College Board Report No. 95-1
ETS RR No. 95-3

Feasibility of Using the SAT® in Academic Guidance

LAWRENCE J. STRICKER, DONALD A. ROCK,
and NANCY W. BURTON

College Entrance Examination Board, New York, 1995

Acknowledgments

Thanks are due to Min hwei Wang for computer programming and Rick L. Morgan and Cathy L. Wendler for reviewing a draft of this report.

Lawrence J. Stricker is a principal research scientist at ETS.
Donald A. Rock is a principal research scientist at ETS.
Nancy W. Burton is a program administrator at ETS.

The College Board is a national nonprofit association that champions educational excellence for all students through the ongoing collaboration of more than 2,900 member schools, colleges, universities, education systems, and organizations. The Board promotes—by means of responsive forums, research, programs, and policy development—universal access to high standards of learning, equity of opportunity, and sufficient financial support so that every student is prepared for success in college and work.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101-0886. The price is $15.

Printed in the United States of America.

Contents

Abstract

This study appraised the validity of SAT scores, in combination with grades in high school courses and the number and difficulty level of these courses, in predicting college grades in various fields of study, with the objective of providing SAT takers with predictions of their academic performance in different fields for guidance purposes. The possible impact of this feedback on the flow of students into specific major fields was also assessed. Data on an entering class at a large state university provided the basis for the study. The SAT and collateral variables were able to predict college grades in different academic areas by taking into account marked variations in grade distributions among fields of study. These predictions could be potentially useful in academic guidance. Students' predicted grades in the various fields and their intended majors were unrelated. Most students received their highest predicted grades in the humanities, although few intended to major in this area. Hence, such feedback would probably stimulate students to take courses or major in the humanities.

Introduction

Accurate predictions of performance in different fields of study are an essential ingredient of academic guidance. Students' decisions about whether to attend college and, if so, which courses to take and which majors to select are affected by a variety of concerns. However, anticipated academic success or failure is not only important in its own right, but also colors interests, confidence, and other determinants of these choices (e.g., Hackett, 1985; Peterson & Roscoe, 1983).

Several efforts have been made over the years to devise test batteries that can be used to make forecasts about success in different courses. An early example is the Differential Aptitude Tests (Psychological Corporation, 1992), first published in 1947, which consist of eight ability tests. The median validity of the individual tests in the current battery in predicting students' grades in courses from eighth to twelfth grade in four areas ranged, for girls and boys, respectively, from .14 and .16 for Clerical Speed and Accuracy to .48 and .52 for Numerical Ability. The median validity for the eight tests was .35 and .37 for girls and boys, respectively (Bennett, Seashore, & Wesman, 1984).

Equally well known is the Washington Pre-College Test (Noeth, 1979), which was used in Washington for several decades. A recent version of this test employed eight ability subtests and high school grades in six areas to predict college grades. This test's median validity was .45 in predicting college grades (primarily for the first two years) in 48 subjects.

Three major studies have investigated the usefulness of other test batteries in forecasting high school or college grades. French (1963) used ability tests, interest scales, and personality scales to predict college grades in freshman courses and in majors. The median validity of the battery was approximately .42 in predicting grades in 10 majors. Its median validity was .46 in making "differential" predictions of these grades (e.g., the correlation between the predicted and actual difference between grades for English and history) (Mollenkopf, 1950).

In a second investigation, French (1964) employed many of the same ability tests and personality scales plus interest information tests to predict senior grades in 10 high school courses. The median validity of this battery, for girls and boys, respectively, was approximately .47 and .45 in predicting course grades. The battery's median validity was .51 and .33, respectively, for the two groups in making differential predictions of grades in five courses.

Katz and Norris (1972) used the PSAT/NMSQT and interest scales to predict self-reported grades in senior-year high school courses in four fields. The median validity in predicting grades for girls and boys, respectively, was .44 and .38 for the PSAT-Verbal (V), .49 and .41 for the PSAT-Mathematical (M), and .22 and .22 for the most relevant interest scale. The median validity for the set of measures for girls and boys, respectively, was .37 and .29 in making differential predictions of grades in these fields.

Katz and Norris also used these same measures plus self-ratings of interests and self-reported senior-year high school grades to predict self-reported freshman college grades. The median validity in predicting grades in four fields, for girls and boys, respectively, was .25 and .24 for the PSAT-V, .24 and .22 for the PSAT-M, .14 and .16 for the most relevant interest scales, .14 and .14 for the most relevant self-rating of interests, and .26 and .32 for the most relevant high school grades. The median validity for the set of measures was .44 and .32, for the two groups, in making differential predictions of grades in these fields.

