DOCUMENT RESUME

ED 446 593 HE 033 433

AUTHOR	Bridgeman, Brent; McCamley-Jenkins, Laura; Ervin, Nancy
TITLE	Predictions of Freshman Grade-Point Average from the Revised and Recentered SAT[R] I: Reasoning Test. College Board Research Report.
INSTITUTION	College Entrance Examination Board, New York, NY.
REPORT NO	ETS-RR-00-1
PUB DATE	2000-00-00
NOTE	20p.
AVAILABLE FROM	College Board Publications, Box 886, New York, NY 10101-0886 (Item #987443; $15, plus $4 postage and handling). Tel: 800-323-7155 (Toll free).
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; College Entrance Examinations; *College Freshmen; *Grade Point Average; Higher Education; *Prediction; *Scoring; Validity
IDENTIFIERS	*Scholastic Assessment Tests

ABSTRACT

The impact of revisions in the content of the Scholastic Assessment Test (SAT) and changes in the score scale on the predictive validity of the SAT were examined. Predictions of freshman grade-point average (FGPA) for the entering class of 1994 (who had taken the old SAT) were compared with predictions for the class of 1995 (who had taken the new SAT I: Reasoning Test). The 1995 scores were evaluated both on the original SAT Program scale and on the recentered scale introduced that year. The changes in the test content and recentering of the score scale had virtually no impact on predictive validity. Other analyses indicated that the SAT I predicts FGPA about equally well across different ethnic groups. Correlations were slightly higher for higher levels of parental education and family income, and grades were more predictable for students with intended majors in math/science (mathematics, engineering, and biological or physical sciences) than for students with other intended majors. Correlations of the SAT I and the composite of SAT I scores and high school grade point average (HSGPA) with FGPA were generally higher for women than for men, although this pattern was reversed at college with very high mean SAT I scores. When a single prediction equation was used for all students, men tended to get lower grades than predicted, and women got higher grades than predicted. African American and Hispanic/Latino men received lower grades than predicted, but women in these groups performed as predicted by the composite. Both men and women with intended majors in math/science got lower grades than would be predicted by an equation based on scores for all enrolled students. (Contains 13 tables, 7 figures, and 18 references.) (Author)

College Board Research Report No. 2000-1
ETS RR No. 00-1

Predictions of Freshman Grade-Point Average From the Revised and Recentered SAT® I: Reasoning Test

Brent Bridgeman, Laura McCamley-Jenkins, and Nancy Ervin

College Entrance Examination Board, New York, 2000

Acknowledgments

The authors thank Rick Morgan, Lawrence Stricker, and Cathy Wendler for very useful comments on an earlier draft of this paper. The editorial assistance of Erin Thomas was also appreciated. Special thanks to the staff at the cooperating colleges who collected and organized the freshman grade data.

Brent Bridgeman is a principal research scientist at ETS.
Laura McCamley-Jenkins is a senior research data analyst at ETS.
Nancy Ervin is a program administrator at ETS.

Additional copies of this report (item #987443) may be obtained from College Board Publications, Box 886, New York, New York 10101-0886, (800) 323-7155. The price is $15. Please include $4 for postage and handling.

Printed in the United States of America.

Contents

Abstract

The impact of revisions in the content of the SAT® and changes in the score scale on the predictive validity of the SAT were examined. Predictions of freshman grade-point average (FGPA) for the entering class of 1994 (who had taken the old SAT) were compared with predictions for the class of 1995 (who had taken the new SAT I: Reasoning Test). The 1995 scores were evaluated both on the original SAT Program scale and on the recentered scale introduced that year. The changes in the test content and recentering of the score scale had virtually no impact on predictive validity. Other analyses indicated that the SAT I predicts FGPA about equally well across different ethnic groups. Correlations were slightly higher for higher levels of parental education and family income, and grades were more predictable for students with intended majors in math/science (mathematics, engineering, and biological or physical sciences) than for students with other intended majors. Correlations of the SAT I and the composite of SAT I scores and high school grade-point average (HSGPA) with FGPA were generally higher for women than for men, although this pattern was reversed at colleges with very high mean SAT I scores. When a single prediction equation was used for all students, men tended to get lower grades than predicted and women got higher grades than predicted. African-American and Hispanic/Latino men received lower grades than predicted, but women in these groups performed as predicted by the composite. Both men and women with intended majors in math/science got lower grades than would be predicted by an equation based on scores for all enrolled students.

