DOCUMENT RESUME

ED 298 851 HE 021 842

AUTHOR	Bloom, Allan M.; And Others
TITLE	Anticipating Mathematics Performance: A Cross-Validation Comparison of AID3 and Regression. AIR 1988 Annual Forum Paper.
PUB DATE	May 88
NOTE	22p.; Paper presented at the Annual Forum of the Association for Institutional Research (28th, Phoenix, AZ, May 15-18, 1988).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Achievement; Achievement Tests; Calculus; *College Mathematics; Higher Education; Institutional Research; Mathematics Tests; Models; Multiple Regression Analysis; *Predictive Validity; Regression (Statistics); *Required Courses; Standardized Tests; *Statistical Analysis; Test Validity
IDENTIFIERS	*AIR Forum; College Board Achievement Tests

ABSTRACT

In response to the increasing importance of student performance in required classes, research was conducted to compare two prediction procedures, linear modeling using multiple regression and nonlinear modeling using AID3. Performance in the first college math course (College Mathematics, Calculus, or Business Calculus Matrices) was the dependent measure. Independent measures included demographics, high school achievement data, and College Board test scores. Using "type of mathematics" as either an i 'ependent dummy variable or as a moderator variable was considere. The data came from over 3,500 students who were beginning freshmen in 1986 and 1987 and who took one of the three math courses. The models were developed on the 1986 beginning freshmen and were compared to the sample on which they were developed and cross-validated on the 1987 freshmen. The following conclusions were reached: (1) the statistical models were useful in predicting performance, particularly among those who are predicted to have difficulty; (2) actual ability to explain performance "shrinks" from the modeling sample to the use sample; and (3) the 70% unexplained variance in performance calls from the developing of more discriminating critical variables. Data are presented in several tables and three appendices. Contains 17 references. (Author/KM)

ED 298851

Anticipating Mathematics Performance:

A Cross-Validation Comparison of AID3 and Regression

Allan M. Bloom

David B. Agard

Gerald W. McLaughlin

Eugene W. Carson

Institutional Research and Planning Analysis
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061-0433

Abstract

Student performance in required classes has become increasingly important. This research compares two prediction procedures, linear modeling using multiple regression and nonlinear modeling using AID3. Performance in first college math course is the dependent measure. Independent measures include demographic, high school, and pre-admissions test scores. Using "type of mathematics" as either an independent dummy variable or as a moderator variable is considered.

The models are developed on entering 1986 beginning freshmen. The models are compared for the sample on which they are developed and also cross validated on the 1987 freshmen class.

Paper presented at the 28th
Association for Institutional Research Forum
Phoenix, Arizona, May 1988

This paper was presented at the Twenty-Eighth Annual Forum of the Association for Institutional Research held at the Hyatt Regency Phoenix Hotel, Phoenix, Arizona, May 15-18, 1988. 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.

Teresa Karolewski
Chair and Editor
Forum Publications Editorial
Advisory Committee

Anticipating Mathematics Performance

Introduction

The quest to predict student performance in specific courses has been with Institutional Research since before there was a discipline with that name. It will doubtless continue. Given the diversity of student and institutional characteristics, no single set of predictor variables is likely to answer more than the most specific of questions for a specific institution (Baird, 1984). As with most such matters, any contribution to the Institutional Research profession involves the method used in addressing the question. While the present research was directed to predicting performance of entering freshmen in their first mathematics courses, we believe the analytic methods apply to the general class of problem, and that application of those methods will advance the body of knowledge that constitutes Institutional Research.

Perspectives

There are several reasons to be able to anticipate academic success and failure. Relative academic success is strongly related to student attrition, even voluntary attrition (Pascarella & Terenzini, 1980). It allows effective academic counseling (Morgan et al, 1985). It permits coherent course and curriculum planning by the faculty (Craney & Armstrong, 1985). Further, being able to anticipate, and therefore possibly to avoid, academic failure could save a student from being depressed (personal observation of the third author, the only one of us who ever actually flunked anything).