Although promising, two important problems with this line of research limit its usefulness. First, the grade criteria were not comparable. The students took different courses, but the grades were aggregated and treated as equivalent. Furthermore, in most of these efforts (the Katz and Norris research is an exception), the

analysis for each grade criterion employed a different subsample of students (those with grades in a course), but treated the samples as equivalent. Second, in most of this research (except for the Noeth study), the focus was on predicting the rank order of grades, without regard to their level (e.g., A, B, C). These two issues are critical because different courses, even those in the same field, vary markedly in their intellectual demands and grading standards (e.g., Goldman & Hewitt, 1975; Prather, Smith, & Kodras, 1979). The situation is further complicated by the considerable self-selection that occurs in course taking (e.g., Ramist, Lewis, & McCamley, 1990; Willingham & Breland, 1982), making the samples of students in different courses nonequivalent. Variations in the difficulty of and grading standards for different courses translate into differences in grade distributions by field, making absolute prediction of students' grade levels as important as relative prediction of their grade rankings.

Workable solutions exist for these problems. First, comparable grades can be derived by treating the unavailable grades for students who did not take particular courses as a missing data problem (Little & Rubin, 1987), with the missing data considered as "missing at random" in the sense that they are predictable from available grades. This procedure was used in a recent study by Stricker, Rock, Burton, Muraki, and Jirele (1994). The missing grades were imputed by applying the EM algorithm to the missing data matrix (Dempster, Laird, & Rubin, 1977), generating a complete set of grades for all students. These imputed grades can be used to obtain comparable GPAs for different fields of study, reflecting how well students would be expected to perform in them. Second, grade level information can be readily preserved by using letter grades in the predictions.

Accordingly, the purpose of the present study was to explore the feasibility of using SAT scores and collateral information in academic guidance, employing imputation procedures and letter grades to deal with the problems of lack of comparability of grade criteria and absence of absolute predictions of grades that affect work in this area. More specifically, the primary aim was to appraise the validity of SAT scores, in combination with grades in high school courses and the number and difficulty level of those courses, in predicting grades in various kinds of college courses, with the objective of providing SAT takers with predictions of their potential academic performance in different fields of study. A secondary aim was to assess the potential impact of this feedback to SAT takers on the flow of students into these fields. This is a matter of some interest in view of concerns about the underrepresentation of women and minorities in the natural sciences and engineering (e.g., Berryman, 1983; National Science Foundation, 1990).

Method

Overview

Imputed grade-point averages (GPAs) in three fields of study were obtained for a sample of college freshmen with at least one course grade in each of the fields; actual GPAs in these fields were also obtained for comparative purposes. Predicted GPAs were estimated from SAT scores and collateral high school variables. The observed and predicted GPAs in the fields were compared to assess the agreement between the two kinds of grades—relative correspondence for grades on the standard four-point scale and absolute correspondence for letter grades. To appraise the potential effect of feedback concerning GPA on course taking, the field with the highest predicted GPA for each student was identified and compared with his or her intended major. This analysis used both the standard grades on a four-point scale and their percentile rank equivalents for the field to represent two models of students' choice of majors: selecting the field in which the highest grade in absolute terms will be earned or the field in which the highest grade in relative terms (vis-à-vis other grades in that field) will be earned, respectively.

Sample

The sample consisted of 981 students (579 women and 402 men) in the Fall 1988 entering class at the main campus of a large state university. The sample was limited to students who (a) were full-time freshmen enrolled in the seven undergraduate schools (three liberal arts colleges and four professional schools); (b) had at least one course grade in each of three fields (humanities, natural science, and social science); and (c) had complete data for the SAT, high school rank (HSR), Student Descriptive Questionnaire (SDQ) variables concerning high school courses (grades; years of study; enrollment in honors, advanced placement, or accelerated courses), and percentage of college-bound seniors in the student's high school. (An exception was made for grades in arts and music because a substantial number of students had not taken any such courses. Students without a grade in these courses were included in the sample, and the mean grade for the sample was substituted for the missing grade.) This is a subsample of the

4,351 students in the same class used in previous studies (Stricker, Rock, & Burton, 1993; Stricker et al., 1994).

GPAs

First-semester grades in all degree-credit courses were used. (No Credit grades assigned to students in one of the liberal arts colleges in lieu of Fs were treated as Fs, and temporary grades were treated the same as permanent ones.) These grades, obtained from the university's longitudinal data base, were also used in the earlier studies (Stricker et al., 1994). The grades were grouped into three categories adapted from the university's subject classification: humanities, natural science, and social science (Stricker, Rock, & Burton, 1991). (Several preliminary analyses were conducted in an attempt to categorize courses empirically, but the results were uninterpretable or unusable: principal-axis factor analyses of imputed department grades for the entire class [N = 4,307] and for the subsample with grades in the humanities, natural science, and social science [N = 1,659]; a factor analysis of available department grades for 29 departments with 84 or more students [N = 3,616]; and a hierarchical cluster analysis of profiles of multiple regression weights for SAT-V, SAT-M, and high school rank, and the intercept against a course grade criterion for 40 individual courses with 35 or more students.) Three versions of these grades were used: continuous grades on the standard 1 to 4 scale, percentile rank equivalents for the field of study (based on data for the same sample), and letter grades.