Key words: SAT validity, over/underprediction, grade adjustments, recentering, gender and ethnic differences

Introduction

A revised SAT, renamed SAT I: Reasoning Test (SAT I), was first offered in March of 1994. The new test continues to assess verbal and mathematical reasoning and uses the same 200–800 score scale as the test it replaced. There were a number of new features introduced in the revised test. The verbal (SAT I–V) section of the test was lengthened by 15 minutes to accommodate more emphasis on critical reading and to allow time for more students to complete all of the questions. A set of questions based on a pair of related passages was added to allow questions that compare and contrast different styles or points of view. Verbal analogies and sentence completion questions were retained from the old test, but antonym questions were dropped. There was an increased emphasis on assessing vocabulary in context. The verbal section now contains 78 questions that are to be answered in 75 minutes; more than half of these questions are based on reading passages. The mathematical (SAT I–M) section was also lengthened by 15 minutes, and a new question type was introduced. The new question type required examinees to enter numerical answers on a special grid. The 60 questions of the revised mathematical section contain 50 multiple-choice questions and 10 of the special grid type. For the new test, examinees are allowed to use their own calculators. Allowing calculators did not result in any major changes in the question specifications, and computations were kept simple so that calculators, while allowed, would not be required.

Over the years, as the population of students taking the SAT changed, average scores were no longer in the middle of the 200–800 score scale. In the 1995 profile of college-bound seniors (College Board/Educational Testing Service, 1995), the mean verbal score was 428 on the original SAT Program scale and the mean mathematical score was 482. Because verbal scores had changed more over time than had mathematical scores, students could easily misinterpret their relative standing in these two academic areas. For example, a student with a 470 on both sections would be well above average on the verbal scale but slightly below average on the mathematical scale. In response to this, a change was made to the SAT I score scale in April of 1995. Scores were realigned so that the middle of both the verbal and mathematical scales would once again be about 500. Recentering affected only the way scores were reported; questions on the recentered test were neither harder nor easier than questions on the test prior to recentering. Scores for the same test could be reported on the original or recentered scales.

The current study was undertaken to determine the effects of the content changes and score recentering on the ability of the SAT I: Reasoning Test to predict freshman grade-point average (FGPA) and to address more generally the factors that may affect correlations of frequently used admission measures with freshman grades. Colleges use SAT I scores as a supplement to other information, notably the high school grade-point average (HSGPA), to make selection decisions. The SAT I and HSGPA were examined both individually and combined as predictors of college grades. The study recognizes that freshman grades are only one indicator of success in college and that much can be gained from considering a broader perspective (Willingham, 1985); nevertheless, the

FGPA is an important indicator because it reflects a cumulative judgment of the quality of college-level academic performance made by a number of faculty members in several different disciplines. Although the four-year average might be a preferable criterion, research reviews suggest that there is little or no difference in the size of validity coefficients based on FGPA and those based on the cumulative four-year average (Burton & Ramist, in preparation; Wilson, 1983).

Because students select colleges and colleges select students, the range of SAT scores and HSGPAs found among the enrolled students at a particular college can be much narrower than the range found in the potential applicant population. This restriction in range tends to reduce correlations with FGPA that can be computed only for enrolled students; the real question of interest is how well the scores predict for potential applicants, not for enrolled students. Therefore, correlations were adjusted to estimate what they would have been if the ranges of SAT I scores and HSGPAs were the same for a given college and the full national cohort of college-bound seniors taking the SAT I.

The impact of the change in test content was assessed by comparing data for students from the entering college freshman class of 1994 who had taken the old SAT, with data for students from the entering freshman class of 1995 who had taken the new SAT I. Although cohort effects were necessarily confounded with effects related to changes in the test content, previous research suggests that changes in correlations with FGPA over a single year tend to be quite small, especially when correlations are averaged over several colleges (Willingham, Lewis, Morgan, & Ramist, 1990). A change of .02 or less in the multiple correlation (predicting FGPA from a combination of SAT I–V, SAT I–M, and HSGPA) would be in the range expected for year-to-year fluctuations; a change of more than .02 could reflect a real difference related to changes in test content. For analyses of the effects of recentering, there was no confounding with cohort differences because only the 1995 cohort was used. Raw scores on the same tests were converted to both the original or recentered score scales; thus, the score of each student in the 1995 cohort was available on both the original SAT Program scale and on the recentered scale.

An additional question of interest was whether the changes in test content and test scale had an impact on over/underprediction, that is, the tendency for predictions based on the total group to be either too high or too low for specific subgroups. Overprediction occurs when a subgroup does not perform as well as predicted; its predicted performance is above, or over, its actual performance. A common finding is that college grades of women are underpredicted and grades of ethnic minorities are overpredicted (Breland, 1979; Linn, 1978; Ramist, Lewis, & McCamley-Jenkins, 1994; Sawyer, 1986). The referenced studies all examined gender and ethnic groups separately rather than examining gender effects within ethnic groups. This leaves open the question of whether FGPA for women from minority groups is over- or underpredicted. A study of African-American and Hispanic/Latina women in three colleges suggested that their scores were slightly underpredicted by SAT scores when the predictions were based on males from the same ethnic group (Pennock-Román, 1994). However, Pennock-Román did not evaluate over/underprediction within gender/ethnic groups when the original predictions were based on the regression for all students.

