Predicting Academic Performance of MBA Program Applicants*

V. Srinivasan, *Stanford University, USA*
Dick R. Wittink, *Yale University, USA*
Bruce M. Zweig, *Consultant, San Francisco*

Abstract : We report on an approach for aiding and strengthening the MBA admissions process that was used at the Stanford University Graduate School of Business. Multiple regression models were used to make predictions of the academic performance of an applicant, if admitted to the MBA program. Applicants who were predicted to have a "substantial chance" of an inadequate academic performance were considered further only if a detailed reading of the application indicated exceptional circumstances or characteristics not adequately captured by the models. From the remaining applicants, selection was made by the admissions officer(s) based mainly on management potential. The present paper focuses on the prediction of academic performance. We discuss in detail the development of criterion variables, the specification of predictor variables, and the development, estimation and validation of models to predict academic performance as well as issues associated with the implementation of the approach.

Keywords : *MBA Program, Academic Performance, Multiple Regression Models*

Introduction

During the past few decades, a tremendous amount of growth has occurred globally in the demand for graduate management education. For many schools, the number of applicants far exceeds the number of places available. Such a situation presents an excellent opportunity as well as a challenge to select the most promising candidates. In many schools, admission decisions are usually made on the basis of overall evaluations of applicants by one or more admissions officers.

An impressive amount of empirical evidence has accumulated in the behavioral literature on decision making showing that actuarial models developed, for instance, through multiple regression are superior in predictive ability to clinical judgments (e.g., an admissions officer's evaluation of potential applicants) (Dawes and Corrigan, 1974). This superiority probably results from two sources. First, the model is likely to be more reliable or consistent, since identical predictions will result for two applicants with identical sets of values for the predictor variables.

*The support of former members of the administration and faculty of the Graduate School of Business, Stanford University are gratefully acknowledged. We thank Professors Robert J. Flanagan, the late Professor Thomas W. Harrell, and Professor James G. March for detailed comments on an earlier version of this paper. We also thank Professor Charles B. Weinberg, now at the University of British Columbia, for his contributions during early stages of the research reported here.

This may not be the case for the judgments of an admissions officer, since the evaluation is probably influenced by the quality of the set of applicants seen just prior to the one presently considered. Furthermore, an admissions officer's evaluations may become less reliable as a result of boredom, fatigue, and excessive workload whereas a model's predictions are not affected by such factors. The evaluation of different applicants by different admissions officers is likely to further decrease the reliability of the evaluation process.

A second reason for the model's superiority is that it is likely to be more valid since it is derived by systematically linking the actual performance of applicants to predictors of that performance. An admissions officer, on the other hand, gets only a limited amount of feedback and does not usually have an opportunity to conduct such a detailed and systematic evaluation. Furthermore, the overall desirability of applicants is usually not stated in terms of explicitly defined criteria. There is potential, therefore, to improve the validity of the admissions process by making the criterion variable(s) explicit. Given explicit criteria, predictions of applicants' scores on the criteria can be made based on the optimal weighting of a set of predictor variables obtainable from the application folder. We believe that model-based predictions aid in improving the reliability and validity of the admissions process while simultaneously facilitating a more efficient allocation of time spent in evaluating the applications.

It is not the intent of the proposed approach to replace an admissions officer's decisions by mechanized decisions. We believe strongly that there is a moral responsibility to read every application carefully. Our aim was merely to aid and strengthen the admission decision process by providing predictions on explicit criteria. In addition to model-based predictions, the admissions officer(s) will take into account the unique characteristics of the applicant, which may not have been adequately captured by the models. Recognizing that a student's education, in the broader sense of the term, is derived in part from the prior experiences of other students in the program, an admissions officer may justifiably admit a candidate based on the applicant's unusual background and/or experience, although the candidate may not be the best in terms of other explicitly defined criteria.

The Overall Approach

In this paper, we report on the approach that was used for aiding and strengthening the MBA admissions process at Stanford University Graduate School of Business through the use of actuarial models to predict academic performance in the MBA program. The models were developed using information derived from the MBA application folder. The predicted academic performance was intended to be used in conjunction with a modified version of the then admission procedure (see Figure 1).

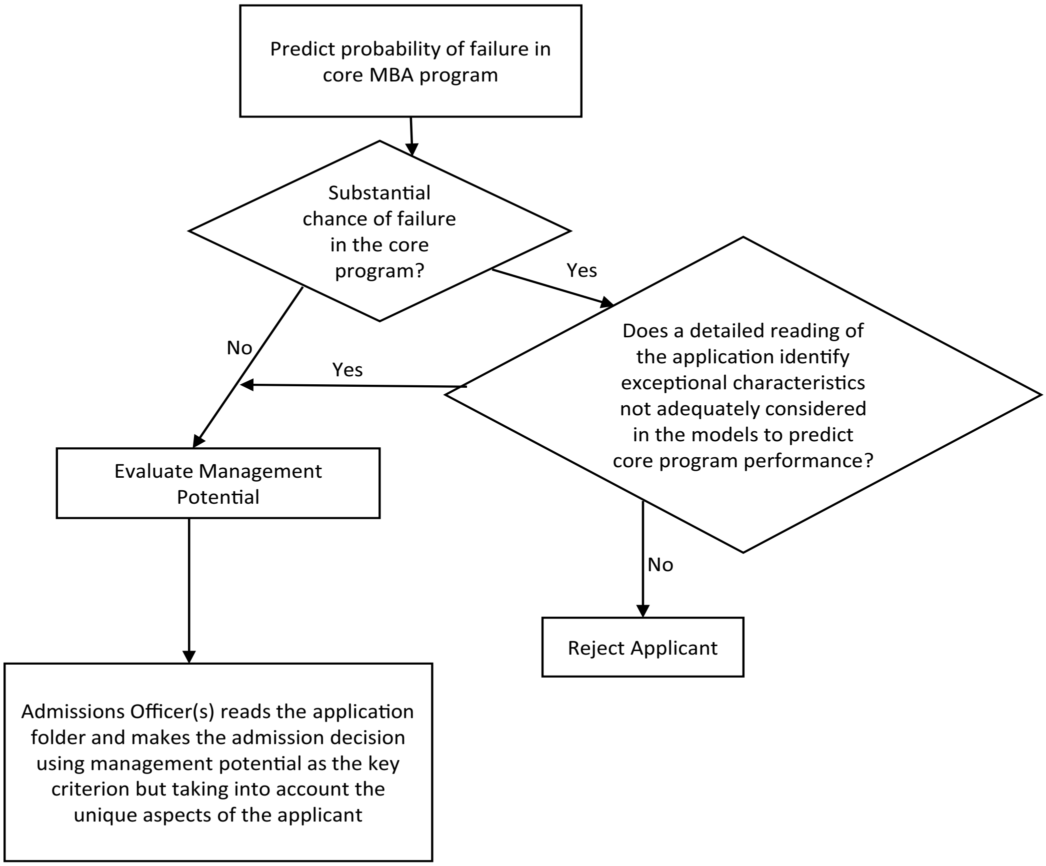


Figure 1. Modified MBA Admissions Process

The proposed procedure aids in the selection of applicants with the highest management potential among those who are academically viable. Performance in the core program, rather than the entire MBA program, is emphasized since the core represents the set of courses considered essential and hence required for every MBA candidate. Furthermore, comparison of academic performance in the entire program is less meaningful since the electives taken vary widely among students. The academic performance model was used primarily for screening out applicants who were predicted to be in academic difficulty, if admitted to the program. However, every application would be read carefully to take into account exceptional characteristics of each applicant, which may not have been adequately captured by the models (see Figure 1).

The procedure examines academic performance prior to evaluating management potential. This is based on the belief (supported by the empirical results reported in Section 7 of this paper) that an effective model for predicting academic performance can be developed from readily available quantifiable information.

Academic Performance: The Stanford Business School in its MBA core program (i.e., the set of required courses) emphasizes basic managerial skills, develops an appreciation for the increasingly complex environment for business and government, and provides an understanding of management problems in the areas of accounting, finance, marketing, and business policy. The managerial skills emphasized include the ability to think through a problem logically, the ability to use behavioral principles, economic theory, and quantitative analysis, and the ability to communicate effectively with others, all of which are essential for effective managerial decision-making and implementation. A satisfactory performance in the core program is thus considered essential for successful completion of the MBA program. If an applicant is predicted to have a substantial chance of failing the core requirements, he/she should not be admitted unless there are exceptional circumstances or characteristics which may not have been adequately captured by the models to predict academic performance. *As will be detailed later, failure in the core program is defined as obtaining a below "Pass" average in one or both of two subsets of core courses.*

Development of Academic Performance Criteria

At the time the models in this paper were developed, the core program at the Stanford MBA program consisted of thirteen courses. The skills emphasized in these courses vary considerably, although there is substantial overlap within subsets of these courses. It may be argued that a student's performance in a given core course depends on that student's skills in the more general area represented by the corresponding subset. Consequently, we studied how to group the core courses. The separation of core courses into subsets was thought to be useful, because the influence of predictor variables on academic performance may depend on the skills emphasized in a subset of courses.

Factor Analysis of Core Courses: The grades for 296 MBA students who graduated in an earlier year were used to determine the subsets of courses. The grading system in use at the Stanford Business School assigns, for each unit of completed coursework, 1.0 point if the grade is H (honors), 0.5 for a P+ (pass plus), 0.0 for a P (pass), -0.3 for a P- (pass minus), and -1.0 for a U (unsatisfactory). Students who exempted a core course were given 0.8 points as an estimate of the grade they might have received, had they taken the course for credit. (The results remain virtually the same when 0.6 is used as the estimate).1 Observations with missing grades on one or more of the courses were deleted from the analysis ("list-wise deletion"). The matrix of (Pearson) correlation coefficients between grades in core courses is shown in Table1.