Predicting performance in freshman mathematics has been of persistent interest in this regard. In Fall 1987, mathematics accounted for 16 percent of freshman level enrollments at Virginia Tech and fully half of the failures (D and F grades) awarded at the freshman level. This differs hardly at all from the situation of 20 years ago. Mathematics and the physical sciences accounted for one fifth of the grades, and two fifths of the failures, at the City University of New York (Bahn et al, 1967).

High school grades (Darling, 1983), standardized tests, and specialized topic tests have been the traditional predictor variables for academic success (Carmichael et al, 1986; Craney & Armstrong, 1985; Welch et al, 1982) despite problems with standardized tests and grades (Baird, 1984; Darling, 1983). The analytic approach tends to be multiple regression, or multiple classification analysis (all listed references), samples are often not large (Carmichael, 1986), and results tend to be expressed in terms of variance explained. Also the success of these estimates depend on student and institution characteristics (Baird, 1984). More recently, researchers have expanded on the traditional procedures, investigating non-linear analytic methods (Sonquist et al, 1970). Others have noted the importance of cross validation of regression analyses (Herzberg, 1969) as compared with investigating a broad range of variables (Merante, 1983). As these improvements, added to adequate sample size, improve the ability to predict academic success or failure adequately, administrators are pressing researchers to express the results in ways the decision makers can use.

Data

The data for this study came from over 3500 freshmen who entered Virginia Tech as beginning freshmen in the years 1986 and 1987 and subsequently took one of three different mathematics courses. The initial group of about 6000 was reduced by about one half based on the availability of scores on a math achievement test from College Board (CEEB). Models are developed for each course as well as with course coded in dummy variables. Other variables similarly coded as dummy variables include gender and race. A large number of high school measures were also included. For the use of AID3 the independent variables were categorized to either five or three levels with equally spaced intervals and equality of size at each level if at all possible. Table 1 gives the description of each independent variable as well as its abbreviation used in regression equations and AID3 diagrams with the number of levels. The models are developed on the half entering in 1986 and cross validated on the half entering in 1987.

The three courses used are College Mathematics (Mathematics), Calculus (Calculus), and Business Calculus Matrices (Matrices). In the following sections the combined model refers the situation where the course has been included as a regressor coded as a dummy variable.

Variables

The independent variables listed in Table 1 are defined below. Note that the regression abbreviations associated with each independent variable are enclosed in parentheses after each of the associated variables.

- Gender (SEXD): The gender of the students was coded with female = 0 and male = 1;

- Ethnic (RD1,RD): The ethnic background of the students was coded to indicate "white", "Asian American", or "other". Two binary variables were used to achieve three unique code combinations for these groupings, e.g., whites were coded 0,1; Asian Americans 1,0; and, all others (Blacks, Spanish Americans, and American Indians) were 0,0. The "other" group was made up mostly of Blacks.

- Course (CD1,CD): The math courses in which the students enrolled at the University. These were the three primary math sequences: calculus, business calculus and matrices, and college mathematics (pre-calculus). Only the performance in the first term (entry course) of each of these sequences was considered. The three courses were coded in a manner similar to the binary ethnic coding.

- SAT Math (SATMAT): This is the highest SAT mathematics score reported to the University.[1] These scores were used to create five groupings with break points beginning at 500 and increasing by 50 to 650

- SAT Verbal (SATVRB): This is the highest SAT verbal score reported to the University. These scores were used to create five groupings with break points beginning at 450 and increasing by 50 to 600

- Math Ach. Test (MACH): The mathematics achievement score (as administered by Educational Testing Service) were used create five groups with the break points starting at 50 and increasing by five for each for grouping.[2]

- H.S. Class Rank (HRANK): The numerical ranking from the top of the class as reported by the high school. The ranks were divided into five groups at the break points of 15, 35, 60, and 120.

- H.S. Class Size (HSIZE): The size of the student's graduating class as reported by the high school. The high schools were divided into five groups with the divisions by school sizes being made at 200, 300, 400, and 500.