Method

Sample

Data for both the 1994 and 1995 entering classes were provided by 23 colleges. Because the same colleges provided data in both years, we could be more confident that any validity changes were attributable to changes in the test content rather than changes in the mix of colleges providing data. The colleges in the sample represented a combination of public and private institutions (13 public and 10 private), including one junior college. One college had only female students. Each of the six College Board geographical regions was represented, but the sample should not be considered to be a nationally representative sample in a strict sampling sense. In particular, most of the colleges were well above average in selectivity and had relatively high scores in their freshman classes. Seven colleges had average combined SAT I scores above 1250, and only two colleges had average scores below 1000. In the sample, average scores on the recentered SAT Program scale were 566 verbal and 581 mathematical compared to 504 and 506 for all college-bound seniors in 1995 (College Board/ Educational Testing Service, 1995). The few students in the 1995 cohort who had scores only from the old SAT were removed from the analyses.

Variables

Colleges were asked to provide the FGPA for all students in the 1994 and 1995 entering classes. In addition, they were asked to provide grades in individual courses, but only seven colleges sent this course-level information. SAT I scores were extracted from SAT Program files at the Educational Testing Service (ETS). If the student had taken the SAT I more than once, the most recent score was used. Demographic information was obtained from the Student Descriptive Questionnaire (SDQ), which about 95 percent of the students voluntarily complete when they register to take tests in the SAT Program. The self-reported HSGPA was also obtained from the SDQ. This HSGPA contains 12 categories from F through A+. This HSGPA was coded such that an F = 0, D- = .7, D = 1.0, D+ = 1.3,... A+ = 4.3. FGPA was similarly coded from 0 to 4.3, though 4.0 was the top score for many colleges that did not use A+ grades. Previous research suggests that using the self-reported HSGPA from the SDQ results in multiple correlations (combining SAT scores and HSGPA to predict FGPA) that are about .03 to .04 points smaller than multiple correlations that use the actual school-reported HSGPA (Freeberg, Rock, & Pollack, 1989).

Procedures

With a few exceptions as noted, correlations of predictors with FGPA were corrected for range restriction with the Pearson-Lawley multivariate correction (Gulliksen, 1950, pp. 165-166). This adjustment requires the national standard deviations (SD) for the predictors as well as their intercorrelations. For the original SAT Program scale, these SDs were as follows: SAT I–V, 112; SAT I–M, 124. For the recentered scale the SDs were: SAT I–V, 110; SAT I–M, 111. The SD for HSGPA was 0.66. Correlations were the same for original and recentered scales and were as follows: SAT I–V with SAT I–M, .71; SAT I–V with HSGPA, .48; SAT I–M with HSGPA, .53. Consistent with the approach taken by Ramist, Lewis, and McCamley-Jenkins (1994), we made coefficients more comparable across colleges, gender categories, college major categories, and ethnic groups by applying a single correction, based on the SDs in the full national group taking the SAT, to each subgroup.

All correlations with FGPA were computed within colleges, weighted by the number of students at that college (or in the relevant subgroup in that college), and averaged across colleges. Similarly, multiple correlations that used more than one predictor were computed within college and then the weighted average taken across colleges. If any predictor in the multiple correlations had a negative weight, the multiple correlation was recomputed with that variable removed.

Over/underprediction was analyzed by making predictions based on all students in a college and then, for each gender within an ethnic subgroup, computing the difference between the predicted and actual FGPA (predicted GPA minus actual GPA). The result is in grade-point units, with positive values indicating overprediction and negative values indicating underprediction. Two colleges were excluded from the averages—one had only female students and the other used a 0–15 scale for FGPA rather than the 0–4 (or 0–4.3) scale used at the other colleges. Procedures that are specific to particular analyses are presented with the results and discussion for those analyses.

Results and Discussion

Correlations of FGPA with the old (1994) SAT original scale, new (1995) SAT I original scale, and SAT I recentered scale are presented in Table 1. For the purpose of comparing predictions from old (1994) and new (1995) content and original and revised score scales, there was no need to adjust for range restriction, and the correlations in Table 1 were not adjusted. (The extent to which unadjusted correlations underrepresent the actual relationship of scores to grades may be seen by comparing the unadjusted correlations in Table 1 with the adjusted correlations in the *Total* column of Table 4.) Table 1 presents results separately for males and females and for gender within each of the four major ethnic groups. Note that the number of students in the *Gender Total* category is not simply the sum of the four ethnic categories shown; it also includes students from other ethnic categories and students who did not specify their ethnicity.

A quick glance down the first column of correlations suggests that neither the content changes nor recentering had much impact on predictive validity. This interpretation is reinforced in Table 2, indicating that the change from the old test to the new recentered test resulted in about as many colleges showing increases as decreases. Because no change was made in the way HSGPA was computed, differences noted for HSGPA presumably reflect cohort effects from 1994 to 1995. As noted previously (e.g., Ramist, Lewis, McCamley-Jenkins, 1994), correlations were generally higher for women than for men, but the impact of changes in the test and score scale on predictive validity appears to be equally trivial for men and women.