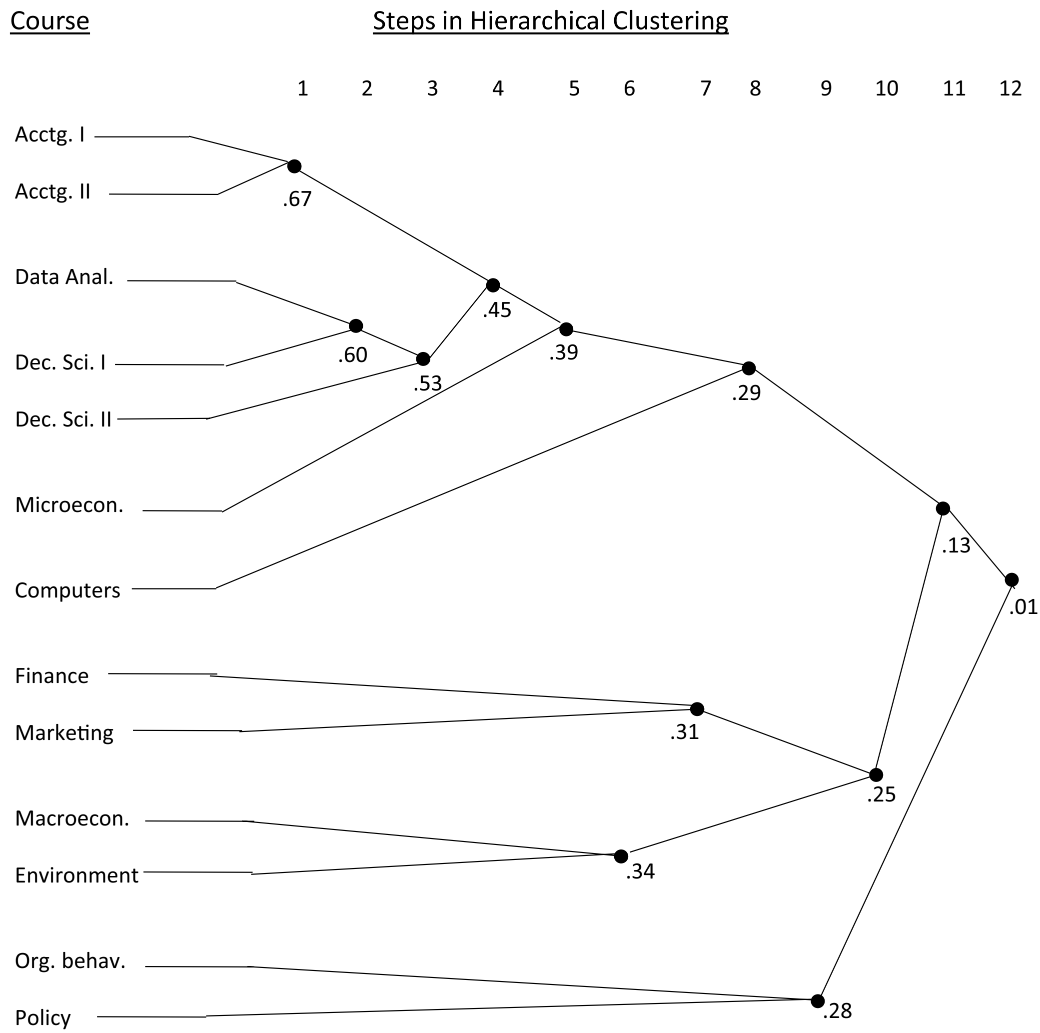
Table 1. Correlations Between Grades in Core Courses (n = 296)													
Courses	**Dec. Sci. I**	**Dec. Sci. II**	**Data Anal.**	**Acctg. I**	**Acctg. II**	**Microecon.**	**Computers**	**Finance**	**Marketing**	**Macroecon**	**Org. Behav.**	**Policy**	**Environment**
Dec. Sci. I	1.00												
Dec. Sci. II	.57	1.00											
Data Anal.	.60	.53	1.00										
Acctg. I	.61	.45	.52	1.00									
Acctg. II	.58	.49	.52	.67	1.00								
Microecon.	.55	.39	.52	.45	.48	1.00							
Computers	.32	.36	.39	.29	.30	.29	1.00						
Finance	.29	.27	.32	.39	.41	.24	.20	1.00					
Marketing	.17	.13	.22	.21	.26	.23	.22	.31	1.00				
Macroecon	.27	.21	.31	.32	.27	.32	.21	.33	.28	1.00			
Org. Behav.	.10	.01	.06	.08	.04	.13	.19	.13	.24	.16	1.00		
Policy	.09	.06	.15	.06	.09	.09	.17	.16	.11	.15	.28	1.00	
Environment	.21	.14	.28	.23	.22	.22	.19	.28	.25	.34	.25	.23	1

The highest correlation is obtained for the two accounting courses (r = 0.67), and the lowest between the course in organizational behavior and the second course in decision sciences (r = 0.01). Based on the common factor analysis model (with estimated communalities) applied to this correlation matrix (Dixon and Brown, 1979), we obtained three factors with eigen value greater than one. Together these factors account for approximately 56 percent of the total variance. The factor loadings obtained after an orthogonal (Varimax) rotation are displayed in Table 2.

Table 2. Factor Analysis of Grades in Core Courses: Three Factor Solution (n = 296)			
	Factor Loadings after Orthogonal Rotation (Varimax)		
Course	Factor I	Factor II	Factor III
Dec. Sci. I	**0.825**	0.130	0.062
Dec. Sci. II	**0.773**	0.005	0.045
Data Anal.	**0.760**	0.196	0.132
Acctg. I	**0.719**	0.340	-0.073
Acctg. II	**0.725**	0.347	-0.082
Microecon.	**0.657**	0.226	0.103
Computers	**0.499**	0.024	0.451
Finance	0.275	**0.674**	-0.011
Marketing	0.077	**0.668**	0.148
Macroecon	0.220	**0.648**	0.104
Environment	0.118	**0.555**	0.356
Org. Behav.	-0.038	0.231	**0.715**
Policy	0.055	0.077	**0.739**

Note: Bold face denotes the factor with the highest loading for that course

The results of factor analysis were essentially replicated by a hierarchical clustering (using the complete linkage method) of the thirteen courses (Dixon and Brown, 1979) as displayed in Figure 2.



Note: Entries represent minimum correlation in the subset formed at the step.

Figure 2. Hierarchical Clustering of Grades in Core Courses (n = 296)

An examination of Table 2 reveals that only two courses load highly on the third factor. To achieve a good degree of reliability of measurement for the factor scores, we used the two-factor solution as a basis for grouping the courses (see Table 3).

Table 3. Factor Analysis of Grades in Core Courses: Two-Factor Solution (n=296)

	Factor Loadings after Orthogonal Rotation (Varimax)	
Course	**Factor I**	**Factor II**
Dec. Sci. I	**0.820**	0.082
Dec. Sci. II	**0.745**	-0.011
Data Anal.	**0.763**	0.182
Acctg. I	**0.780**	0.132
Acctg. II	**0.787**	0.129
Microecon.	**0.674**	0.188
Computers	**0.435**	0.316
Finance	0.414	**0.430**
Marketing	0.203	**0.555**
Macroecon	0.342	**0.501**
Org. Behav.	-0.071	**0.684**
Policy	-0.017	**0.591**
Environment	0.193	**0.629**

We label the group of courses loading most heavily on the first factor as the quantitative set of courses (QUANT), with the remaining core courses labeled as managerial (MGMT). The results have a substantial amount of face validity. It should, however, be noted that for some courses the loadings on the two factors do not differ greatly. For example for Finance, the loading on the second factor barely exceeds the loading on the first factor.

Academic Performance Criteria. For each student, the average grades QUANT and MGMT were computed by taking the weighted arithmetic mean of the grades obtained for all courses in that subset, with the number of credit units for the courses, serving as the weights.2 This procedure, in contrast to the more complicated procedure of computing the factor scores for QUANT and MGMT, was used for simplicity in interpretation. The two procedures, however, can be expected to yield similar results [Srinivasan, 1977].

In addition to QUANT and MGMT, which correspond to the core MBA program, a third criterion variable relating to academic performance was constructed by computing the weighted average course grade across all elective courses (ELECT) taken at the Graduate School of Business. Comparability of ELECT scores across students is, of course, limited, because of the wide variability in the course contents of elective courses. Nevertheless, this variable is informative about academic performance in subjects of the students' own choosing, and is expected to be related to management potential (Marshall 1964, Williams and Harrell, 1964). However, we reemphasize that only the performance in the core program (viz., MGMT and QUANT) are relevant to screening out those who are predicted to be in academic difficulty (see Figure 1).

Reliability of MGMT and QUANT: Two graduating classes were chosen as the sample for the development of a model to predict academic performance. For these two graduating classes a total of 629 students enrolled in the MBA program. To determine the reliability of the QUANT and MGMT variables, alpha coefficients were computed (Cronbach 1970, p. 161). For this computation, students with no grade for one or more courses included in a criterion variable were eliminated. For QUANT, the reliability score obtained was 0.854 (n = 110), whereas for MGMT this score was 0.574 (n = 333).3 The lower reliability for MGMT can be explained by two considerations. First, the variability in grades for MGMT courses tends to be smaller compared to QUANT courses. (Reliability is inversely related to variability). Second, MGMT may be thought of as a combination of two constructs. As discussed earlier, the factor analysis of core course grades suggested that two of the courses on which MGMT is based, should, in theory, be considered separately as a third factor (see Table 2). The alpha coefficients can be viewed as approximate upper limits for the R2 values in models to predict MGMT and QUANT.[4]

Specification of Predictor Variables

To determine the variables that could serve as predictors in models designed to predict academic performance, we examined the following sources: (i) previously published studies in this area, Deckro and Woundenberg (1977), Gayle and Jones (1973), Harrell (1972), Harrell and Harrell (1973), Harrell, Harrell, McIntyre and Weinberg (1977), Livingston (1971), Marshall (1964), Page and West (1969), Pfeffer (1977), Pitcher (1971), Srinivasan and Weinstein (1973), Weinstein and Srinivasan (1974), Williams and Harrell (1964), (ii) a list of rating scales used by the admissions officers at the Stanford Business School (Lieberman, 1977), and (iii) the application form, for an exhaustive list of variables.

Potential predictor variables based on preadmission information were categorized, according to ease of accessibility, as follows:

(i) Directly Available or Codable (AC) variables: Information is available in numerical form, for example, date of birth, and can be entered directly into a database, or can be coded or calculated in a straightforward manner, for example, the number of months of full-time work experience codable from the employment history provided by the applicant. See Table 4 for a list of the AC predictor variables.

Table 4. List of AC (Available or Codable) Variables [a]

ADV	degree received as a graduate student (= 1 if received an advanced degree; = 0 otherwise)
AGE	age of student at time of enrollment in the MBA program (in months)
AGESQ	the squared value for AGE
CES	candidate excellence by school index [b]
CLASS	year of graduation from MBA program (0= earlier year; 1 = later year)
DUMJi	undergraduate major area (DUMJ i = 0 for all i, except
except:	DUMJ1 = 1 if math., stat., or computer sci.
	DUMJ2 = 1 if behavioral sciences
	DUMJ3 = 1 if other sciences
	DUMJ4 = 1 if engineering
	DUMJ5 = if business administration
	DUMJ6 = 1 if accounting
	DUMJ7 = 1 if economics
	DUMJ8 = 1 if liberal arts
	DUMJ9 = 1 if political science)
EXPBUS	full-time business experience in months
EXPMIL	full-time military experience in months
EXPOTH	full-time other experience in months
EXPPT	part-time work experience in months
FOREIGN	indicator for country of citizenship (= 1 if foreign; = 0 if US)
GMATMS	number of times student has taken graduate management admission test (GMAT) (= 0 if taken once; = 1 if taken more than once)
GMATQ	graduate management admission test score (quantitative) [c]
GMATV	graduate management admission test score (verbal) [c]
GPA1	undergraduate grade point average as a freshman
GPA2	undergraduate grade point average as a sophomore [d]
GPA3	undergraduate grade point average as a junior [d]
GPA4	undergraduate grade point average as a senior [d]
GRADWK	number of months of study in a graduate school
GRADYR	year of graduation from college
MARDUM	marital status (= 1 if married; = 0 otherwise)
MAXSAL	maximum monthly salary ($) at time of application (= 0 if no prior experience, or no salary data available) (cf. NOSAL)
MILDUM	indicator for military experience (= 1 if veteran; = 0 otherwise)
NOSAL	indicator for non-availability of salary data (= 1 if no salary data, although the applicant has had prior work experience, = 0 otherwise)
SUMEMP	summer work experience while in college in months
TOEFL	test of English as a foreign language (total) score

[a] The variables corresponding to the sources of financing of undergraduate education (proportions of financing by parents, scholarship, loans, employment, and other) are excluded since these data were incomplete for a large proportion of the observations.