- Rank Percentile (PTLE): The class rank percentile was computed by dividing class rank by class size.

- Total Units (TOTALU): This is the total number of high school units of academic courses attempted by each student. The units are in Carnegie units of a full year of study for each course.[3]

- Total Adv. Units (TOTALAU): The total number of high school units of advanced level courses in which each student enrolled. Advanced level courses are those identified by the high school as being Advanced Placement, Honors, Gifted and Talented, or International Baccalaureate courses.[4]

---

[1] If more than one SAT score was reported, the highest score was used.

[2] There were very few math achievement tests at the B level. Thus, no distinction was made to analyze the math achievement scores independently. If sufficient numbers of each type of examination scores are available, consideration should be given to a weighting or separation of these scores.

[3] The source of these data were the the high school transcripts for each new freshmen student. These data were entered into a transcript data base. All academic courses were entered and only academic courses are included. That is, physical education, driver education, vocational and similar courses are not included.

[4] The source of these data is the the same transcript data base discussed in the previous footnote.

- Total G.P.A. (TOTALGPA): The high school overall grade point average as reported by the high school.[5] Note that this may include non-academic courses as defined above and is not the grade average for only the academic courses entered into the transcript data base.

- Math Units (MATHU): The total number of high school units of math in which the student enrolled.

- Math Adv. Units (MATHAU): The total number of advanced level high school units of math in which the student enrolled. Advanced level courses are the same as those defined under the variable Total Adv. Units above.

- Math G.P.A. (MATHGPA): The grade point average for all the high school mathematics courses in which the student enrolled.

- Science Units (SCIENCEU): The total number of high school units of science courses in which the student enrolled. Science courses are laboratory based and typically include chemistry, biology, and physics.

- Sci. Adv. Units (SCIAU): The number of units of advanced science courses in which the student enrolled. The science courses are the same as the laboratory sciences defined above.

- Science G.P.A. (SCIGPA): The grade point average for all the high school science courses in which the student enrolled.

---

[5] All grade point averages used in this study were based on a 4.0 scale with A = 4.0 and F = 0.0 quality credits per unit.

Table 1

Independent Variables and Model Abbreviations

Description	Regression Abbr.	No. of Abbr. levels (AID3)
Gender	SEXD	2
Ethnic	RD1,RD2	3
Course	CD1,CD2	3
SAT Math	SATMAT	5
SAT Verbal	SATVRB	5
Math Ach. Test	MACH	5
H.S. Class Rank	HRANK	5
H.S. Class Size	HSIZE	5
Rank Percentile	PTLE	5
Total Units	TOTALU	5
Total Adv. Units	TOTALAU	3
Total G.P.A.	TOTALGPA	5
Math Units	MATHU	3
Math Adv. Units	MATHAU	3
Math G.P.A.	MATHGPA	5
Science Units	SCIENCEU	3
Sci. Adv. Units	SCIAU	3
Science G.P.A.	SCIGPA	5

Analysis

In this section we present the results from the development of the AID3 and regression models. The first question that arises concerns the equality of the regression models for each course (see Rao, pp. 237-238). Table 2 gives the error sum of squares and degrees of freedom for the full regression models of each course and the regression model with the courses pooled together. Combined (1) is the pooled model with no information on course, Combined (2) has course identification as a dummy variable (0,1). Table 3 gives the ANOVA table for the test of equality of regression equations using the model with dummy variables for courses.

Table 2

Sums of Squares and Degrees of Freedom of Regression Models

Model	Error SS	Error df
Mathematics	372.43	362
Calculus	654.82	717
Matrices	421.57	412
Combined (1)	1809.89	1529
Combined (2)	1492.42	1527

Table 3

ANOVA Table for Test of Equality of Regressions

Source	df	SS	MS	F-value	p-value
Deviation from hypothesis	36	43.60	1.2110	1.25	$p > .05$
Separate regressions	1491	1448.82	.9717		
Combined regression	1527	1492.42	1.1837		

From these results we see that separate regressions for each course are not appropriate. However, for completeness and comparative purposes, we include results from the separate models in the remainder of the study. Table 4 gives the full model R-square for the combined as well as the individual course regression models along with the respective sample sizes and the number of regressors.