Predictions Within Gender and Ethnic Groups

Because of the somewhat smaller sample sizes, some year-to-year fluctuations in the gender-within-ethnic-group correlations should be expected. Changes from 1994 to 1995 in correlations of HSGPA with FGPA indicate only cohort effects. As indicated in the uncorrected correlations in Table 1, shifts of .05 or greater were found in 3 of the 8 gender/ethnic groups (two

TABLE 1

Correlations With FGPA for SAT (1994), SAT I (1995) on Original Scale, and Recentered SAT I

			Correlation With Freshman Grade-Point Average									
			Gender Total		*African American*		*Asian American*		*Hispanic/Latino*		*White*	
Score	*Version*	*Total*	M	F	M	F	M	F	M	F	M	F
Number of Students	old (1994)	45,100	21,733	23,367	1,074	1,761	3,767	4,047	1,400	1,825	14,545	14,607
	new (1995)	48,039	23,925	25,114	1,148	1,826	3,778	4,087	1,517	1,934	15,249	15,920
SAT–V	old	.29	.28	.32	.28	.29	.23	.27	.21	.29	.27	.29
	new, OS	.30	.28	.32	.23	.29	.24	.26	.19	.29	.28	.30
	new, R	.30	.28	.32	.23	.29	.24	.26	.19	.29	.28	.30
SAT–M	old	.29	.30	.34	.29	.33	.28	.33	.25	.25	.27	.30
	new, OS	.30	.32	.35	.31	.34	.32	.31	.19	.31	.30	.31
	new, R	.30	.32	.35	.30	.34	.32	.32	.19	.31	.30	.31
HSGPA	old	.36	.37	.35	.29	.34	.32	.28	.32	.34	.37	.35
	new	.36	.37	.34	.34	.29	.28	.26	.30	.29	.38	.34
SAT	old	.34	.34	.39	.35	.38	.32	.38	.30	.34	.32	.34
	new, OS	.35	.35	.39	.35	.37	.36	.37	.24	.37	.33	.35
	new, R	.35	.35	.39	.34	.37	.36	.37	.24	.37	.33	.35
SAT + HSGPA	old	.45	.45	.46	.43	.47	.43	.43	.42	.44	.44	.43
	new, OS	.44	.44	.45	.45	.44	.44	.42	.38	.44	.43	.43
	new, R	.44	.44	.45	.45	.44	.44	.43	.38	.44	.44	.43

Note: OS = original scale; R = recentered scale. For HSGPA, old and new refer only to 1994 and 1995 samples respectively, not to content changes.

down and one up). The only single test score correlation showing a change greater than .05 was SAT–M in the Hispanic/Latino sample, with a decline of .06 for Hispanic/Latino men and an increase of .06 for Hispanic/Latina women. The SAT combined score was relatively stable; the only change greater than .05 was an apparent decline for Hispanic/Latino men of .06, from .30 to .24. Because the SAT combined correlation increased by .03 for Hispanic/Latina women, it would be hard to make an argument that the new content was somehow less predictive for Hispanic/Latino students generally. Adjusting the values in Table 1 for range restriction still showed essentially the same decline for Hispanic/Latino men (from .50 to .43). However, there was not a uniform decline across colleges. In the 8 colleges with at least 50 Hispanic/Latino men in both years, the correlation declined in 5 but increased in the other 3. Two colleges in the sample had large Hispanic/Latino populations (more than 200 students in both 1994 and 1995). In one of these colleges, a large state university in the Southwest, the adjusted correlation for Hispanic/Latino men declined from .57 to .50. However, in the other college, a large state university in the West, the adjusted correlation increased from .37 to .46. Thus, there does not appear to be a uniform declining trend or increasing trend in any subgroup, and the observed differences in the mean correlations may simply reflect year-to-year variation.

TABLE 2

Number of Colleges With Validity Increase or Decrease From SAT (1994)/Original Scale to SAT I (1995)/Recentered Scale

	Verbal	*Math*	*HSGPA*	*SAT + HSGPA*
# Colleges With Validity Increase	11	13	8	9
# Colleges With Validity Decrease	10	10	10	13
# Colleges Validity Unchanged	2	0	5	1

Correlations for Colleges in Three Score Ranges

Correlations for colleges in three score ranges are shown in Table 3. The score ranges were based on the mean SAT I combined score in the college, using the recentered score scale. The first category included colleges with mean scores of less than 1050, the middle category range was 1050–1250, and the colleges in the highest category had mean SAT combined scores above 1250. The correlations were adjusted for restriction in range. Such adjustments are especially useful when comparisons are being made across categories in which there is more range restriction in one category than another; there is typically greater restriction in the highest score category because colleges in this category tend to be most selective. Within any one score category, the changes in the test and score scale had little or no impact on validity. A