[b] This index of quality of the undergraduate school was published by the Educational Testing Service. It is computed as the average Graduate Management Admission Test Score (total) of all test takers from the applicant's undergraduate school

[c] The most recent score was used for applicants who had taken the test more than once.

[d] On a four-point scale with D = 1, C = 2, B = 3, and A = 4.

Rating scale (R) Variables: These are qualitative variables for which systematic rating procedures were developed to obtain quantitative assessments; for example, in evaluating an applicant's leadership activity as an undergraduate. A detailed examination of 30 application folders was useful in defining the R-variables and in providing detailed instructions to raters (for an illustration, see the instructions for coding "level of work experience" in Appendix B). Inter-rater reliabilities, computed using the 30 applicants, were used as diagnostics to improve the definitions and to clarify the instructions. In addition, composites were created as weighted sums of the rated variables. For instance, level of experience was coded for each job held by an applicant on a 1-9 ordinal scale (Appendix B). To summarize the experience of the applicant, while simultaneously taking into account the level of experience, a weighted total experience (EXPSUM) variable was defined by summing over the different jobs the number of months of experience, multiplied by a weight reflecting the level of experience for each job. The weights were arrived at by averaging the subjective judgments of the members of the research team. See Table 5 for a listing of the R- variables.

Table 5. R-(Rating Scale) Predictor Variables [a]

ACHIEVE	relative achievements [b]
COMMU	communication ability [b]
CREATE	four indicator variables for creativity demonstrated by evidence of creating artistic work, publication, invention, or business venture
DEBATE	number of years of participation in debate club
EXPGRO	career growth prior to entry = (10 - EXPHI) / (total number of months of
EXPHI	highest level of work experience [a] prior to entry
EXPSUM	sum of work experience in months, weighted by the level of experience [a]
EXPTOT	sum of work experience in months, unweight
GOAL	ability of candidate to define career goals
INITIA	demonstrated initiative and drive
INSIGHT	insightfulness into own strengths and weaknesses [b]
LEADHI	number of years of leadership activity at high level while at college (for example, college elective leadership)
LEADLO	number of years of leadership activity at low level while at college (for example, officer of residence hall)
LEADSU	sum of LEADHI and LEADLO, weighted by 1.0 and 0.4
OTHACT	number of years of other activities (for example, social and service clubs)
PRESENT	quality of presentation of case for admission
RELATE	relationships with other people [b]
S	socioeconomic status of applicant's family
SPORTHI	number of years of sports participation at high level while at college (for example, varsity sports)
SPORTLO	number of years of sports participation at low level while at college (for example,
SPORTME	number of years of sports participation at medium level while at college (for example, intramural sports)
SPORTSU	sum of SPORTHI, SPORTMED and SPORTLO, weighted by 1.00, 0.38 and 0.19
UNDSTD	candidate's demonstrated understanding of the MBA program
WHYMB	clarity of candidate's statement of reason for wanting to pursue an MBA program

[a] See Appendix B for a listing of the levels of work experience and associated weights. For brevity in presentation, the definitions of the remaining R- variables are not provided in this paper.

[b] This variable, derived from recommendation letters, is defined as the weighted average of the scores provided by at most four evaluators of the candidate; the weights are determined by nature and frequency of contact: 1.0 if daily or weekly, at work or at school; 0.5 if monthly, at work or at school; 0 otherwise.

Reliability of the R-variables: For the rating scale variables to be useful as predictors, we need both intra- and inter-rater reliability. Based on additional samples of thirty applicants each, (Pearson) correlation coefficients were computed for most of the R-variables. If there is more than one measure of the same variable (e.g., experience), only the measure considered most relevant was used in the computation of reliabilities.

Table 6. Reliability Scores for the R-Variables

Variable	Intra-rater[a] reliability (n = 30)	Inter-rater[a] reliability (n = 30)
ACHIEVE	.724	.413
COMMUN	.633	.749
CREATE	B	b
DEBATE	.149	b
EXPSUM	.999	.886
GOAL	-.049	.343
INITIA	.505	.452
INSIGHT	B	b
LEADSUM	.967	.782
OTHACT	.987	.826
PRESENT	.726	.485
RELATE	.306	.647
SES	.923	.604
SPORTSUM	.997	.883
UNDSTD	B	.639
WHYMBA	-.128	.248

The results reported in Table 6, suggest that the reliability is high for some variables, but unacceptably low for some other variables. In particular, GOAL and WHYMBA suffer from a lack of consistency. Both variables are obtained from information scattered through the application folder. It may be expected therefore that these variables posed serious difficulties for the raters. DEBATE also fares poorly, but there is a very large proportion of zero values for this variable. For this variable, minor differences between the raters or ratings can have an enormous influence on the reliability score. For a few variables, no reliability score could be computed due to the lack of variation across applicants in the sample. In general, variables which have been indicated in the previous literature to be the most relevant had high enough reliability scores to be of potential value as predictor

[a] The reliability scores in the two columns were computed for two different sets of applicants. Intra-rater reliabilities were obtained for one judge (A) who made the independent ratings separated by six months. Two different judges (B and C) were used to obtain inter-rater reliabilities.
[b] A blank denotes that the reliability score could not be computed since there was no variation on this variable.

variables. Further improvements in the definitions and instructions may be useful in improving the reliabilities. The application form may also have to be modified to elicit some of the information more readily.

Developing Models to Predict Academic Performance

As remarked in the earlier section, we hypothesized that given a model to predict academic performance using the AC-variables, the incremental predictive power of the R-variables would be slight. Thus, the initial model development was restricted to the AC-variables listed in Table 4.

Estimation and Validation Samples: The total sample consisted of students from two graduating classes (n = 629). Since model development was expected to consist of a large number of steps, it seemed desirable to set aside a fraction of these observations as a holdout sample to examine the model's predictive validity. For this purpose, twenty percent of the observations were randomly selected to constitute the validation sample. The remaining 80% of the observations constitute the estimation sample.

Treatment of Missing Data: The following strategy was adopted for handling missing data: an observation would be excluded from analysis, if the data on the dependent variable or on one or more of the predictor variables in that model were missing ("list-wise deletion"). Consequently, the number of observations will vary, in general, from one model to another. However, for the undergraduate GPA variables for which there was a relatively small proportion of missing data, a different strategy was used. To illustrate, consider individuals for whom only GPA4 was missing. An equation was developed, predicting GPA4 based on the values for GPA1, GPA2, and GPA3, using all observations for which GPA1 through GPA4 were available. This equation was then used to predict the value of GPA4 for individuals for whom only GPA4 was missing.

Model Development: The models should allow for possible subgroup (e.g., gender, race) differences in the relationship between academic performance criteria and the predictor variables. Since the number of observations for minorities, women and foreigners were small, we chose to first develop the models on all white male U.S. citizens in the estimation sample (i.e., after excluding 20% of the observations set aside as the holdout sample for validation purposes).

For each of the three criterion variables (MGMT, QUANT and ELECT), the initial multiple regression model specification included the variables listed in Table 4. The models were updated separately step by step by deleting variables with insufficient explanatory power (as indicated by low values for the t- statistic). When there was more than one such variable to be deleted, the deletion was done in the reverse order of importance of variables as indicated by the previous literature cited earlier. Also, GRADYR was eliminated due to the extreme amount of overlap (collinearity) with AGE. Predictor variables were added to the model, if necessary, so that the final models had the property that the incremental F of every excluded predictor is statistically insignificant. To the extent possible, models were made more parsimonious by grouping variables, which had nearly equal regression coefficients. For instance, if the regression coefficients for GPA1, GPA3, and GPA4 were nearly equal in a particular model, these variables were replaced by GPA134, the average of GPA1, GPA3, and GPA4, for that model.[5]

To minimize the number of separate indicator (dummy) variables for undergraduate major, a scheme was developed (see Figure 3) to group majors based on their similarities with respect to preparation for the MBA program.

Indicator Variable Number (Table 4)	**Undergraduate Subject Major**	**Grouping**
1	Math., Stat., or Computer Science	F = 3.80[a] (ELECT)
3	Other Sciences	
4	Engineering	
5	Business	
6	Accounting	
7	Economics	
2	Behavioral Sciences	F = 3.09[a] (ELECT)
8	Liberal Arts	
9	Political Science	
Legend: }	Indicates the categories to be grouped. The absence of an F-value indicates that the null hypothesis` of equal regression coefficients for the majors belonging to a group could not be rejected at any significance level since the F-value was below 1.	

Figure 3. Grouping Undergrad. Majors for MGMT, QUANT and ELECT Models (n = 480)

[a] These values are not statistically significant at the 5% level for the ELECT model. (The F-values for MGMT and QUANT are below 1.) A conservative approach dictates against the grouping these categories for ELECT model.

The appropriateness of this scheme for reducing the number of subgroups from nine to five was statistically tested for the MGMT, QUANT, and ELECT models separately. Based on an F-statistic, the null hypothesis of equal parameter values for the majors belonging to a subgroup as per the scheme of Figure 4 could not be rejected, except possibly for two cases involving ELECT as the criterion variable. Thus, for MGMT and QUANT the number of subgroups was reduced to five, whereas for ELECT seven subgroups remained. Further grouping of undergraduate majors was done in the analysis if the regression coefficients for two or more undergraduate majors were nearly equal. For instance, because the coefficients for DUMJ2 (behavioral sciences), DUMJ8 (liberal arts) and DUMJ9 (political science) were nearly the same, the three variables were replaced by the indicator variable DUMJ289 which takes the value one for behavioral science, liberal arts, or political science undergraduate majors, and zero otherwise.

Examination of Subgroup Differences: After various iterations to develop a model for each of the criterion variables, based on the largest subgroup of white male U.S. applicants only, tests were developed to determine whether the relationships depend on gender, marital status, and minority status. The results suggested that there were no statistically significant differences in the relationship between females and males, or between married graduates and others. There was also no difference in the effects of predictor variables when whites were compared with minority subgroups of the U.S. population.

Having developed a model for all U.S. applicants in the estimation sample, observations representing foreign applicants were added. However, for most foreigners, neither the college excellence score (CES) (see Table 4 for a definition) nor undergraduate GPA (on a comparable scale) could be obtained. Consequently, it was necessary to add an indicator variable FOREIGN, which could compensate on the average for the assigned values of zero to foreigners for CES and the GPA variables. We also examined differences in the models across foreign nations using per capita GNP and illiteracy rate as additional predictors. However, these additional predictors did not produce statistically significant effects.

Empirical Results

Estimated Multiple Regression Models: Beta weights (standardized regression coefficients) and other statistics for the final models, using AC-variables only, are shown in Table 7.