Actual GPAs

Actual GPAs were based on available grades, weighted by the number of credit hours per course.

Imputed GPAs

Imputed GPAs were based on imputed grades for departments. Maximum likelihood estimates of the GPAs in each of 51 departments were obtained with the BMDP AM Program (Frane, 1990), using actual GPAs for the departments from a larger sample of 1,659 students with course grades in humanities, natural science, and social science. (Thirty-five other departments were excluded because they had fewer than four GPAs or no variation in GPAs; the 678 additional students in this analysis were excluded from the study sample because of missing data for the SAT, HSR, or SDQ.) An unweighted GPA was computed.

Predictors

The predictors were:

1. SAT-Verbal (scaled scores)
2. SAT-Mathematical (scaled scores)
3. HSR (converted to normalized T scores)
4. Percentage of college-bound seniors in high school
5. Years of high school courses in arts and music
6. Years of high school courses in English
7. Years of high school courses in foreign and classical languages
8. Years of high school courses in mathematics
9. Years of high school courses in natural science
10. Years of high school courses in social science and history
11. High school honors, advanced placement, or accelerated course in arts and music
12. High school honors, advanced placement, or accelerated course in English
13. High school honors, advanced placement, or accelerated course in foreign and classical languages
14. High school honors, advanced placement, or accelerated course in mathematics
15. High school honors, advanced placement, or accelerated course in natural science
16. High school honors, advanced placement, or accelerated course in social science and history
17. Average grade for high school courses in arts and music
18. Average grade for high school courses in English
19. Average grade for high school courses in foreign and classical languages
20. Average grade for high school courses in mathematics
21. Average grade for high school courses in natural science
22. Average grade for high school courses in social science and history
23. Average grade available for high school courses in arts and music

The data for all variables, except the percentage of college-bound seniors in the student's high school, were obtained from the university's longitudinal data base. The data on high school courses came originally from the SDQ, completed by students when they registered for the SAT. The data on college-bound seniors came from the AI Profile Survey of high school officials, conducted by Educational Testing Service in 1988. Average

grade available for high school courses in arts and music was included to adjust for the effect of missing data on this grade variable by capitalizing on the information inherent in their presence or absence (J. Cohen & P. Cohen, 1983). Data on this variable were missing for 316 students (32.2 percent of the sample).

Other Variables

Other variables in the study were sex, ethnicity (white, all others), and intended major (humanities, natural science [including architecture and engineering], and social science). The data for these variables were obtained from the university's longitudinal data base; the data for college majors came originally from the SDQ. Majors were grouped into categories derived from the SDQ classification (Stricker et al., 1991).

Analyses

A product-moment intercorrelation matrix was computed for the predictors and the actual and imputed GPAs; multiple regressions with the GPAs were calculated, using the entire set of predictors; and predicted GPAs were obtained. Predicted GPAs were converted to percentile ranks and letter grades. (The intercorrelation matrix for the predictors and the observed and predicted GPAs are available from the first author; the standardized regression coefficients for predicting actual and imputed GPAs in the three fields from the predictors are in the Appendix.)

Parallel analyses were conducted for observed and predicted GPAs and, in turn, for standard, percentile rank, and letter-grade versions of these GPAs. Associations between standard GPAs were assessed by product-moment correlations; corresponding relationships between letter-grade GPAs were appraised by weighted Kappa coefficients (Cohen, 1968). (The "disagreement weights" were 0 for the diagonal cell, 1 for the first cell off the diagonal, 2 for the second cell, etc.) Associations between the field of study with the highest predicted GPA for each student and his or her intended major were evaluated by unweighted Kappa coefficients (Cohen, 1960).

Results and Discussion

Correspondence Between Observed and Predicted GPAs

The correlations between the observed and predicted standard GPAs for the three fields of study—actual and imputed versions—are presented in Table 1. The associations between the letter-grade GPAs for the fields are shown in Table 2.

The observed and predicted actual GPAs correlated moderately (r = .43 to .45), but these correlations may be attenuated by the unreliability of the observed GPAs, for each is typically based on a single grade. The corresponding observed and predicted letter grades also agreed moderately (Kw = .20 to .22).

The observed and predicted imputed GPAs correlated moderately (r = .49 to .52). The corresponding observed and predicted letter grades also agreed moderately (Kw = .25 to .27).