Table 3

Adjusted Correlations With Freshman GPA for Colleges in Three Score Ranges

Score	Score Range (V + M)	Version: Old (1994)	New (1995), OS	New, R
SAT–V	<1050	.47	.46	.46
	1050–1250	.46	.45	.46
	>1250	.57	.54	.55
SAT–M	<1050	.47	.45	.46
	1050–1250	.47	.46	.47
	>1250	.58	.57	.57
HSGPA	<1050	.58	.58	.58
	1050–1250	.55	.52	.53
	>1250	.59	.60	.60
SAT	<1050	.51	.50	.50
	1050–1250	.50	.49	.50
	>1250	.63	.60	.61
SAT + HSGPA	<1050	.62	.62	.62
	1050–1250	.60	.58	.59
	>1250	.69	.69	.69

Note: Five colleges and 4,490 students in <1050 category; 11 colleges and 37,033 students in 1050–1250 category; 7 colleges and 6,516 students in >1250 category. OS = original scale; R = recentered scale.

previous study of score recentering (Morgan, 1994), suggested that any impact on predictive validity should be quite small, but whatever effects are noted should be positive; Table 3 supports that conclusion.

Consistent with previous findings (Ramist, Lewis, and McCamley-Jenkins, 1994), correlations tended to be higher for the institutions in the highest score category. In addition, the SAT increment, that is, the extent to which SAT scores improve predictions over HSGPA alone, tended to be greatest for colleges in the highest category. For the recentered scores on the new test, the SAT increment was .04 in the lowest category, .06 in the middle category, and .09 in the highest category.

Table 4 shows the ethnic/gender breakdown for the score categories in Table 3. Because the old test, new test, and recentered scale correlations were all virtually the same, only the correlations for the recentered scores on the new test are included in the table. The pattern of higher correlations in the colleges in the highest score category was not replicated in all groups. Note, for example, that the SAT I combined correlation for the highest category colleges was no higher than for the lowest category colleges for African-American, Asian-American, and Hispanic/Latina females and for African-American males

Table 4

Adjusted Correlations of Recentered Scores With Freshman Grade-Point Average for Colleges in Three Score Categories

Score	Score Category (V+M)	Total	Gender Total M	Gender Total F	African American M	African American F	Asian American M	Asian American F	Hispanic/Latino M	Hispanic/Latino F	White M	White F
Number of Students	<1050	4,490	2,001	2,489	269	320	76	68	170	268	1,398	1,733
	1050–1250	37,033	17,589	19,444	697	1,305	3,560	3,550	1,220	1,556	11,448	12,063
	>1250	6,516	3,335	3,181	182	201	442	469	127	110	2,403	2,174
SAT I–V	<1050	.46	.43	.50	.48	.49	.42	.58	.48	.49	.42	.49
	1050–1250	.46	.44	.49	.32	.48	.50	.50	.39	.48	.42	.47
	>1250	.55	.56	.55	.43	.46	.54	.55	.46	.33	.55	.50
	Total	.47	.46	.50	.37	.48	.50	.50	.39	.46	.44	.48
SAT I–M	<1050	.46	.41	.53	.52	.21	.38	.66	.52	.51	.39	.52
	1050–1250	.47	.47	.51	.40	.52	.57	.54	.40	.50	.44	.48
	>1250	.57	.59	.57	.46	.52	.58	.56	.56	.29	.59	.53
	Total	.48	.49	.52	.44	.52	.57	.55	.40	.48	.46	.49
HSGPA	<1050	.58	.54	.60	.47	.53	.38	.66	.47	.53	.55	.61
	1050–1250	.53	.52	.52	.45	.47	.55	.52	.48	.47	.51	.51
	>1250	.60	.60	.58	.49	.51	.54	.49	.62	.38	.61	.58
	Total	.54	.53	.53	.46	.49	.55	.52	.49	.48	.53	.53
SAT I	<1050	.50	.46	.56	.55	.55	.47	.68	.55	.55	.44	.55
	1050–1250	.50	.50	.55	.42	.55	.59	.57	.43	.55	.47	.51
	>1250	.61	.63	.61	.53	.54	.61	.61	.59	.41	.67	.56
	Total	.52	.51	.56	.47	.55	.59	.58	.44	.53	.49	.53
SAT I + HSGPA	<1050	.62	.58	.66	.59	.62	.55	.77	.59	.62	.58	.66
	1050–1250	.59	.58	.61	.52	.60	.66	.63	.54	.60	.56	.59
	>1250	.69	.70	.68	.61	.61	.67	.67	.72	.66	.70	.65
	Total	.61	.60	.62	.55	.61	.66	.64	.55	.61	.58	.60

as well. In the relatively large white student population sample, the correlations in the colleges in the highest category were much higher for males (.23 and .20 higher compared to the low and middle groups, respectively) but only marginally higher for females (.01 and .05, respectively). As indicated in Figure 1, correlations for the combined ethnic groups suggest that the SAT I combined score is a better predictor for women than for men at the colleges in the lowest score category, but that it predicts FGPA equally well for men and women at the colleges in the highest category. These data are consistent with the argument that behaviors unrelated to the developed abilities measured by the SAT, such as failing to attend class or complete assignments on time, may be more common in males and therefore make male grades more difficult to predict (Stricker, Rock, & Burton, 1993). Because males at most colleges in the highest score category may be as likely as females to attend class and complete assignments, tested abilities should be equally valid for men and women at these institutions.