Predictor Variable[b]	Criterion Variable[a]		
	MGMT	QUANT	ELECT
AGE	0.20 (4.70)	-	1.59 (3.10)
AGESQ	-	-	-1.37 (-2.67)
CES	0.73 (3.89)	0.16 (1.12)	0.60 (3.16)
CLASS	-	-0.05 (-1.89)	-
DUMJ2456[b]	-	-	-0.09 (-2.32)
DUMJ289[b]	-	-0.07 (2.17)	-
DUMJ7	0.14 (3.60)	-	-
EXPBUS	-	0.12 (3.93)	-
FOREIGN	1.31 (5.65)	0.89 (5.06)	0.96 (3.55)
FGMATV[b]	-	-	0.32 (1.90)
GMATV	0.30 (6.93)	0.13 (3.40)	0.09 (1.69)
GMATQ	-	0.54 (15.09)	0.13 (2.95)
GMATMS	-	-0.09 (-2.89)	-
GPA134[b]	0.62 (5.17)	0.70 (7.81)	-
GPA4	-	-	0.78 (6.85)
Adjusted R^2	0.183	0.532	0.173
Standard Error of Estimate[c]	0.198	0.229	0.177
Sample Size	596	587	602

Table 7. Beta Weights and Other Statistics for Final Models (Total Sample)

Although the models were developed (i.e., predictor variables decided) on the basis of the estimation sample (80% of the total data), Table 7 presents the results based on all the data. In general, the beta weights indicate the relative importance or relative explanatory power of predictor variables. However, the very high correlation between AGE and AGESQ may make such interpretations inaccurate in the ELECT model, when the age variables are compared to other predictors.

[a] Numbers in the main table are beta weights, and in parentheses the t-ratios. Since the models were developed based on the estimation subsample of the total sample, the t-ratios should be interpreted only as crude indications of the statistical significance. A blank (-) denotes that the predictor variable does not appear in the model for the corresponding criterion variable.

[b] See Table 4 for a definition of these variables except as noted below:
DUMJ2456 = 1 if the undergraduate major is behavioral sciences, engineering, business administration, or accounting, and zero otherwise
DUMJ289 = 1 if the undergraduate major is behavioral sciences, liberal arts, or political sciences, and zero otherwise
FGMATV = FOREIGN * GMATV; i.e., FGMATV = GMATV for foreigners, and zero otherwise
GPA134 = average GPA for freshman, junior and senior years

[c] The values for MGMT, QUANT, and ELECT range from -1 to 1.

As expected, the variable FOREIGN, which compensates for the assignment of CES = 0 and GPA= 0, for foreign applicants, is a prominent predictor variable in all three equations. The induced correlation between FOREIGN and CES and FOREIGN and GPA makes the use of beta weights (to infer relative importance) inaccurate when FOREIGN, CES or GPA are compared with other predictors. A comparison of the present results to the results for U.S. applicants only (thus excluding FOREIGN from the models), indicates, however, that the qualitative conclusions below continue to hold within the U.S. subpopulation.

For the MGMT model, undergraduate grade point average, as measured by GPA134 (average GPA for freshman, junior and senior years), verbal score on the graduate management admission test (GMATV), the college excellence score (CES), and age are important predictors.6 The positive coefficient for DUMJ7 indicates that economics undergraduates' performance on MGMT is better, on the average, than other undergraduates, holding the remaining predictor variables constant.

For QUANT, undergraduate grade point average (GPA134) and the quantitative score on the graduate management admissions test (GMATQ) are the more important variables. The verbal score on the graduate management admissions test (GMATV) and, amount of business experience (EXPBUS) also have explanatory power in this model. In QUANT, undergraduates who major in behavioral science, liberal arts, or political science perform poorer, on the average, in comparison to other undergraduates, holding the remaining predictor variables constant. However, since the models are compensatory, a liberal arts major with a higher GPA (and/or GMATQ) would be predicted to have a higher QUANT than an economics major with a lower GPA (and/or GMATQ). The negative coefficient for GMATMS indicates that it is appropriate to adjust the prediction downward if an applicant has taken the GMAT more than once. (Recall that the most recent scores were used if the applicant had taken the GMAT more than once.) The negative coefficient for CLASS indicates that after adjusting for other predictors, the mean QUANT score for one graduating class is smaller than for another, holding all other predictors equal.

The adjusted R^2 for the QUANT model is considerably higher than that for MGMT. However, it should be noted that the amount of variation in QUANT is also greater than that for MGMT. In fact, the estimated standard error (standard deviation of residuals) for the QUANT model exceeds the corresponding value for MGMT despite the higher adjusted R2. The model for ELECT has both AGE and AGESQ in the equation. This indicates that the age of an applicant tends to have a positive effect on academic performance in elective courses, but at a diminishing rate. (For ELECT the optimum age, i.e. the age at which the total effect of AGE and AGESQ is at a maximum, holding the other predictors constant, is 34 years).7 The large (absolute) values of the beta weights for AGE and AGESQ results from their considerable collinearity. Undergraduate grade point average, measured now by only the average grade in the senior year, and the college excellence score (CES) are important predictor variables. Both verbal and quantitative GMAT scores are included in the equation. Furthermore, an interaction effect is incorporated by the creation of FGMATV (= FOREIGN*GMATV). Thus, for a foreigner, the effect of GMATV is greater and is obtained by adding the slope coefficients for GMATV and FGMATV. (For U.S. applicants the corresponding effect is given by the slope coefficient for GMATV only). In terms of academic performance in elective courses, students with an undergraduate major in behavioral science, engineering, business, or accounting score not as well, on the average, as other undergraduates, holding the effects of other predictor variables constant. The adjusted R2 for ELECT is slightly smaller compared to MGMT, although the fit of the ELECT model is slightly better than the MGMT model in an absolute sense (as measured by the standard error of estimate).

Incorporating the R-variables: Given the AC-variables selected for each of the three models, it is of interest to determine how much the explanatory power of each model can be increased with the addition of the R-variables. As initially expected, the adjusted R2 values, after including those R-variables with absolute t-ratios in excess of two, are not much higher than the figures reported in Table 7. (The increases in adjusted R2 for QUANT, ELECT, and MGMT are 0.01, 0.03 and 0.04, respectively.) In each model, however, EXPSUM and PRESENT (see Table 5 for definitions) added significantly to the proportion of variation in the criterion variables explained by the equation. In the model for MGMT, an additional predictor, CREATE (see Table 5) had a significant effect.8 By far the most promising variable for MGMT or QUANT is the weighted experience variable, EXPSUM. However, in considering the usefulness of collecting information on rating scales for predicting academic performance we have to weigh the cost of collecting this information for all applicants against the expected benefit in terms of greater predictive accuracy. We examine this issue further by comparing the predictive validity, using holdout samples, of the models with and without the R-variables, in a later section of this paper.

Seemingly Unrelated Regressions: If the value for the error term in one model, for a given observation (student) tends to be correlated with the value for the error in another model for the same observation, there is an opportunity to obtain more efficient (albeit asymptotically) estimates of the regression parameters by estimating the equations jointly (Zellner, 1962).

Variable	MGMT	QUANT	ELECT
MGMT	1.000		
QUANT	0.389	1.000	
ELECT	0.593	0.443	1.000

Table 8. Correlations of Residuals for Three Regression Equations (n = 383)[a]

The correlations, shown in Table 8, vary from 0.389 (between MGMT and QUANT) to 0.593 (between MGMT and ELECT). However, the results obtained by simultaneously estimating the parameters of the three equations (Hall and Hall, 1980) did not differ systematically from the results obtained by estimating each equation separately, except that the coefficients tended to be slightly closer to zero.

Testing the Assumptions of Multiple Regression: The procedure used for estimating the parameters in a given regression equation provides best linear unbiased estimates if certain assumptions are satisfied. One assumption is that the variance of the error term in the equation is constant across observations (homoscedasticity). It may be argued that this assumption is violated because the values for the criterion variables are based on a varying number of courses. For MGMT and QUANT, students may have exempted one or more core courses. If a core course is exempted, the student has to obtain the required number of credits by taking additional elective courses. It is reasonable to expect that the uncertainty (error) associated with a given grade point average is a function of the number of courses on which the average grade is based. However, a procedure which weighted each criterion variable score by the number of courses (a weighted regression version of generalized least squares) left the results essentially unchanged.

For each criterion variable, plots of the residuals against the predicted values of the criterion variable did not reveal any evidence of nonlinearity or heteroscedasticity. To examine the assumption of normality, the Kolmogorov-Smirnov test (Siegel, 1956) was used. For each criterion, the null hypothesis of normality of the residuals could not be rejected ($p > 0.40$ for each of the three equations).

When performing a regression analysis on a set of data, one should also be concerned about the possibility that some extreme observations have a great influence on the regression coefficients obtained. An examination of the residuals did not reveal more outliers than what would be expected by chance alone. Related to this issue is the fact that the predictions for some individuals may be quite poor. We have attempted to determine whether there is anything systematic about large differences between actual and predicted values for a criterion variable. In a later

[a] The correlations were computed before including observations on foreign students in the sample.

section we discuss some possible reasons for obtaining sizable differences, using observations from the validation sample.

Curtailment: The analysis so far has been carried out on applicants who were admitted and who enrolled in the MBA program. This group is selected on the basis of certain criteria and is, therefore, likely to be systematically different from other applicants. Therefore, if we consider only the enrollees, we may not capture the variation in predictor variables in the entire applicant pool, which is the population to which the models are to be applied. Such curtailment problems (Lord and Novick, 1968) have been addressed previously by Srinivasan and Weinstein (1973) in a context similar to the present one. Basically, restricted variation in predictor variables tends to reduce the values for the t-ratios and beta weights. Corrections for curtailment can be made by assuming that the regression coefficients relating the criterion variables to the predictor variables are the same for enrolled students and other applicants.

To examine the incidence of curtailment, we collected data on the predictor variables for a subset of the remaining applicants (i.e., those who were rejected and those who were accepted but did not enroll). Define $Q = S_A/S_E$, where SA is the estimated standard deviation of a predictor variable in the entire applicant pool[9] and S_E is the standard deviation of the same predictor variable for all enrollees. If Q is substantially above 1, curtailment is indicated; on the other hand, if Q is substantially below 1, values for the t-statistics and beta weights based on the enrollees may be overstated.

For each potential predictor variable (Tables 4 and 5) a Q-value was computed. If the Q-value was substantially above 1 ($Q > 1.2$) and the predictor was not included in the model, the t-value and the beta weight were examined by including the predictor variable in the final model (Table 7). In all cases these t-values and beta weights were small, so that even with an adjustment for curtailment these predictor variables would not be statistically significant. Similarly, for predictor variables included in a given model, the t-values would continue to be statistically significant even if an adjustment was made for Q-values substantially below 1 ($Q < 0.8$) as was the case for a few predictor variables. Consequently, no curtailment correction appears to be necessary.

Goodness of fit corrected for reliability: As explained earlier, we obtained reliability coefficients for two of the criterion variables, MGMT and QUANT. These reliability coefficients can be used to correct the adjusted R2 for the final models presented in Table 7. Specifically, the reliability scores serve as upper limits for R2 since no predictors can explain the error components of the criteria.10 The correction in adjusted R2 can be made by partitioning the variance (A) in the criterion variable into three components: variance explained by the model (B), model error variance

(C), and measurement error variance (D). Adjusted R2 corrected for reliability is defined by the ratio B/(A-D). The reliability score (Cronbach alpha) α is an estimate of (A-D)/A, and adjusted R^2 obtained for the multiple regression equals $\overline{R}^2$= B/A. Thus, adjusted R2 corrected for reliability is given by

$B/(A-D) = (B/A) \div [(A-D)/A] = \overline{R}^2/\alpha$,

or 0.319 for MGMT and 0.642 for QUANT (see Table 9).