In brief, all the observed GPAs—the actual and imputed GPAs and their letter-grade forms—were moderately related to the corresponding predicted GPAs.

Correspondence Between Fields: Observed GPAs

Means and standard deviations of the observed standard GPAs—actual and imputed versions—for the three fields of study are reported in Table 3 and the intercorrelations between the GPAs for the fields are shown in Table 4. Distributions of the letter-grade forms of these GPAs appear in Table 5, and associations between the letter grades for the fields are presented in Table 6.

TABLE 1

Correlations Between Observed and Predicted GPAs

	GPA	
Field	*Actual*	*Imputed*
Humanities	.44	.49
Natural science	.43	.51
Social science	.45	.52

Note: N = 981.

TABLE 2

Kappas Between Observed and Predicted Letter Grades

	Letter Grade	
Field	*Actual*	*Imputed*
Humanities	.22	.27
Natural science	.20	.25
Social science	.21	.26

Note: N = 981.

TABLE 3

Means and Standard Deviations of Observed GPAs for Three Fields of Study

	GPA			
	Actual		*Imputed*	
Field	*Mean*	S.D.	*Mean*	S.D.
Humanities	3.06	.70	3.11	.28
Natural science	2.31	1.13	2.73	.43
Social science	2.61	.99	2.80	.49

Note: $N = 981$.

For actual GPAs, the mean GPA for humanities (3.06) was substantially higher than the other two GPAs, and natural science (2.31) was somewhat lower than social science (2.61). The letter grades followed a similar pattern: the median was B for humanities, B– for natural science, and B– for social science. The three GPAs correlated moderately (r = .41 to .48), but these correlations may be attenuated by the unreliability of the GPAs, as noted previously. In contrast, the letter grades agreed slightly for humanities and natural science (Kw = .17), but agreed moderately for humanities and social science and for natural science and social science (Kw = .25 to .31), reflecting level differences in grade distributions, most pronounced for humanities and natural science.

The pattern of means for imputed GPAs was similar to that for actual GPAs: humanities was highest (3.11) and natural science was lowest (2.73), but social science (2.80) was only slightly higher than natural science. And the median letter grades were identical to those for available GPAs. But, unlike the moderate correlations for available GPAs, the imputed GPAs correlated very highly (r = .89 to .95). The letter grades agreed moderately for humanities and natural science, and for humanities and social science (Kw = .24 and .34), but agreed highly for natural science and social science (Kw = .74), mirroring differences in grade levels for humanities and natural science and for humanities and social science, and similar distributions for natural science and social science.

TABLE 4

Intercorrelations Between Observed GPAs for Three Fields of Study

Field	*(1)*	*(2)*	*(3)*
(1) Humanities		.41	.47
(2) Natural science	.89		.48
(3) Social science	.90	.95	

Note: $N = 981$. Correlations for actual GPAs appear above the diagonal; correlations for imputed GPAs appear below the diagonal.

TABLE 5

Percentage Distributions of Observed Letter Grades for Three Fields of Study

	Actual			*Imputed*		
Letter Grade	*Humanities*	*Natural Science*	*Social Science*	*Humanities*	*Natural Science*	*Social Science*
A+						
A	12.6	10.8	13.0			
A–	24.0	8.6	11.9	6.6	1.7	6.6
B+	7.2	3.7	3.7	40.7	15.5	18.7
B	27.6	14.9	17.6	37.9	25.9	23.4
B–	16.2	12.8	17.2	12.9	28.1	27.0
C+	2.3	4.5	2.4	1.1	19.3	15.1
C	6.5	18.8	19.3	.7	6.5	5.5
C–	1.0	3.6	3.1		2.5	2.7
D	1.3	13.9	8.1		.4	1.0
F	1.1	8.6	3.7			

Note: $N = 981$. Percentages may not add to 100.0 because of rounding.

Divergences between the imputed and actual means for natural science (2.73 compared to 2.31) and social science (2.80 compared to 2.61) reflect adjustments in grading standards produced by imputing grades for all students in every department. The imputed means for these fields differed, although the imputed and actual means for individual departments were generally very similar, because students were not uniformly distributed across departments, i.e., Simpson's (1951) paradox. Substantively, students were clustered in courses in natural science and social science departments with strict grading standards. And the lack of differences between imputed and actual means for humanities (3.11 compared to 3.06) suggests that students are equally likely to take courses in humanities departments with strict and with lenient grading standards.

In short, the actual and imputed versions of observed GPA functioned similarly. The GPAs for the three fields of study correlated appreciably, especially the imputed GPAs, but they had markedly different means. Despite the similar rankings of the GPAs, the corresponding letter grades generally agreed much less because of level differences in grade distributions.