Table 4

Full Model R^2 for Regression Models

Model	N	No. of vbls.	R^2
Combined	1548	20	.3425
Mathematics	381	18	.3007
Calculus	736	18	.4110
Matrices	431	18	.2757

Because of the non-significance of many regressors in the full model, a forward selection procedure was employed to develop the regression models. A .10 significance level was set for entry into the model.

Table 5 gives the order of entrance of variables in the forward selection procedure for the combined and individual course models for both AID3 and regression. The sequence of entry is left blank if the variable did not enter.

Table 5

Order of Entry for Variables for AID3 and Regression Analyses

Variable	AID3				Regression			
	Comb	Math	Cal	Matr	Comb	Math	Cal	Matr
Course	2	*	*	*	2	*	*	*
Gender	4				11			
Ethnic					14		5	
SAT Math	6			3	8	5		
SAT Verbal		6	4		7	4		
Math Ach	3	1	1	2	3	1	2	1
HS Rank				6	9			
Hs Size	9	5		7	6		4	5
Rank Percent	13	4						
Total Units	10				12			
Total Adv Unit	11	7		5	10			
Total GPA	1	3	2		1	6	1	2
Math Units	5	2	5		5	3	6	
Math Adv Unit	7		6					
Math GPA	12		3	4	4		3	3
Science Unit			7	1				
Sci Adv Unit							7	
Sci GPA	8				13	2		6

* Since groups split on this variable, it could not enter in these models

A first comparison between the AID3 and regression model shows that the early splits in the AID3 branching process are similar to the variables entered early in the forward selection procedure. The most interesting variables are performance on the Math Achievement Test (MACH) and Total GPA. These variables enter in 15 out of 16 of the models, often as one of the first three variables.

Because of the non-significance of many regressors in the full model, a forward selection procedure was employed to develop the regression models. A .10 significance level was set for entry into the model.

Table 5 gives the order of entrance of variables in the forward selection procedure for the combined and individual course models for both AID3 and regression. The sequence of entry is left blank if the variable did not enter.

Table 5

Order of Entry for Variables for AID3 and Regression Analyses

	AID3				Regression			
Variable	Comb	Math	Cal	Matr	Comb	Math	Cal	Matr
Course	2	*	*	*	2	*	*	*
Gender	4				11			
Ethnic					14		5	
SAT Math	6			3	8	5		
SAT Verbal		6	4		7	4		
Math Ach	3	1	1	2	3	1	2	1
HS Rank				6	9			
Hs Size	9	5		7	6		4	5
Rank Percent	13	4						
Total Units	10				12			
Total Adv Unit	11	7		5	10			
Total GPA	1	3	2		1	6	1	2
Math Units	5	2	5		5	3	6	
Math Adv Unit	7		6					
Math GPA	12		3	4	4		3	3
Science Unit			7	1				
Sci Adv Unit							7	
Sci GPA	8				13	2		6

* Since groups split on this variable, it could not enter in these models

A first comparison between the AID3 and regression model shows that the early splits in the AID3 branching process are similar to the variables entered early in the forward selection procedure. The most interesting variables are performance on the Math Achievement Test (MACH) and Total GPA. These variables enter in 15 out of 16 of the models, often as one of the first three variables.

The number of mathematics units (MATHU) and grades in mathematics (MATHGPA) also seem to make a significant, unique contribution to explaining math performance. In the combined models, course identifier is extremely important.