	Criterion Variable	
	MGMT	QUANT
Adjusted R^2 (from Table 7)	0.183	0.532
Cronbach α (estimated reliability) from the earlier result under "Reliability of MGMT & QUANT"	0.574	0.854
Adjusted R^2 corrected for reliability = Adjusted R^2/Cronbach α	0.319	0.623
$\frac{\textit{Adjusted } R^2}{\textit{Cronbach } \alpha}$		

Table 9. Adjusted R2 Corrected for Reliability

These corrected values suggest that there is considerable room for improving the explanatory power of the models, particularly for the model to predict MGMT. The above correction for reliability was not done for ELECT, since its Cronbach α could not be determined (cf. endnote 4).

Model Validation

Cross-Validation: As indicated earlier, 20% of the observations representing the graduates in the two graduating classes were set aside to examine the predictive validity of the academic performance models developed from the estimation sample (i.e., the remaining 80% of the total sample). Results were computed separately for the models with AC- variables only (see Table 7 for the list of predictor variables included in the models) and the models with selected R-variables added. The R- variables added are EXPSUM and PRESENT for all three equations, and, in addition, CREATE for the MGMT equation only.

The cross-validation results were computed separately for U.S. white males, U.S. females, U.S. minorities, and foreigners, as well as for all individuals together. For each criterion variable we show in Table 10, SY2, the variance in the scores across individuals in the validation sample, MSE, the mean squared error of prediction, r^2, the squared (Pearson) correlation coefficient, and n, the sample size. Note that predictions are made only for individuals with no missing data on the predictor variables for a given model. The sample sizes fluctuate for this reason across the three models.

	U.S. white males	U.S. females	U.S. minorities	Foreign	All
Criterion Variable: MGMT					
$S_Y^2(x100)$	5.76	4.58	6.00	4.45	5.57
MSE (x100)	5.06[a] (5.02)[b]	3.13 (3.69)	3.31 (2.92)	4.20 (4.54)	4.54 (4.58)
r^2	0.12 (0.15)	0.32 (0.08)	0.75 (0.81)	0.10 (0.08)	0.18 (0.17)
n	56[c] (54)	23[c] (22)	17[c] (15)	19 (18)	115[c] (110)
Criterion Variable: QUANT					
$S_Y^2(x100)$	11.09	10.05	12.96	8.35	11.09
MSE (x100)	5.52 (5.57)	5.57 (6.00)	4.67 (4.45)	9.80 (9.30)	6.05 (6.05)
r^2	0.50 (0.44)	0.36 (0.32)	0.68 (0.70)	0.01 (0.03)	0.45 (0.42)
n	56 (54)	23 (22)	17 (15)	18 (17)	114 (109)
Criterion Variable: ELECT					
$S_Y^2(x100)$	3.72	3.65	4.37	3.76	3.76
MSE (x100)	3.13 (3.03)	3.53 (3.84)	3.42 (3.20)	4.16 (4.16)	3.39 (3.35)
r^2	0.16 (0.18)	0.07 (0.01)	0.25 (0.22)	0.04 (0.06)	0.10 (0.09)
n	57 (55)	24 (23)	17 (15)	19 (18)	117 (112)

Table 10. Predictive Validity of Academic Performance Models (Hold-out Sample)

As indicated earlier, the ultimate interest is in predicting the probability of failure in the core program. This criterion is stated in absolute terms, as opposed to a criterion which measures the performance of an applicant relative to other applicants. Thus, the mean squared error is a more appropriate measure of predictive validity than the squared correlation coefficient. Nevertheless, based on either measure as shown in Table 10, there is a tendency for the minorities to be more predictable than the average, except that the mean squared error for minorities in the ELECT model is about the same as average.

An examination of Table 10 reveals that for foreigners the mean squared error is larger than average, except when MGMT is the criterion variable. In fact, for QUANT and ELECT the mean squared error for foreigners exceeds the variance of the criterion variable. This is probably due to the fact that no scores for CES or for the GPA variables are available for foreigners, for which the model compensates with an average value through the indicator variable FOREIGN. The implication is that closer attention should be paid to foreign applicants. A study to provide CES or similar ratings for foreign undergraduate schools and a scheme to convert foreign grades to equivalent GPA's would appear to be worthwhile, despite the complexity and difficulty of the proposed task. Even though the validation

---

[a] This number represents the result based on the AC-variables listed in Table 7.

[b] The number in parentheses represents results obtained by adding selected R-variables to the equation (EXPSUM and PRESENT for all three models and, in addition, CREATE for MGMT model only).

[c] The subgroups are not mutually exclusive, because U.S. women belonging to a minority are counted in both groups. The categories are also not collectively exhaustive, because individuals who did not identify minority status are not included in any subgroup, but are included in the total.

results for foreigners are poor, the models may still be useful for foreigners since prediction is intended to be made for the foreign applicant population rather than the more homogeneous foreigners who enrolled. (The results in Table 10 indicate that the enrolled foreign students are more homogeneous, in terms of the variance SY2, than the enrolled U.S. students, for both MGMT and QUANT.)

Effect of Incorporating the R-Variables: By comparing the validation results for the models based on AC-variables only with the results when selected R-variables are added (the latter values are shown in parentheses in Table 10), we see that there is little, if any, gain from the addition. When all observations are grouped together, the mean squared error virtually does not change when the R-variables are added, while the squared correlation decreases in each case. There is some variation across subgroups, but this could easily be due to sampling fluctuations. We conclude that the R-variables do not provide incremental power in predicting academic performance. This does not mean that the R-variables, by themselves, do not predict academic performance. (In fact, they do predict, although not anywhere as well as the AC-variables.) It only means that given the relatively easy to collect data on AC-variables are used for prediction, the more difficult to collect data on R-variables provide virtually no incremental predictive power. Consequently, we conclude that the R-variables need not be considered for predicting academic performance.

Predicted versus Actual Quintiles: To determine whether there are identifiable systematic characteristics about individuals who perform far better or far worse than expected, we grouped observations in the validation sample into quintiles (five ranked categories each with 20% of the data), first in terms of actual performance and second in terms of predicted performance, for each criterion variable. The actual and predicted values were cross-tabulated to obtain further insight into the predictive validity of the models, and to isolate observations for which the predictions were very poor. In an overall sense, the percentages of hits (a hit is obtained when an observation is in the same quintile for observed and predicted scores) are: 29 percent for MGMT, 35 percent for QUANT, and 32 percent for ELECT, compared to an expected percentage of 20 percent for a naive (random) model. For each criterion variable, this result is statistically significant ($n \geq 114$, $p< .01$, one-tailed test).

Given our objective of predicting those in academic difficulty, the percentage of observations with an actual score in the lowest quintile out of all those predicted to be in the lowest quintile is of particular interest. Using this criterion, the MGMT and QUANT models perform quite well with 11/23 or 48 percent for MGMT and 11/22 or 50 percent for QUANT. (The reverse percentages, i.e., those with predicted scores in the lowest quintile out of all those with actual scores in the lowest quintile, were 55 percent and 65 percent for MGMT and QUANT respectively). However, for ELECT the performance was relatively poor with 6/23 or 26 percent correctly predicted ln the lowest quintile. (The reverse percentage for ELECT was 29). Since the objective is to predict those with academic difficulty in the core program, our validation results, averaging about a 55% hit rate, are very encouraging.

A detailed examination of student application folders was undertaken for individuals whose predicted quintile differed by more than two quintiles from the actual quintile. For example, for MGMT, two students were predicted to be in the lowest quintile whereas they ranked in the highest quintile based on their actual scores. Similarly, one student's actual performance ranked in the lowest quintile even though he or she was predicted to be in the highest quintile. For QUANT, such extreme cases did not occur, however. On the basis of the detailed examination, we were not able to identify systematic information in terms of potential predictors available from the application folder to explain dramatic differences between actual and predicted performance. However, on the basis of the detailed examination, we speculate that a student's actual performance is likely to be influenced by the number of core courses exempted. Given that the grades MGMT and QUANT exclude exempted core courses, we postulate that there is a tendency for a student's measured performance in core courses to be depressed (lower than predicted) as the number of core courses exempted increases. This is understandable since a student will tend to exempt a course if he or she already knows the subject matter, and by taking the course for credit could have improved the average performance. An additional consideration for MGMT is that a student's actual performance is likely to be better than predicted, if the student has superior articulation skills in his or her written expression. This is consistent with the explanatory power of the R-variable PRESENT (see Table 5). However, the extent to which these and other reasons can account for some of the prediction error is difficult to determine. Other possible reasons include: differences in students' level of motivation to perform academically, involvement in outside activities such as social clubs or sports, personal difficulties, medical reasons, ability of a student to get along with and get help from classmates, and adaptation to the business school culture.

Predictive Validity for the Graduates of the subsequent year: To obtain further evidence of the models' predictive validity, and to examine the stability of the models' coefficients over time, data were gathered on the subsequent graduating class. Most of the AC-variables are routinely included in computer files constructed for incoming classes. Grades are recorded and stored in separate files. Therefore, no attempt was made to collect the information from the original application folders. However, for some predictor variables the available information was not as detailed as necessary for the computation of the final models reported in Table 7.

For this reason, slightly different versions of the final models were estimated from the two-year data in order to predict academic performance for the subsequent class based on the readily available information. The adjusted R2 values obtained with this version were, as expected, slightly lower than the values obtained for the set of predictors presented in Table 7. Based on the (80%) estimation sample from the earlier years from which the models were estimated, the reductions in adjusted R2 were 0.016 for MGMT, 0.032 for QUANT, and 0.003 for ELECT.

	Two years Holdout Sample	Third year Validation Sample	Two years Estimation Sample
Criterion Variable: MGMT			
$S_Y^2(x100)$	5.57	4.67	4.59
MSE (x100)	4.54	3.84	3.75
r^2	0.18	0.18	0.18
n	115	222[a]	466
Criterion Variable: QUANT			
$S_Y^2(x100)$	11.09	9.73	11.29
MSE (x100)	6.05	5.38	5.10
r^2	0.45	0.48	0.55
n	114	222[a]	459
Criterion Variable ELECT			
$S_Y^2(x100)$	3.76	3.31	3.84
MSE (x100)	3.39	2.96	3.14
r^2	0.10	0.15	0.18
n	117	220[a]	470

Table 11. Predictive Validity for two-years Holdout Sample and third year Sample (With Comparison to two-year Estimation Sample)

The predictive validity results, for the revised model, on the subsequent year are shown in Table 11, along with the overall predictive validity results obtained for the final AC-model with the two-year holdout sample. To get an idea of the shrinkage, i.e., reduction in predictive power when going from estimation to validation, the results for the two-year estimation sample are displayed alongside. Basically, the results are very similar for the two validation samples. The squared correlation coefficients are slightly higher when the subsequent year's data are used. The mean squared error is lower for each model using the subsequent year,

[a] The sample sizes are substantially below the number of students who enrolled for this class. This is due to the fact that data on one or more predictor variables were missing for many observations.

although the variability in the dependent variable (as measured by) is also lower. We conclude that the model holds up very well in the subsequent year.