For the SAT I and HSGPA composite, the same pattern seen for SAT I scores alone was repeated—grades of females were predicted more accurately in the colleges with relatively low mean test scores, but grades of males were predicted more accurately at the colleges with high mean test scores. This pattern was especially evident in the sample of white examinees. In the colleges in the lowest category, the correlation was higher for females by .08, but in the colleges in the highest category, the correlation was .05 higher for males. The previous Ramist et al. (1994) study found a similar pattern with higher correlations for women at the colleges with lower mean test scores, and a very small male advantage at the schools in their highest category; they did not provide information on gender within ethnic groups within mean score categories. The same pattern could also be observed in the uncorrected correlations for the colleges studied by Pennock-Román (1994). The two colleges in

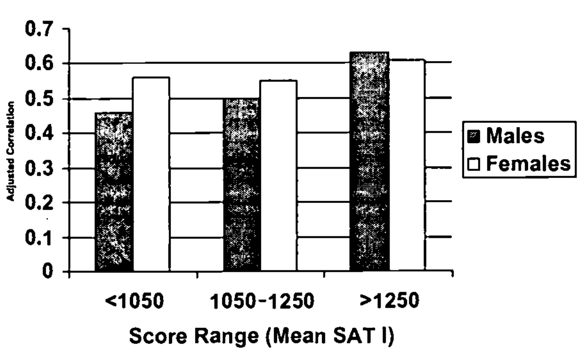

Figure 1. Adjusted correlations of SAT I scores with FGPA for males and females in colleges in three score ranges.

her sample that would be classified in the highest category in our sample both showed higher correlations for white males than for white females. In the two colleges with lower mean scores, the pattern was reversed.

Correlations in Individual Colleges

Means and standard deviations of the predictor and criterion scores for each college in the sample are presented in Table 5 along with the adjusted correlations for HSGPA, the SAT I combined, and the SAT I and HSGPA composite. The colleges are arranged in ascending order of the total SAT I combined mean score. The FGPA for the last college listed was on a 0–15 scale; all other colleges were on a 0–4.3 scale. The last column of the table indicates the range in the ability of the SAT I and HSGPA composite to predict freshman grades. The lowest adjusted correlation, and the only correlation in the .40s, was .47; there were four correlations in the .50s, twelve in the .60s, and six in the .70s.

Over- and Underpredictions

Table 6 presents the over- and underpredictions of FGPA. Results for the new content, original scale and new content, recentered scale were virtually identical, so for this analysis only the results for the recentered scale are presented for the new test (SAT I). For the full sample (across ethnic and gender groups), the over/underprediction was exactly the same for the new test as for the old. For the smaller gender within ethnic group breakdowns, differences were what would be expected from sample variations. The greatest differences were in the samples of African-American female students, with apparently less overprediction, or a shift from overprediction to underprediction, with the new test. However, because the shift in predictions from HSGPA alone was in the same direction and about as large, the change seems to reflect sample differences and not test content differences.

Consistent with previous findings (e.g., Pennock-Román, 1994; Ramist et al., 1994), there was a modest underprediction of women's grades and the complimentary overprediction of men's grades. As indicated in Figure 2, for the three ethnic minority groups studied, there was virtually no over- or underprediction of women's grades from the combination of SAT I scores and HSGPA. There was moderate to substantial overprediction of men's grades. For all of the three minority groups, but especially for the African-American and Hispanic/Latino groups, there was overprediction of grades for men, i.e., men did not perform as well in college as would be expected from their high school grades

TABLE 5

Means, Standard Deviations, and Adjusted Correlations With FGPA for Recentered SAT I Scores