TABLE 6

Kappas Between Observed Letter Grades for Three Fields of Study

Field	*(1)*	*(2)*	*(3)*
(1) Humanities		.17	.25
(2) Natural science	.24		.31
(3) Social science	.34	.74	

Note: $N = 981$. Kappas for actual GPAs appear above the diagonal; Kappas for imputed GPAs appear below the diagonal.

TABLE 7

Means and Standard Deviations of Predicted GPAs for Three Fields of Study

	GPA			
	Actual		*Imputed*	
Field	*Mean*	S.D.	*Mean*	S.D.
Humanities	3.06	.31	3.11	.14
Natural science	2.31	.49	2.73	.22
Social science	2.61	.45	2.80	.25

Note: $N = 981$.

Correspondence Between Fields: Predicted GPAs

Means and standard deviations of the predicted standard GPAs for the three fields of study, based on actual and imputed GPAs, are presented in Table 7, and the intercorrelations between the GPAs for the fields are shown in Table 8. Distributions of the letter-grade versions of these GPAs appear in Table 9, and associations between the letter grades for the fields are reported in Table 10.

For predicted GPAs estimated from actual GPAs, the means for the fields were identical to the corresponding actual GPAs but were larger on a relative basis because the standard deviations were roughly half as great. The median letter grade was B for humanities, C+ for natural science, and B– for social science. The three predicted GPAs correlated highly (r = .84 to .90), but the letter grades agreed minimally for humanities and natural science (Kw = .06) and moderately for humanities and social science and for natural science and social science (Kw = .23 to .43), reflecting the level differences in grade distributions.

The results for predicted GPAs estimated from imputed GPAs were similar. The means were identical to the corresponding imputed GPAs, but were larger in relative terms because the standard deviations were halved. The median letter grade was B for humanities, B– for natural science, and B– for social science. The three predicted GPAs correlated highly (r = .98 to .99). The letter grades agreed minimally for humanities and natural science and for humanities and social science (Kw = .04 to .12), but agreed highly for natural science and social science (Kw = .72), reflecting similarities and differences in grade distributions in the three fields.

In sum, the predicted GPAs presented the same general picture as the observed GPAs: the rankings of the three fields of study were very similar, but the corresponding letter grades usually differed markedly because of divergent levels in grade distributions.

TABLE 8

Intercorrelations of Predicted GPAs for Three Fields of Study

Field	*(1)*	*(2)*	*(3)*
(1) Humanities		.84	.90
(2) Natural science	.98		.85
(3) Social science	.97	.99	

Note: $N = 981$. Correlations for predicted actual GPAs appear above the diagonal; correlations for predicted imputed GPAs appear below the diagonal.

TABLE 9

Percentage Distributions of Predicted Letter Grades for Three Fields of Study

	Actual			*Imputed*		
Letter Grade	*Humanities*	*Natural Science*	*Social Science*	*Humanities*	*Natural Science*	*Social Science*
A+						
A	.4		.1			
A–	7.5	.5	1.9	.1		.2
B+	32.8	3.9	11.9	39.8	3.2	9.4
B	35.2	9.6	17.2	57.8	26.8	33.7
B–	20.7	21.8	29.3	2.3	56.7	46.0
C+	3.3	27.3	24.9		12.8	10.1
C	.1	23.2	10.4		.5	.6
C–		9.2	3.7			
D		4.5	.6			
F						

Note: $N = 981$. Percentages may not add to 100.0 because of rounding.

Correspondence Between Predicted GPAs and Intended Majors

Distributions of the field of study with the highest GPA for each student, using predicted GPAs based on actual and imputed GPAs, and standard and percentile rank versions of these GPAs, are presented in Table 11. Cor-

TABLE 10

Kappas Between Predicted Letter Grades for Three Fields of Study

Field	*(1)*	*(2)*	*(3)*
(1) Humanities		.06	.23
(2) Natural science	.04		.43
(3) Social science	.12	.72	

Note: $N = 981$. Kappas for predicted actual GPAs appear above the diagonal; Kappas for predicted imputed GPAs appear below the diagonal.