Stability of regression coefficients: To determine whether the regression coefficients for the two-year data differed significantly from the coefficients for the subsequent year's data, a statistical test was carried out (Chow 1960, Fisher 1970). For each of the three models, the null hypothesis of no change in the parameters could not be rejected at the five percent significance level.

Stability of Factor Structure: In an earlier section we discussed the development of academic performance criteria. Two subsets of core courses (corresponding to MGMT and QUANT) were defined, based on factor analysis of the grades for thirteen core courses for the class from the earliest year. Over time, the content of these courses may change, which may necessitate revisions in the definitions of MGMT and QUANT. Empirical evidence of the need for revisions was obtained by factor analyzing the grades for the class of the subsequent year.

	Factor Loadings After Orthogonal Rotation (Varimax)	
Course	Factor I	Factor II
Dec. Sci. I	**0.730**	0.175
Dec. Sci. II	**0.702**	0.179
Data Anal.	**0.647**	0.255
Acctg. I	**0.765**	0.115
Acctg. II	**0.805**	0.135
Microecon.	**0.697**	0.236
Computers	0.243	**0.399**
Finance	**0.476**	0.216
Marketing	0.236	**0.541**
Macroecon	**0.448**	0.401
Org. Behav.	0.194	**0.401**
Policy	-0.021	**0.451**
Environment	0.133	**0.583**

Note: Bold-facing denotes the factor that has a higher loading for that course.

Table 12. Factor Analysis of Grades in Core Courses: Two Factor Solution for the Subsequent Class (n = 250)

Based on a comparison of the factor loadings for the subsequent year's two-factor solution displayed in Table 12 with the results for the earlier of the two years in Table 3, we see that Finance and Macroeconomics now load more heavily on the first factor, whereas Computers now obtains the higher loading on the second factor. All three changes pertain to courses for which the classification was uncertain in the earlier analysis. We did not ascertain to what extent the changes may be due to sampling error as opposed to changes in instructors and/or course content. The factor structure of Table 12 remained essentially unaffected when the orthogonal rotation (Varimax) was replaced by an oblique rotation.

If it were desirable to change the definitions of the criterion variables (groupings of courses (QUANT and MGMT) in light of the possibly significant change in the factor structure from Table 3 to Table 12, regression analyses could have been carried out with the newly defined criterion variables as per Table 12. However, to the extent that QUANT and MGMT are robust constructs, such changes in the course groupings should not affect the predictive validity of the models developed in Table 7 very much. Using data on the subsequent year's graduates and the models developed for the original criterion variables (see Table 7) to predict the newly created criterion variables (based on Table 12), we obtained r2 values of 0.43 for QUANT and 0.18 for MGMT (compared with 0.48 and 0.18 for the original variables, as shown in Table 11). We conclude that the underlying factors appear to be robust over the years compared.

Implementation

The models developed in the earlier sections enable us to predict the MGMT and QUANT scores for each applicant. Depending on school policy, these predictions can be used in different ways to select applicants. For instance, an overall core program GPA can be obtained by averaging MGMT and QUANT. (If desired, the core program GPA may be replaced by the overall GPA by considering ELECT in addition to MGMT and QUANT.) Applicants may be screened by considering predicted GPA.

At Stanford, the emphasis has been on selecting applicants with the highest management potential among those who are academically viable, i.e., among those who are unlikely to fail in the core program. (Failures are unlikely in ELECT.) We define failure (in the core program) as obtaining a below zero grade point average in MGMT or QUANT or both. (The grading system was detailed earlier.) Although this definition of failure is more stringent than the MBA degree requirements, it can be shown to be a reasonable approximation to the Academic Standards Requirement in the MBA program. The present definition, in comparison to defining failure based on the overall core program GPA, is more likely to identify those in considerable academic difficulty. Furthermore, given the objective of educating the managers of the future, it seems desirable to require an applicant to be proficient in the different skills corresponding to both MGMT and QUANT. By simply taking the average performance across all core courses as a criterion variable, we would have allowed students to compensate an unacceptable performance in one area with a superior performance in another.

Probability of Failure P: For each applicant, we may use the models developed earlier to assess the probability that he/she will fail, if admitted to the program (i.e., QUANT < 0, or MGMT < 0, or both). The technical details of computing the probability P of failure are described in Appendix A.

Critical Probability CP: If an applicant has a substantial chance of failing, i.e., P is too high, the proposed admissions process (see Figure 1) would declare that he/she should not be admitted unless a detailed reading of the application identifies exceptional characteristics not adequately captured by the models. Obviously, an exceptional factor should not be one of the AC or R variables listed in Tables 4 and 5, since these variables have already been considered in developing the models. An illustration of an exceptional characteristic would be a superior performance in courses recently taken but which are not included in the undergraduate grade point averages used in the models, or there was a death in the family. The numerical value corresponding to "an unacceptably high probability of failure," is defined as the critical probability CP. Thus if CP were chosen to be 0.4, those candidates with probability of failure greater than 0.4 are screened out unless exceptional characteristics are identified. The remaining applications (i.e., P $<$ CP) are evaluated further for their management potential.

There are two conflicting implications of decreasing (or increasing) the critical probability CP. To simplify the discussion, assume that all admitted candidates enroll. By choosing a smaller value for CP, the academic screen for selection becomes tighter thereby decreasing the (expected) percentage of the admitted class that will fail. On the other hand, a lower CP means a larger percentage of the applicants would be screened out; the (expected) percentage of those screened out who would have passed the core program had they been admitted, becomes larger. Thus decreasing CP decreases the percent of bad admit decisions but it also increases the percentage of bad reject decisions.

For the two enrolled classes we have summarized in Table 13 the implications of (hypothetically) using alternative values for the critical probability CP. For example,

Critical Probability CP	Percentage Rejected	Percentage Failing[a] Out of Those Accepted	Percentage Passing[b] Out of Those Rejected
.05	85	8	78
.10	73	8	75
.20	50	7	66
.30	33	8	56
.40	24	11	50
.50	18	12	40
.60	12	14	33
.95	1 (n = 6)	19	0

Table 13. Implications of Alternative Critical Probabilities Applied to the Two-year Enrollment Population

if all individuals with a probability of 0.30 or more of failing MGMT, QUANT or both (CP = 0.30) were eliminated, 33% of these classes would not have been admitted. Of those remaining ("accepted") only 8 percent have an inadequate performance in MGMT, QUANT or both. In other words, by using a CP of 0.30, the percent failed in the two years enrolled class could have been reduced from 19% (corresponding to CP = 0.95) to 8%. On the other hand, of those eliminated ("rejected"), 56 percent actually had an adequate performance in the core courses. Raising the critical probability to 0.50 (i.e. eliminate individuals if $P \geq 0.50$) decreases the percentage rejected to 18 percent. Among those eliminated, 40 percent would have had an adequate performance in MGMT and QUANT. However, of those remaining ("accepted") the percentage of students with an inadequate performance increases to 12 percent. Looking at Table 12, it is clear that there is no advantage of decreasing CP below 0.30.

Let us define the presented population as the subset of the applicant population whose predicted probability of failure is less than CP, i.e., the presentation population is what remains of the applicant population after screening out those whose probability of failure is at least CP. This is the population from which most admit decisions will be made based on management potential. The problem of choosing a value for CP may be viewed as reducing the percentage in the presented population failing one or both subsets of core courses but without greatly affecting the population in terms of managerial potential, diversity of backgrounds, and other factors.

[a] A student fails the core program if QUANT < 0 and/or MGMT < 0

[b] A student passes the core program if QUANT ≥ 0 and MGMT ≥ 0

Effect of Screening Rule on Presented Population: To determine the effect of screening based on predicted academic performance on the presented population, we constructed the applicant population by weighting the sample observations on enrollees, non-enrollees, and rejects by the ratio of population to sample size. For three screening rules considered, i.e. critical probability of 0.50, 0.40, and 0.30, we have summarized the general nature of the results using a few selected variables in Table 14

Variable[a]	**Applicant Population**	**Presented Population Corresponding to Critical Probability (CP)**		
		CP = 0.50	**CP = 0.40**	**CP = 0.30**
Presented population size as a percent of applicant population	--	69.8	62.9	53.1
AGE (months)	322	318	319	317
GPATOT[b]	3.17	3.28	3.30	3.33
GMATV	35.6	37.7	38.3	38.3
GMATQ	34.2	37.5	38.2	39.3
GRADWK (months)	4.1	3.7	3.9	3.8
EXPBUS (months)	13.1	12.1	12.0	10.4
GOAL[c]	2.5	2.6	2.6	2.6
WHYMBA[c]	2.4	2.5	2.5	2.6
PRESENT[c]	2.0	2.1	2.2	2.3

Table 14. Effect of Screening Rule on Presentation Population for Selected Variables (Based on the Two-Year Applicant Population)

Comparing the population of all applicants with the presentation population corresponding to CP = 0.40, we note the following results. The presented population is reduced from 4955 U.S. applicants11 to 3118 or to approximately 63 percent. This reduced population is better academically as indicated by, for example, an increase in average GPA from 3.17 to 3.30, and increases in the GMAT scores from 35.6 to 38.3 for the verbal part, and from 34.2 to 38.2 for the quantitative part. The presented population is also better in terms of the qualitative variables GOAL,

[a] See Tables 4 and 5 for definitions of these variables. The entries in the Table represent the average value for the variable in the population defined by the screening rule.
[b] On a scale defined by D = 1, C = 2, B = 3, A = 4.
[c] On a 5 point rating scale with 1 denoting the least desirable rating and 5 denoting the most desirable rating.

WHYMBA, PRESENT (see Table 5), i.e. on the average goals are expressed more clearly, the relevance of an MBA degree is better articulated and the overall case for admission is better presented. On the other hand, the presentation population is younger and somewhat less experienced. However, such less desirable effects are not drastic, considering the large ratio of the presented population to the number of admits. About 5,200 candidates apply to the MBA program. For CP = 0.4, the presented population is 62.9% of the applicant population (see Table 14). Thus the presented population is 0.629 x 5,200 = 3,270. To enroll a class of about 310, usually about 415 are admitted. Thus the admissions officer(s) has to select 415 out of the 3,270 predicted to be academically viable. Note the enormously large size of the academically viable pool of applicants in relation to the number of admits. Thus there is ample opportunity to increase the experience level of the admitted class by approximately selecting from the large pool of academically viable applicants.[12]

The proposed modification in the admission procedure calls for selecting those with the highest management potential from the presented population. As the critical probability is lowered, the academic screen becomes tighter and the size of the presented population decreases. Consequently, the selection for highest management potential is achieved from a smaller subset of applicants thereby lowering the average management potential of those selected. Thus the cost of using lower values of CP can be assessed by determining the corresponding reduction in average management potential of the selected candidates. To quantify this effect, a Monté Carlo simulation was carried out. Imagine a population of applicants to the MBA program equal to 5,200. Let X denote management potential, so that if academic performance were not a consideration, the 415 applicants with the highest values of X among the 5,200 will be admitted. With X being measured on an interval scale, and normally distributed with mean of 6 and standard deviation of 1,13 the average score for the 415 best applicants out of the total population was found to be 7.85. If a CP = 0.40 were used, only 3,270 of the 5,200 applicants would be considered as academically viable (62.9 percent of 5,200 -- see Table 14). Thus, 415 applicants who have the largest values of X among the 3,270 would be admitted. In this case the average score for these 415 applicants was 7.63. Both scores can be compared to an average of 6.00 based on random selection (naive rule). Thus, for the presented population we obtain an index of 88 percent (7.63-6.00)/(7.85-6.00) of the maximum achievable average management potential. The corresponding indices for CP = 0.5 and 0.3 are 90.5 and 84 percent, respectively.[14]

The computed index should be viewed as a conservative estimate of the achievable average management potential under the screening rule. In the Monté Carlo simulation, academic performance was assumed to be independent of managerial potential. Actually these two variables are likely to be positively correlated

(Weinstein and Srinivasan, 1974). Furthermore, the current admissions procedures do not maximize average managerial potential without considering academic

potential. Finally, there is always some error in assessing managerial potential (as well as in the estimated academic performance). Consequently, those admitted based on estimated managerial potential are not necessarily the best in terms of actual managerial potential.