College Type/Region	SAT I Mean	N	SAT I–V Mean	SAT I–V SD	SAT I–M Mean	SAT I–M SD	HSGPA Mean	HSGPA SD	FGPA Mean	FGPA SD	Adjusted Correlations HSGPA	Adjusted Correlations V+M	Adjusted Correlations V+M+H
Public/Southwest	850	468	426	85	424	85	3.10	0.58	2.24	0.92	0.53	0.58	0.63
Private Jr. Col./South	962	210	487	78	475	69	3.11	0.53	2.28	0.84	0.68	0.60	0.73
Public/Southwest	1009	1759	506	74	503	71	3.48	0.43	2.45	0.84	0.60	0.49	0.63
Private/Southwest	1010	264	510	81	500	76	3.37	0.53	2.69	0.84	0.49	0.49	0.56
Public/West	1036	1789	515	83	521	87	3.37	0.52	2.77	0.71	0.57	0.47	0.61
Public/Middle States	1063	10862	527	93	536	99	3.38	0.57	2.64	0.83	0.47	0.40	0.50
Public/Middle States	1073	1254	522	85	550	83	3.43	0.41	2.60	0.79	0.50	0.48	0.56
Public/New England	1079	2047	535	87	544	86	3.21	0.47	2.55	0.77	0.49	0.48	0.55
Public/West	1098	2327	524	87	574	84	3.61	0.42	2.75	0.64	0.50	0.54	0.60
Private/South	1106	2186	548	77	558	80	3.47	0.53	2.59	0.87	0.62	0.53	0.66
Public/West	1155	2532	561	92	594	81	3.72	0.40	2.82	0.65	0.54	0.54	0.61
Public/South	1201	2824	601	83	600	80	3.94	0.35	2.84	0.65	0.63	0.65	0.73
Public/West	1204	2602	582	88	622	77	3.82	0.36	2.95	0.54	0.55	0.55	0.63
Public/Southwest	1207	5146	595	83	612	80	3.78	0.43	2.86	0.81	0.57	0.57	0.65
Private/Middle States	1207	2280	606	88	601	83	3.59	0.46	3.10	0.69	0.44	0.37	0.47
Public/West	1209	2973	590	81	618	83	3.84	0.39	3.01	0.58	0.59	0.63	0.70
Private/New England	1264	394	639	65	624	68	3.61	0.38	2.95	0.63	0.62	0.55	0.67
Public/South	1265	1616	611	75	653	69	3.81	0.41	2.78	0.67	0.68	0.67	0.77
Private/South	1277	1193	632	71	645	69	3.79	0.43	2.84	0.73	0.55	0.52	0.61
Private Women/Mid. St.	1280	355	650	70	630	58	3.71	0.34	3.24	0.46	0.52	0.56	0.62
Private/New England	1282	319	643	76	638	69	3.74	0.37	3.06	0.53	0.65	0.59	0.71
Private/Midwest	1325	1516	656	69	669	74	3.95	0.34	3.21	0.49	0.55	0.55	0.62
Private/New England	1456	1123	731	64	726	61	4.09	0.25	12.28	1.73	0.62	0.72	0.77

and/or SAT I scores. Note that in these groups the overprediction was as great for HSGPA by itself as for the SAT I by itself. In the most highly selective colleges, the underprediction of women's grades from the SAT I and HSGPA composite was slightly less, ranging from -.04 to -.05.

TABLE 6

Over- (+) and Underprediction (-) of FGPA

Score	Version	Gender Total M	Gender Total F	African American M	African American F	Asian American M	Asian American F	Hispanic/Latino M	Hispanic/Latino F	White M	White F
Number of Students	Old (1994)	21,100	22,830	1,024	1,707	3,646	3,923	1,358	1,787	14,199	14,361
Number of Students	New (1995), R	22,327	24,589	1,121	1,770	3,675	3,994	1,480	1,903	14,887	15,713
SAT–V	Old	+.08	-.07	+.23	+.13	+.01	-.09	+.19	+.09	+.07	-.12
	New, R	+.08	-.07	+.26	+.06	+.01	-.07	+.20	+.09	+.07	-.11
SAT–M	Old	+.12	-.12	+.24	+.05	+.13	-.05	+.21	+.04	+.11	-.18
	New, R	+.13	-.12	+.26	-.02	+.14	-.03	+.22	+.02	+.11	-.17
HSGPA	Old	+.04	-.04	+.20	+.18	+.04	-.02	+.20	+.17	+.02	-.10
	New	+.04	-.04	+.22	+.12	+.04	-.01	+.20	+.15	+.01	-.09
SAT	Old	+.11	-.10	+.20	+.04	+.08	-.08	+.17	+.02	+.11	-.15
	New, R	+.11	-.10	+.22	-.03	+.09	-.06	+.19	+.01	+.11	-.14
SAT + HSGPA	Old	+.08	-.07	+.13	+.04	+.07	-.05	+.14	+.04	+.07	-.11
	New, R	+.08	-.07	+.14	-.01	+.07	-.03	+.15	+.02	+.07	-.09

Note: For HSGPA, version refers only to sample (Old = 1994; New = 1995). R = recentered.

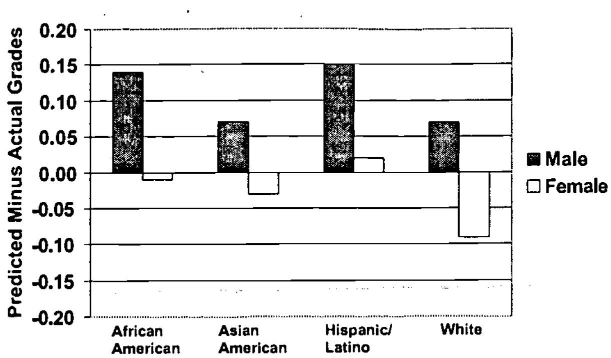

Figure 2. Over- (+) and underprediction (-) of FGPA from SAT I and HSGPA composite for four ethnic groups.