TABLE 11

Percentage Distributions of Fields of Study with Highest Predicted GPA and Intended Major

	Standard GPA		*Percentile Rank GPA*		*Intended*
Field	*Actual*	*Imputed*	*Actual*	*Imputed*	*Major*
Total (N = 981)					
Humanities	98.4	99.7	32.9	34.4	9.9
Natural science			36.1	32.8	32.9
Social science	1.4	.3	30.4	31.3	55.3
Women (N = 579)					
Humanities	99.0	99.8	43.0	43.9	12.4
Natural science			27.1	21.2	27.0
Social science	.9	.2	29.0	33.9	58.7
Men (N = 402)					
Humanities	97.5	99.5	18.4	20.6	6.2
Natural science			49.0	49.5	41.5
Social science	2.2	.5	32.3	27.6	50.4
Whites (N = 761)					
Humanities	97.9	99.9	31.7	32.5	10.3
Natural science			37.1	34.3	31.7
Social science	1.8	.1	30.6	31.4	56.0
Others (N = 220)					
Humanities	100.0	99.1	37.3	40.9	8.3
Natural science			32.7	27.7	37.1
Social science		.9	29.5	30.9	53.2

Note: Percentages may not add up to 100.0 because of rounding, ties for predicted GPAs, or "Other" or "None" for intended majors.

TABLE 12

Kappas Between Field of Study with Highest Predicted GPA and Intended Major

	Standard GPA				*Percentile Rank GPA*			
	Actual		*Imputed*		*Actual*		*Imputed*	
Sample	N	K	N	K	N	K	N	K
Total	884	.00	885	.00	879	.03	873	.06
Women	522	−.01	523	.00	518	.02	519	.03
Men	362	.00	362	.00	361	.01	354	.07
Whites	682	−.01	683	.00	678	.05	672	.07
Others	202	.00	202	.00	201	−.03	201	.06

responding distributions of intended majors are also shown in this table. The associations between these fields and intended majors are reported in Table 12. The data in these tables are presented for the total sample and for sex and ethnic subgroups.

For standard versions of predicted GPAs estimated from actual GPAs, humanities was the field with the highest GPA for most students in the total sample (98.4 percent). In contrast, few students in the total sample chose this major (9.9 percent). The pattern was similar for the sex and ethnic subgroups. The field with the highest GPA and intended major agreed minimally for the total sample (K = .00). Agreement was also minimal for the sex and ethnic subgroups.

The corresponding results for the percentile rank versions of these GPAs were very different. The three fields were about evenly divided in the total sample. Natural science was the field with the highest GPA for the most students (36.1 percent), close to the proportion choosing that major (32.9 percent). The proportion of students with the highest GPA in a field and the proportion choosing the same major differed markedly for humanities (32.9 percent with the highest GPA compared to 9.9 percent with intended major) and social science (30.4 percent compared to 55.3 percent).

The percentile rank results were similar for the ethnic subgroups but differed for the sex subgroups. For women, humanities was the field with the highest GPA for a disproportionately large number of students (43.0 percent), well above the proportion choosing that major (12.4 percent). And the proportion of women with the highest GPA in a field and the proportion choosing the same major differed markedly for social science (29.0 percent compared to 58.7 percent). For men, natural science was the field with the highest GPA for a disproportionately large number of students (49.0 percent), close to the proportion choosing that major (41.5 percent), and humanities was the field with the highest GPA for a disproportionately small number of men (18.4 percent), close to the proportion choosing that major (6.2 percent). The proportion of men with the highest GPA in a field and the proportion choosing the same major differed markedly for social science (32.3 percent compared to 50.4 percent).

In these analyses of percentile ranks, the field with the highest GPA and intended major agreed minimally in the total sample (K = .03); agreement was minimal in the subsamples as well.

All the findings for predicted GPAs based on imputed GPAs were similar to those for predicted GPAs based on actual GPAs. Humanities was the field with the highest standard GPA for virtually all students in the total sample (99.7 percent). The field with the highest GPA and intended major agreed minimally in the total sample (K = .00). And the results were similar for the sex and ethnic subgroups. Furthermore, for the percentile rank versions of these GPAs, the three fields were approximately evenly divided. The pattern was similar for ethnic subgroups but diverged for sex subgroups, much as it did for predicted GPAs based on actual GPAs. And the field with the highest GPA and intended major agreed minimally for the total sample and the subgroups (K = .06 for the total sample).

In summary, the findings for predicted GPAs based on actual GPAs and predicted GPAs based on imputed

GPAs were very similar. Almost all students were predicted to receive their highest grades, in absolute terms, in the humanities, reflecting the higher level of the grade distributions in that field, although few students planned to major in the humanities. And the field in which students were predicted to receive their highest grades and their intended major did not agree at all.

Eliminating differences in fields of study due to differences in the levels of their grade distributions by comparing students' performance to that of other students in the same field drastically changed the predictions: although roughly the same proportions of students were predicted to receive their highest grades, in relative terms, in humanities, natural science, and social science, sharp differences were produced between women and men, with more women predicted to receive their highest grades in the humanities and more men predicted to receive their highest grades in natural science. But, again, the field of study in which students were predicted to receive their highest grades did not agree with their intended major. All these findings should be regarded cautiously because the differences in percentile ranks may be very unreliable.