Considering a more quantitative comparison, Table 6 gives the number of variables included and the R-square for the forward selection procedure. Also, Table 6 gives the number of final groups and the average size in the AID3 procedure and the R-square found by a one-way analysis of variance on the final groups. From this we see the two prediction procedures are very comparable in explaining the variability. AID3 performs better in the course models due to the ability of the procedure to achieve a smaller number of observations in each group. Before addressing the performance of the procedures in predicting success we wish to exhibit some descriptive statistics depicting the accuracy of the procedures. Table 7 gives the mean and median of the absolute differences between the actual grade and the predicted grade for each procedure. Again we see the slight edge AID3 has shown over the regression model.

Table 6

R-squares of AID3 Versus Forward Selection

	AID3			Regression	
Model	No. groups	Average size	R^2	No. vbls.	R^2
Combined	30	51.6	.3025	15	.3406
Mathematics	12	31.75	.3077	6	.2937
Calculus	18	40.9	.4096	7	.3964
Matrices	14	30.8	.2918	6	.2545

Table 7

Absolute Differences Between Actual and Predicted Grade

Model	AID3		Regression	
	Mean	Median	Mean	Median
Combined	.8151	.6808	.7938	.6834
Mathematics	.7899	.6962	.8003	.6994
Calculus	.7639	.6394	.7869	.6983
Matrices	.7586	.6045	.8034	.6867

Our main objective in this study is to assess the ability of each procedure to predict success in a math class. Success is defined as achieving a grade of at least a "C". Predicted success is defined as having a grade of C or higher as a predicted grade. Table 8 gives the actual proportion of students that were successful as well as the proportions of success predicted by the two procedures. We see that in all cases the two procedures were overly optimistic in predicting success. This comes from means above the cut point and resulting regression to the mean.

Table 8

Actual and Predicted Proportions of Success

Model	Actual	AID3	Regression
Combined	.741	.803	.854
Mathematics	.756	.929	.900
Calculus	.694	.728	.755
Matrices	.807	.942	.947

For a test of the two procedures Table 9 presents the test statistic and p-value for a test of equality of proportion of success. For example, in a test that the proportions of success predicted by the two procedures are equivalent in Table 8, we find for the combined model class a Z statistic of 5.32 with p-value of $< .0001$.

Table 9

Test of Equality of Proportion of Success by the Two Procedures

Model	Z	p-value
Combined	5.32	<.0001
Mathematics	2.02	.0217
Calculus	1.67	.0475
Matrices	.45	.3264

We see that except for the college mathematics class the regression procedure predicted a greater proportion of success. This represents a superiority of AID3.

Cross Validation

In this section we examine the behavior of the two procedures' prediction equations on a new data set consisting of students entering in 1987. Table 10 gives the sample size, actual proportion of successes, and the proportion of successes of the new data set predicted by the developed models.

Table 10

Actual and Predicted Proportion of Success

Model	N	Actual	AID3	Regression
Combined	1733	.780	.932	.874
Mathematics	343	.831	1.000	.930
Calculus	816	.703	.763	.814
Matrices	574	.857	1.000	.956

Not unlike the development of the models, we have both the prediction equations exhibiting on overly optimistic rate of success. In comparison with Table 8, the proportions are reversed for the combined model.

Considering the accuracy of the two prediction equations Table 11 gives the percentages of times the model correctly predicted the actual performance of students who were successes (failures) and the total percentage of correct predictions. For example, in model development the combined regression model predicted success for 93.11 percent of the students that were actual success. Similarly, 36.66 percent of students that actually failed were predicted to fail. Also note the moderately low percentages of correct predictions for students who were not successful; this coincides with the overly optimistic success rates found in Table 10. As evidenced here both prediction equations maintain levels of correctly predicting performance on the new set of students similar to the levels attained in the model development.