Combining the results of Tables 13 and 14, we find that the use of a critical probability of 0.3 or 0.4 is likely to significantly reduce the percentage failing in the core program but not likely to lead to a significant loss in terms of the achievable management potential of the admitted class. Of course, the possibility remains that the faculty may correspondingly increase the academic rigor of the program and hence the percentage failing may not drop as much as expected. This phenomenon is, however, less likely at Stanford since the proportion of fail grades is not based on any pre-specified distribution.

Model Updating: As long as information on all AC-variables included in the academic performance models is routinely available on a year-to-year basis, it is straightforward to update the coefficients of the models to predict MGMT, QUANT and ELECT. However, based on the results reported in Table 11 there is no evidence that updating of the coefficients is necessary every year. More substantial effort is required to update the models, i.e., to determine empirically the potential value of changing the set of AC-variables for predicting any of the three criterion variables. When the models are updated, the definitions of the criterion variables MGMT and QUANT may have to be changed to reflect changes in course content and structure. We speculate, however, that unless such changes are expected to be dramatic, model updating could be carried out once every five years.

Summary and Conclusions

With the growth in the demand for graduate education in management, the task of selecting the most promising candidates from the applicants to a graduate school of business has become even more challenging. Admission decisions are usually made based on “clinical judgments,” i.e., overall evaluations of applicants by one or more admissions officers. An impressive amount of empirical evidence has accumulated in the behavioral literature on decision making demonstrating that actuarial models developed, for example, by multiple regression analysis, are superior in predictive ability to clinical judgments. The superiority of actuarial models derives from the consistency inherent in model-based predictions, and from the fact that the model is obtained by systematically linking information about actual performance to predictors of that performance.

In this paper we outline an approach for aiding and strengthening the MBA admissions process. The proposal does not advocate replacing the current process by a mechanized decisions. Under the proposal, every application will continue to be read carefully in order to consider applicants' unique characteristics. The proposal would aid the admissions process by first predicting each applicant's academic performance in the MBA core program using information readily available from an application folder. If the prediction indicates that the applicant, if admitted, has a "substantial" chance of an unacceptable performance in the MBA core program, he/she would be considered further only if a detailed reading of the application identifies exceptional circumstances or characteristics not adequately captured by the model. Admission decisions are then made by the admissions officer(s) from the remaining pool of applicants based mainly on the applicants' likely management potential. Attention is also paid to the fact that an individual, because of his/her unique back ground and experience can make significant contributions to the program.

The present paper focuses its attention on predicting academic performance. The development of criterion (dependent) and predictor variables for the academic performance models was discussed in detail. Factor analysis was used to define two separate academic performance criterion variables; MGMT, defined as the GPA (Grade Point Average) in managerially oriented MBA core courses, and QUANT, defined as the GPA in quantitatively oriented core courses. A third criterion variable, ELECT, is the GPA in all elective courses taken in the business school. The set of potential predictor variables incorporates all information which is available in numerical form or can be coded directly from the application. We showed that other information which can be captured through the use of rating scales does not add materially to the predictive power of the academic performance models.

The models were developed from a subset of applicants who graduated in two years. No statistically significant differences in the models were observed between females, minorities, foreigners, and U.S. white male applicants. An analytical examination of the correlation between errors in the three academic performance models, and the possible reduction in variation on predictor variables for the admitted class (compared to all applicants) showed that the results remain essentially unaltered. The assumptions of the models were verified, and the models validated by comparing the predictions with the actual GPA's, using a holdout sample of graduates in the two years, and with data on the graduates in a third year, and by considering the stability of the factor structure (for the criterion variables) as well as the stability of the models' coefficients. We conclude that the models provide useful predictions, and the models hold up well in the subsequent year.

The results displayed in Table 7 indicate that an index of undergraduate school quality provided by the Educational Testing Service, undergraduate GPA (excluding the sophomore year) and the verbal score on the Graduate Management Admission Test (GMAT) are significant predictors of MGMT, QUANT and ELECT. In addition, the GMAT quantitative score is a very important predictor for QUANT. As expected, GMAT verbal score is more predictive of MGMT than QUANT. Prior experience in business is a significant predictor for QUANT. The variable AGE, which is a surrogate for business and non-business experience, is significantly positively related to both MGMT and ELECT. Undergraduate major areas are differentially related to MGMT, QUANT and ELECT. The (adjusted) coefficient of determination ($\bar{R}2$) is approximately 0.18 for MGMT and ELECT and 0.53 for QUANT. (As indicated earlier, these numbers roughly hold up on cross-validation on a holdout sample and on a subsequent year.) After adjusting for the unreliability in the MGMT and QUANT grades, the coefficients of determination become 0.32 and 0.62, respectively. In terms of the ability to identify the applicants likely to be in academic difficulty, the model has approximately a 55% success rate, i.e., approximately one half of those predicted to be in the bottom quintile (20%) of the class do, in fact, actually end up in the bottom quintile.

We indicate how the probability of unacceptable performance is calculated. Those applicants with calculated probability less than the critical probability and/or with unique characteristics constitute the "presented population," i.e., those who will be presented for the evaluation of management potential, and from whom the final selection will be made. We examined the effect of using different values for the critical probability on the "presented population." The results indicate that the academic quality of the admitted applicants can be enhanced without greatly affecting the achievable average management potential.

The decision aid was used at the Stanford University Graduate School of Business for over fifteen years. Because quantitative information may receive an unreasonably high attention compared to qualitative information, the admissions office did not receive the actual probability of failure, but applicants were merely divided into three categories: Acceptable risk ($P<0.3$), Unacceptable risk ($P>0.4$), and questionable risk ($0.3<= P <=0.4$). The questionable risk category requires further digging into the admissions folder regarding the likely academic performance. As stated earlier, every application was read carefully by at least one admissions officer. The models were not updated and hence not currently used in the admissions process after the MBA program course structure underwent a major change under which each core course was offered at three different levels based on the prior preparation of the student in the subject matter covered by that core course.

APPENDIX A

Determination of the Probability P of Failure in the MBA Core Program

As explained in the text, performance in the MBA core program is measured by its two components, MGMT and QUANT, denoting the average grades in the management oriented courses and quantitatively oriented courses, respectively. Since a failure (i.e., average grade falling below zero) in either group of courses is considered to be failure in the core program, we define

$$P = \text{Probability of failure in the core program}$$

$$= \text{Prob (MGMT}<0 \text{ and/or QUANT}<0\text{) so that}$$

$$P = 1 - \text{Prob (MGMT} \geq 0 \text{ and QUANT} \geq 0\text{)} \quad (1)$$

Let X and Y denote the random variables measuring the actual performance in MGMT and QUANT, respectively, for a specific applicant if he/she were to be admitted to the MBA program. By using the values on the predictor variables for this applicant (available from the application) and the regression coefficients corresponding to the beta weights reported in Table 7, we can determine the predicted values X' and Y'. The uncertainty of these predictions is indicated by the estimated standard errors of prediction SX' and SY', which depend on the specific values of the predictor variables. However, these estimates can be taken to be approximately equal to the estimated standard errors of estimate (standard deviations of the residuals), provided on the bottom of Table 7. In general, the standard error of prediction exceeds the standard error of estimate. Nevertheless, the difference between these two standard errors was found in our study to be negligible. Srinivasan (1977, Eq. (45)) has proven theoretically that the above difference, expressed as a fraction, is given by the ratio of the number of predictor variables to the number of observations; this ratio is approximately .01 in the present study.

The actual performance (X,Y) of an applicant, if admitted to the program, can be thought of as randomly distributed with means (X',Y') and standard deviations (SX', SY'). From the Table 8, the correlation between the prediction errors in MGMT and QUANT is $r = 0.389$. In Section 6, it was indicated that the residuals of the multiple regression models satisfied the assumption of normality. Consequently, the actual performance (X,Y) of the applicant may be assumed to be bivariate normally distributed with means (X',Y') standard deviations (SX', SY'), and correlation r. From Eq. (1) we have

$$P = 1 - \text{Prob}\ (X \geq 0 \text{ and } Y \geq 0) \quad (2)$$

The probability that $X \geq 0$ and $Y \geq 0$ can be determined by numerically integrating the bivariate normal distribution over the positive quadrant of the X-Y plane (i.e., over $X \geq 0$ and $Y \geq 0$). The probability of failure is then determined from Eq. (2). Curve AB in Figure 4 gives the locus of all (X',Y') for which P = 0.4.

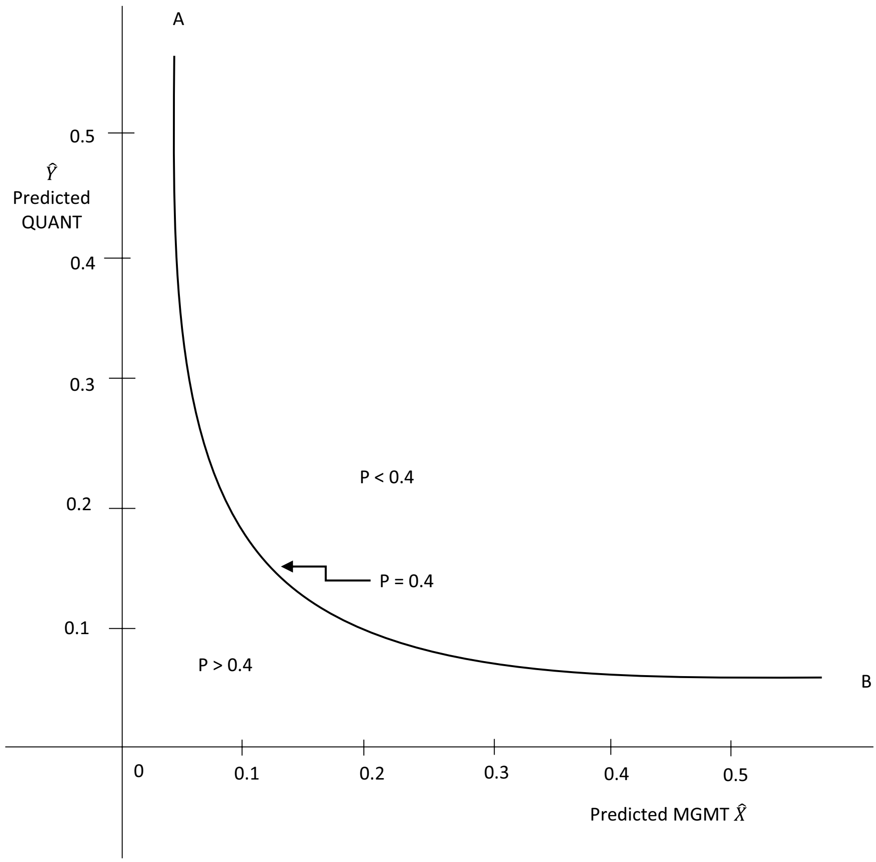


Figure 4. Values of Predicted QUANT (Y') and Predicted MGMT (Y') for which Probability of Failure (P) = 0.4

This is obtained by connecting by a smooth curve, the points (X',Y') for those applicants whose predicted probability P, as determined above, is approximately equal to 0.4. Consequently, if we adopt a critical probability CP = 0.4 as the admission policy, then those applicants with P > 0.4 will be screened out unless exceptional circumstances warrant that such applicants be considered further.