Conclusions

A key finding of this study was that the SAT and other variables based on high school performance did predict college grades in different fields of study by taking into account marked variations in grade distributions among the fields. These predictions of letter grades could be potentially useful to students in making decisions about college courses and majors. Even though students are probably aware of the differences in grade distributions from field to field, they may still be unable to forecast their own grades in them.

This outcome of the study is attributable to variations in grade distributions from field to field combined with the general ability of the SAT and the other variables to predict academic performance. The rankings of observed grades in each field were very similar, as were the rankings of predicted grades, making it impossible to predict differences in students' performance from one field to the next, except for the crucial fact that the levels of the grade distributions varied from field to field. These results underscore the value of attending to absolute grade levels in making predictions of college performance.

Another important finding was that students' predicted grades in the different fields and their intended majors were virtually unrelated: the distributions of the fields with the highest predicted grades and intended majors were markedly different, and the two were minimally associated, regardless of whether the highest absolute grade or the highest relative grade was considered. Choice of major probably has many determinants besides anticipated grades, such as satisfaction and expected economic return (see the review by Rever, 1973), but this observed lack of association may reflect, at least in part, students' ignorance of their probable level of performance in different academic fields (Rohr & Ayers, 1973). Providing students with usable predictions of their performance in different fields would be expected to increase the congruence between the grades forecast and the majors chosen.

The extent to which students' course taking or selection of majors would actually be affected by feedback about their predicted college grades is unknown. Insofar as the grade predictions would have any effect, the present results suggest that the clearest impact would be to stimulate taking courses or majoring in the humanities and to discourage it in social science; the effects on natural science are less clear-cut, but it is evident that such grade predictions would not increase enrollments in that field. The findings of this study also imply that these impacts would be similar for white and other students, but might be magnified for the sexes, with dramatic increases in the number of women studying the humanities. It is apparent that the underrepresentation of women and minority students majoring in natural science would not be corrected (American College Testing Program, 1993; College Board, 1993; Dey, Astin, & Korn, 1991).

These initial results about the ability of the SAT and collateral variables to predict college grades are promising, but follow-up research is needed to address several important issues. First, the highly similar findings for actual and imputed GPA offer no basis for choosing between them; the latter is much more defensible on theoretical grounds but also far more complicated to obtain. Second, the differences in predicted grades, especially those employing percentile ranks, may be highly unreliable, as mentioned previously, and their stability requires evaluation. Third, cross validation is needed to establish the accuracy of the grade predictions, although the ratio of number of variables to number of students suggests that the shrinkage will be minimal. Fourth, the present prediction system could probably be substantially simplified, reducing the number of predictors involved. This study attempted to achieve maximum accuracy, without regard to the number of predictors used. Fifth, further attempts to identify a greater variety of fields of study are called for, beyond humanities, natural science, and social science. More highly differentiated predictions, if they prove to be feasible, would clearly be extremely valuable in aca-

demic guidance. Sixth, the ability to generalize the present findings beyond freshman grades and a single institution needs to be established. In particular, the accuracy of the predictions based on available GPAs may be seriously underestimated because of the unreliability of these GPAs, as noted earlier. All these proposed efforts to extend and verify the present findings could readily be accomplished by employing the larger data set for 45 colleges used by Ramist et al. (1990).

References

American College Testing Program. (1993). *The high school profile report, normative data.* Iowa City, IA: Author.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1984). *Technical supplement, Differential Aptitude Tests, forms V and W.* Cleveland, OH: Psychological Corporation.

Berryman, S. E. (1983). *Who will do science?* New York: Rockefeller Foundation.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Cohen, J. (1968). Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70,* 213–220.

Cohen, J., & Cohen, P. (1983). *Applied multiple correlation/regression analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

College Board. (1993). *College-bound seniors—1993 profile of SAT and Achievement Test takers.* New York: Author.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39,* 1–22.

Dey, E. L., Astin, A. W., & Korn, W. S. (1991). *The American freshman: Twenty-five year trends.* Los Angeles: University of California at Los Angeles, Higher Education Research Institute.

Frane, J. (1990). Description and estimation of missing data. In W. J. Dixon (Ed.), *BMDP statistical software manual* (pp. 813–829). Los Angeles, CA: BMDP.

French, J. W. (1963). Comparative prediction of college major-field grades by pure-factor aptitude, interest, and personality measures. *Educational and Psychological Measurement, 23,* 767–774.

French, J. W. (1964). Comparative prediction of high-school grades by pure-factor aptitude, information, and personality measures. *Educational and Psychological Measurement, 24,* 321–329.

Goldman, R. D., & Hewitt, B. N. (1975). Adaptation-level as an explanation for differential standards in college grading. *Journal of Educational Measurement, 12,* 149–161.