Table 11

Model Percentage in Correctly Predicting Actual Performance

Model Development	A-C	D-F	Total
<u>Regression</u>			
Combined	93.11	36.66	78.49
Mathematics	94.44	23.66	77.17
Calculus	88.26	53.33	77.58
Matrices	97.41	16.87	81.90
<u>AID3</u>			
Combined	89.63	46.38	78.42
Mathematics	96.88	19.35	77.95
Calculus	85.91	56.89	77.04
Matrices	97.99	21.69	83.29
Validation			
<u>Regression</u>			
Combined	93.49	34.03	80.38
Mathematics	95.44	18.97	82.51
Calculus	90.07	39.26	75.00
Matrices	97.56	15.85	85.89
<u>AID3</u>			
Combined	96.45	19.11	79.40
Mathematics	100.00	0.00	83.09
Calculus	85.37	45.04	73.41
Matrices	100.00	0.00	85.71

Table 12 presents the R-square of the prediction equations on the new data set.

Table 12

R^2 of Prediction Equations on the New Data Set

Model	AID3	Regression
Combined	.2063	.3241
Mathematics	.1483	.1326
Calculus	.2691	.3256
Matrices	.2193	.2713

Comparing these R-squares with those of Table 4, we see the regression models have done better in maintaining the R-square achieved when developing the model.

A final result from the validation is the proportion who are in a category given they are predicted to be in that category. These results are shown in Table 13. For example, if a student is predicted to be at risk, then regardless of the procedure, the student is about 60 percent sure to make a D or F in the course. Conversely, about 80 percent of those predicted to be at a C or higher will make a C or higher.

Table 13

Percent Correct Prediction Given Anticipated Performance

	Predicted F	Predicted S
Regression		
Combined	59.6	83.4
Individual	59.2	82.8
AID3		
Combined	60.3	80.8
Individual	56.5	82.3

Conclusions

The use of statistical models to anticipate performance in a first mathematics course can produce useful results. While some adjustment might be made for the overly optimistic expectation, those who are predicted to have difficulty are much more likely to have difficulty than their colleagues (60 percent versus 20 percent). Furthermore, our research supports the beliefs that courses and grades in high school and standardized test scores are important in this estimate (Craney & Armstrong, 1985; Welch, Anderson & Harris, 1982).

A second conclusion is that actual ability to explain performance "shrinks" from the modeling sample to the use sample. While this is not a new concept, it seems that previous research has failed to consider this shrinkage.

In our case, the shrinkage was of sufficient magnitude to result in the simplest model, combined with dummy variables, being the best model.

One final conclusion, however, must be less positive than the first two. With respect to the 70 percent unexplained variance in performance, we defer to Merante (1983) who, when considering similar studies, concludes that we must aspire to "The development of critical variables that will allow us to consider each individual within his or her respective environment and predict his or her behavior within the specific situation."

References

Bahn, C. et al. Differential failure rates in college: Implicit educational concepts. In G. A. Drewry (Ed), Proceedings of the 7th Annual AIR Forum, Tallahassee, FL: The Association for Institutional Research, 1967.

Baird, L. L. Predicting predictability: The influence of student and institutional characteristics on the prediction of grades. Research in Higher Education, 1984, 21 (3), 261-279.

Bloom, A. M. et al. "AID"-ing academic program evaluation: The automatic interation detector as analysis tool. Paper presented at the 24th Annual AIR Forum, Fort Worth, May 1984.

Carmichael, J. R. et al. Predictors of first-year chemistry grades for black Americans. Journal of Chemical Education, April 1986, 63 (4), 333-336.

Craney, C. L. & Armstrong, R. W. Predictors of grades in general chemistry for allied health students. Journal of Chemical Education, February 1985, 62 (2), 127-129.

Darling, A. L. Canadian admissions: An update and summary of research findings. College and University, Summer 1983, 58 (4), 383-394.

Herzberg, P. A. The parameters of cross validation. Monograph 16, The Psychometric Society, June 1969, 34 (2, Part 2), 5.

Institute for Social Research. OSIRIS III (Volume 1, Release 2), Ann Arbor, MI: University of Michigan, 1973.

Jones, L. V. The influence on mathematic test scores by ethnicity and and sex, of prior achievement and high school mathematics courses. Journal for Research in Mathematics Education, May 1987, 18 (3), 180-186.