Thus all applicants whose predicted values for MGMT and QUANT fall below the curve AB will most likely be screened out under such a policy. The applicants for whom the predicted values for MGMT and QUANT fall above the curve AB, will be considered for admission primarily based on their management potential.

The curve AB may be interpreted intuitively as follows. Applicants whose predicted (X’,Y’) fall near A are predicted to score highly on QUANT. They are, therefore, unlikely to fail in QUANT so that there is approximately a 40% chance that their actual MGMT score (X) will be negative. (Note that there is a 50% chance for a negative MGMT score if point A is moved left to lie on the Y-axis.) Similarly, applicants whose predicted (X’,Y’) fall near B, are unlikely to fail in MGMT, but have approximately a 40% chance of failing in QUANT. Applicants whose predicted values (X’,Y’) fall midway between A and B on the curve have approximately the same chance of failing MGMT as of failing QUANT. The possibility of such applicants failing in both MGMT and QUANT is incorporated by the curvature in the middle of AB.

.

APPENDIX B

Levels of Prior Work Experience[a]

1. Executive Responsibility with an organization with 250 or more employees. Executive responsibility means that the applicant had control over policies which affect multiple functions of an organization or divisions of an organization. This might include policies affecting budgeting, employment, production, marketing, and planning for an organization. Employment in private and nonprofit agencies would include the president and the vice presidents of an organization. Employment in the public sector would include department and agency directors and chief deputy directors, and in some multi-faceted agencies the deputy directors.

2. Executive Responsibility with an organization with less than 250 employees. See employment type 1 for a description of executive responsibility. This includes substantial and successful self- employment.

3. Management Responsibility for an activity or function requiring the direct supervision of 50 or more employees. These are essentially line management responsibilities. They require the management and operation of one function of an organization, and usually involve carrying out policies established at the executive level. They generally exclude jobs which are scientific or technical in nature or involve an administrative specialty, although they might include the

[a] To obtain EXPSUM, the number of months of experience in each category is multiplied by the weights (for categories 1 through 9 respectively): 2.00, 1.71, 1.00, 0.86, 0.67, 0.50, 0.40, 0.25 and 0.25, and the products are added.

management of such activities. Occupations include middle to low level officers in government, corporations, and nonprofit agencies. They include managers, supervisors, and department heads and their assistant department heads in industrial establishments.

Experience may be in a wide range of fields, such as: agriculture, forestry and fishing; mining; construction; manufacturing (as a production supervisor, branch manager, superintendent, or supervisor); transportation, communications and the utilities industry (as an operations manager, maintenance supervisor, or station superintendent); finance, insurance, and real estate (as a controller, brokerage office manager, or operations officer); service industry (as a hotel manager, or restaurant manager); military officer.

4. Management Responsibility for an activity or function requiring the direct supervision of up to 50 employees other than clerical personnel. See employment type 3 for a description of management responsibility.

5. Professional and Technical Support to business, industry, and government. Employment of this type includes high level staff specialists who influence an organization at a substantial level. This includes employment as industrial and engineering psychologists, lawyers, systems analysts, organization development and management specialists, chemists, economists, engineers, statisticians, market research analysts, sociologists, and intelligence specialists.

6. Administrative Specialists in business, industry, and government. Occupations in this area require a knowledge of particular functions within an organization rather than a knowledge of the operations of an organization. Occupations of this type involve the more routine non-clerical duties of an organization.

Jobs include those of accountant, auditor, budget analyst, computer analyst, purchasing agent, buyer, programmer, field representative, advertising manager, public relations officer, lobbyist, job analyst, personnel officer, inspector, investigator, administrative assistant, and technical writer.

7. Administrative/Management Trainee or Intern

8. Clerical and Sales. This includes such occupations as secretary, bookkeeper, office clerk, and salesperson.

9. Other. This includes such occupations as poet, actor, teacher, laborer, truck driver, waitress, barber, pilot, homemaker, miner (i.e., much of the work force of the world).

REFERENCES

Chow, G. C. (1960), "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica,* 28, 591-605.

Cronbach, Lee J. (1970), Essentials of Psychological Testing (third edition), New York: Harper & Row.

Dawes, R.M. and B. Corrigan (1974), "Linear Models in Decision Making," *Psychological Bulletin*, 81, 95-106.

Deckro, R. F. and H. W. Woundenberg, (1977), "M.B.A. Admission Criteria and Academic Success," *Decision Sciences*, 8, 765-769.

Dixon, W. J. and M. B. Brown (1979), *eds., Biomedical Computer Programs, P-Series,* Berkeley, CA: University of California Press.

Fisher, F. M. (1970), "Tests of Equality between Sets of Coefficients in Two Linear Regressions: An Expository Note," *Econometrica*, 38, 361-366.

Gayle, J. B. and T. H. Jones (1973), "Admission Standards for Graduate Study in Management," *Decision Sciences*, 4, 421-425.

Hall, B. H. and R. E. Hall (1980), *Time Series Processor (Version 3.5)*, Stanford, CA: National Bureau of Economic Research.

Harrell, M. S., T. W. Harrell, S. McIntyre, and C. Weinberg (1977), "Predicting Compensation among MBA Graduates Five and Ten Years after Graduation," *Journal of Applied Psychology,* 62, 636-640.

Harrell, T. W. (1972), "High Earning MBA's," *Personnel Psychology,* 25, 523-530.

Harrell, T. W. and M. S. Harrell (1973), "The Personality of MBA's who Reach General Management Early," *Personnel Psychology*, 26, 127-134.

Lieberman, L. (1977) "Admissions Reader Questionnaire Summary (Step 1)," Admissions Research Group, Graduate School of Business, Stanford University.

Livingston, J, S. (1971), "The Myth of the Well-Educated Manager," *Harvard Business Review*, 49(1), 79-89.

Lord, F. M. and M. R. Novick (1968), *Statistical Theories of Mental Test Scores*, Reading, MA: Addison-Wesley Publishing Company.

Marshall, G. L. (1964) “Predicting Executive Achievement,” Unpublished doctoral dissertation, Boston, MA: Harvard Business School.

Nie, N. H., C.H. Hull, J. G. Jenkins, K. Steinbrenner and D. H. Bent (1975), *eds., Statistical Package for the Social Sciences* (second edition), New York: McGraw-Hill Book Company.

Page, A. N. and R. E. West (1969), “Evaluating Student Performance in Graduate Schools of Business,” *Journal of Business*, 42, 36-41.

Pfeffer, J. (1977), “Effects of an MBA and Socioeconomic Origins on Business School Graduates’ Salaries,” *Journal of Applied Psychology*, 62, 698-705.

Pitcher, B. (1971), “Summary Report of Validity Studies Carried out by ETS for Graduate Schools of Business, 1954-1970,” Princeton, N.J.: Educational Testing Service.

Siegel, S. (1956), *Nonparametric Statistics for the Behavioral Sciences*, New York: McGraw-Hill Book Company.

Srinivasan, V. (1977), “A Theoretical Comparison of the Predictive Power of the Multiple Regression and Equal Weighting Procedures,” Research Paper No. 347, Stanford, CA: Graduate School of Business, Stanford University.

Srinivasan, V. and A. G. Weinstein (1973), “Effects of Curtailment on an Admissions Model for a Graduate Management Program,” *Journal of Applied Psychology*, 58, 339-346.

Weinstein, A. G. and V. Srinivasan (1974), “Predicting Managerial Success of Master of Business Administration (MBA) Graduates,” *Journal of Applied Psychology*, 59, 207-212.

Williams, F. J. and T. W. Harrell (1964), “Predicting Success in Business,” *Journal of Applied Psychology*, 48, 164-167.

Zellner, A. (1962), “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias,” *Journal of the American Statistical Association*, 57, 348-368

END NOTES

[1] Students who fail a core course have to pass the requirements of the course by taking it again or by passing an exemption exam. The present analysis uses only the grade received during the first time a core course was taken.

[2] Since the number of credit units for exempted courses is equal to zero, this amounts to computing a weighted average grade using only the courses taken for credit.

[3] The sample size available for the computation of Cronbach alpha for QUANT is considerably smaller than the sample size available for MGMT due to the popularity of taking the exemption examination for one of the courses included under QUANT.

[4] For ELECT, the set of courses taken by students varies widely. It is not meaningful, therefore, to compute the Cronbach alpha for this variable.

[5] An F statistic was computed to "verify" such a grouping of variables. However, this test statistic is to be interpreted only as a crude indication because the same regression coefficients were used to decide which variables to combine, as well as to "test" the validity of the obtained grouping.

[6] The regression coefficient corresponding to the grade point average during the sophomore year (GPA2) was not statistically significant in any of the three academic performance models. This may be a result of the so-called "Sophomore Slump" phenomenon.

[7] This result was obtained by using the actual regression coefficients for AGE and AGESQ (not reported in Table 7) and not the beta weights (reported in Table 7). AGESQ, when added to the equations for MGMT and QUANT, resulted in coefficient that was not statistically significant.

[8] In the regression model, CREATE is an indicator variable which takes on the value one if an applicant had created his or her own artistic work, publication, invention, or business venture, and zero otherwise.

[9] This standard deviation is obtained by weighting the observations for enrollees, non-enrollees (but admitted to the program), and rejected applicants by the sizes of these three groups among all applicants.

[10] This argument is not exact since Cronbach alpha coefficient is not equal to reliability but only a lower bound on reliability. As a rough check of the extent to which the alpha coefficient may be smaller than the true reliability, six different

Guttman reliability coefficients (Nie et. a1, 1975) were also computed. The maximum of the six Guttman coefficients exceeded the corresponding Cronbach alpha only by 0.006 for MGMT and by 0.005 for QUANT. Although this is no proof, these results lead us to believe that the Cronbach alpha is likely to be a tight lower bound on reliability for MGMT and QUANT.

[11] Since information on the predictor variables was collected for a subset of rejected U.S. applicants only, the results do not include foreign applicants.

[12] This, however, is not true for many other schools which do not have as large an applicant/admit ratio. For such schools, admission decisions may be better made by simultaneous consideration of predicted academic performance and management potential.

[13] The normally distributed variable was approximated by adding 12 uniformly distributed independent random numbers with range (0, 1). The critical results below, in terms of the index of maximum achievable management potential, are unaffected by the arbitrary choices of mean = 6 and standard deviation=1.

[14] If X is assumed to be uniformly distributed instead, the indices corresponding to CP = 0.3, 0.4 and 0.5 are 97%, 94.8% and 92%, respectively.