Factors Affecting Student Persistence: A Longitudinal Study. (Report No. 92-4). Iowa City, Iowa: American College Testing Program.

Hosmer, D. & Lemeshow, S. (1989). Applied Logistic Regression. New York: John Wiley & Sons, Inc.

Huesman, R., Moore, J., Druva-Roush, C., Wang, X., & Huang, C. (1994). A Comparison of Decision Theoretic Approaches to Studying Student Retention. A paper presented at the Annual Forum of the Association for Institutional Research in New Orleans, Lousinana.

Kim, J.and Mueller, C. (1978). Factor Analysis: Statistical Methods and Practical Issues. (Sage University Paper series on Quantitative Applications in the Social Sciences, Series no. 07-14). Beverly Hills and London: Sage Publications.

Molnar, D. (1993). The Impact of Mission Effectiveness on Student Retention. A paper presented at the Annual Forum of the Association for Institutional Research in Chicago, Illinois.

Ott, M. D. (1988). An analysis of predictors of early academic dismissal. Research in Higher Education, 28(1), 34-48.

Pascarella, E. T., Duby, P. B., Miller, V. A., & Rasher, S. P. (1981). Pre enrollment variables and academic performance as predictors of freshman year persistence, early withdrawal, and stopout behavior in an urban, nonresidential university. Research in Higher Education, 15(4), 329-49.

Pascarella, E. T., and Terenzini, P. T. (1980). Predicting freshman persistence and voluntary dropout decisions from a theoretical model. Journal of Higher Education, 51(1), 60-75.

Stevens, J. (1992). Applied Multivariate Statistics for the Social Sciences: second edition. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Terenzini, P. T. and Pascarella, E. T. (1977). Voluntary freshman attrition and patterns of social and academic integration in a university: A test of a conceptual model. Research in Higher Education, 6, 25-43.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. Review of Educational Research, 45, 89-125.

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