Leonard, D. E. & Niebuhr, B. R. Admissions variables as predictors of performance in basic science coursework. Occupational Therapy Journal of Research, March/April 1986, 6 (2), 105-113.

Merante, J. A. Predicting student success in college: What does the research say? NASSP Bulletin February 1983, 67 (460), 41-46.

Morgan, R. G. et al. Predicting student success in intermediate accounting I. Journal of Education for Business, November 1985, 61 (2), 80-84.

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

Rao, C. R. Linear statistical inference and its applications. New York, NY. John Wiley & Sons Inc., 1965.

Sonquist, J. A. Multivariate Model Building: The Validation of a Search Strategy. Ann Arbor, MI: Institute for Social Research, University of Michigan, 1970.

Sonquist, J. A. et al. Searching for Structure: An approach to analysis of substantial bodies of micro-data and documentation for a computer program. Ann Arbor, MI: Institute for Social Research, University of Michigan, 1973.

Welch, W. W. et al. The effects of schooling on mathematics achievement. American Educational Research Journal, Spring, 1982, 19 (1), 145-153.

Appendix 1

Coefficients and Standard Errors of Forward Selection Procedures

Model	Variable	Coefficient	Standard Error
Combined	INTERCEPT	-3.634	
	TOTALGPA	.67	.116
	CD1	-1.363	.084
	MACH	.05	.005
	MATHGPA	.237	.059
	MATHU	.188	.044
	HSIZE	.0007	.0002
	CDZ	-.291	.074
	SATVRB	-.013	.004
	SATMAT	.0185	.005
	HRANK	-.0015	.0006
	MATHAU	.0835	.033
	SEXD	-.124	.06
	TOTALU	-.040	.014
	SCIGPA	.086	.045
	RD1	.199	.112
Mathematics	INTERCEPT	-3.589	
	MACH	.057	.01
	SCIGPA	.323	.106
	MATHK	.282	.071
	SATVRB	-.0285	.008
	SATMAT	.021	.01
	TOTALGPA	.422	.198
Calculus	INTERCEPT	-6.633	
	TOTALGPA	1.165	.128
	MACH	.058	.005
	MATHGPA	.294	.077
	HSIZE	.0005	.0002
	RD1	.395	.139
	MATHU	.122	.039
	SCIAU	-.153	.007
Matrices	INTERCEPT	-3.304	
	MACH	.054	.007
	TOTALGPA	.433	.203
	MATHGPA	.346	.1165
	SEXD	-.193	.1065
	HSIZE	.0005	.0002
	SCIGPA	.141	.082

Appendix 2

Relative Frequency Distribution of Actual Grades

1986	A	B	C	D	F	Sample Size
Combined	16.28	30.36	27.45	14.34	11.56	1548
Mathematics	18.90	28.08	28.61	15.22	9.19	381
Calculus	13.18	28.13	28.13	16.03	14.54	736
Matrices	19.26	36.19	25.29	10.67	8.58	431
1987						
Combined	18.87	33.18	25.91	12.46	9.58	1733
Mathematics	21.57	37.90	23.62	10.20	6.71	343
Calculus	13.85	28.55	27.94	15.56	14.09	816
Matrices	24.39	36.93	24.39	9.41	4.88	574

Appendix 3

Table of Performance: Actual/Predicted

Model Development	F/F	F/S	S/F	S/S	Sample Size
Regression					
Combined	147	254	79	1068	1548
Mathematics	22	71	16	272	381
Calculus	120	105	60	451	736
Matrices	14	69	9	339	431
AID3					
Combined	186	215	119	1028	1548
Mathematics	18	75	9	279	381
Calculus	128	97	72	439	736
Matrices	18	65	7	341	431
Validation					
Regression					
Combined	130	252	88	1263	1733
Mathematics	11	47	13	272	343
Calculus	95	147	57	517	816
Matrices	13	69	12	480	574
AID3					
Combined	73	309	48	1303	1733
Mathematics	0	58	0	285	343
Calculus	109	133	84	490	816
Matrices	0	82	0	492	574