Hackett, G. (1985). Role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis. *Journal of Educational Psychology, 32,* 47–56.

Katz, M., & Norris, L. (1972). The contribution of academic interest measures to the differential prediction of marks. *Journal of Educational Measurement, 9,* 1–11.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Mollenkopf, W. G. (1950). Predicted differences and differences between predictions. *Psychometrika, 15,* 409–417.

National Science Foundation. (1990). *Women and minorities in science and engineering* (NSF 90-301). Washington, DC: Author.

Noeth, R. J. (1979). *Counselor's guide and technical summary, 1979-80, Washington Pre-College Test.* Seattle, WA: Washington Pre-College Testing Program.

Peterson, K. L., & Roscoe, B. (1983). Factors influencing selections of major by college females. *College Student Journal, 17,* 29–34.

Prather, J. E., Smith, G., & Kodras, J. E. (1979). A longitudinal study of grades in 144 undergraduate courses. *Research in Higher Education, 10,* 11–24.

Psychological Corporation. (1992). *Technical manual, Differential Aptitude Tests* (5th ed.). San Antonio, TX: Author.

Ramist, L., Lewis, C., & McCamley, L. (1990). Implications of using freshman GPA for the predictive validity of the SAT. In W. W. Willingham, C. Lewis, R. Morgan, & L. Ramist (Eds.), *Predicting college grades: An analysis of institutional trends over two decades* (pp. 253–288). Princeton, NJ: Educational Testing Service.

Rever, P. R. (1973). *Scientific and technical careers: Factors influencing development during the educational years* (Monograph 12). Iowa City, IA: American College Testing Program.

Rohr, M. E., & Ayers, J. B. (1973). Relationship of student grade expectations, selected characteristics, and academic performance. *Journal of Experimental Education, 41* (3), 58–62.

Simpson, E. H. (1951). The interpretation of interaction contingency tables. *Journal of the Royal Statistical Society,* Series B, *13,* 238–241.

Stricker, L. J., Rock, D. A., & Burton, N. W. (1991). *Sex differences in SAT predictions of college grades* (College Board Rep. No. 91-2; ETS Research Rep. No. 91-38). New York: College Entrance Examination Board.

Stricker, L. J., Rock, D. A., & Burton, N. W. (1993). Sex differences in predictions of college grades from Scholastic Aptitude Test scores. *Journal of Educational Psychology, 85,* 710–718.

Stricker, L. J., Rock, D. A., Burton, N. W., Muraki, E., & Jirele, T. J. (1994). Adjusting college grade point average criteria for variations in grading standards: A comparison of methods. *Journal of Applied Psychology, 79,* 178–183.

Willingham, W. W., & Breland, H. M. (1982). *Personal qualities and college admissions.* New York: College Entrance Examination Board.

Appendix

TABLE A-1

Standardized Regression Coefficients for Predicting Actual and Imputed GPAs for Three Fields of Study

	Actual GPA			*Imputed GPA*		
Predictor	*Humanities*	*Natural Science*	*Social Science*	*Humanities*	*Natural Science*	*Social Science*
SAT-Verbal	.12	.08	.17	.12	.13	.17
SAT-Mathematical	.07	.22	.06	.13	.19	.13
High school rank	.19	.18	.16	.24	.23	.24
Percentage of college-bound seniors	.08	.08	.05	.09	.09	.09
Years of high school courses in:						
Arts and music	.03	−.08	.03	−.03	−.02	.01
English	.08	.09	.06	.11	.09	.10
Foreign and classical languages	.05	−.01	.04	.03	.01	.02
Mathematics	.00	−.09	−.09	−.07	−.07	−.07
Natural science	−.07	−.07	.02	−.06	−.05	−.04
Social science and history	−.01	−.02	.04	.00	.01	.01
High school honors course in:						
Arts and music	.03	.04	.01	.06	.04	.03
English	.11	.00	.09	.08	.06	.09
Foreign and classical languages	−.04	.00	.01	−.02	.01	.01
Mathematics	−.10	−.03	.01	−.06	−.04	−.05
Natural science	−.04	−.01	−.04	−.03	−.02	−.04
Social science and history	.00	.01	.00	.01	.01	.01
Average grade for high school courses in:						
Arts and music	−.04	.00	−.01	−.03	−.01	−.02
English	.09	.05	.05	.07	.07	.07
Foreign and classical languages	.07	.03	−.03	.06	.04	.02
Mathematics	−.03	−.01	.03	.01	.02	.02
Natural science	.04	.09	.05	.06	.06	.06
Social science and history	.06	.02	.11	.05	.04	.06
Average grade available for high school courses in arts and music	.03	−.05	.02	.00	−.01	.01