Correlations Within Parental Education and Income Categories

Socioeconomic categories, such as the highest educational degree earned by either parent, interact with ethnic categories in a way that makes it difficult to attribute results to ethnic as opposed to socioeconomic categories. In an attempt to disentangle these effects, correlations were run, adjusted for range restriction, for the ethnic/gender groups separately in four parental education categories. The parent education categories were derived from self-reported responses on the Student Descriptive Questionnaire. Students responded for both mother's and father's education level, and the parent with the highest level was used. There are four categories: high school diploma or less (HS), some college but no college degree (HS+), bachelor's degree (B), and graduate degree (G). Students with parents who had some graduate work but no graduate degree were included in the bachelor's degree category. For this and all subsequent analyses, only the recentered scores for the SAT I were used.

As shown in Figure 3 and Table 7, across ethnic groups college grades tend to be more predictable for students whose parents have more education. Within each parental education category, grades were most predictable for Asian-American males, but within-category trends were less clear for the other groups. For example, within the college degree category, SAT I and HSGPA composite correlations were just as high for African-American males as for white males, but correlations for African-American females appeared to be relatively low. In the high school diploma category, SAT I and HSGPA composite correlations were as high for African-American females as for white females. As indicated in

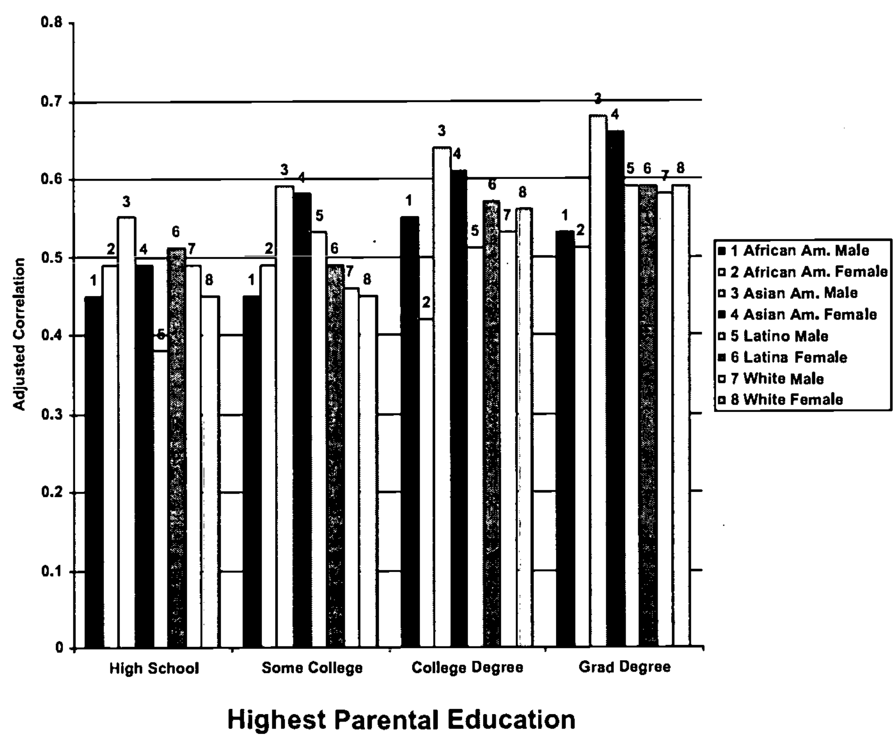

Figure 3. Adjusted correlations of the SAT I + HSGPA composite with FGPA by parental education level for males and females in four ethnic groups.

TABLE 7

Adjusted Correlations by Highest Parent Education Level Between Various Scores and FGPA

Score	Parent Education[a]	Total	Gender Total M	Gender Total F	African American M	African American F	Asian American M	Asian American F	Hispanic/Latino M	Hispanic/Latino F	White M	White F
Number of Students	HS	6,211	2,829	3,382	200	329	674	747	425	568	1,439	1,622
	HS+	9,774	4,395	5,379	346	628	578	694	393	565	2,944	3,303
	B	14,940	7,249	7,691	308	470	1,203	1,268	339	373	5,120	5,295
	G	15,986	7,886	8,100	287	364	1,258	1,309	357	422	5,614	5,608
SAT I	HS	.41	.40	.44	.42	.47	.52	.47	.26	.44	.36	.38
	HS+	.40	.40	.43	.36	.47	.54	.52	.41	.46	.35	.39
	B	.46	.46	.49	.51	.38	.55	.57	.40	.53	.44	.47
	G	.49	.49	.52	.42	.43	.61	.57	.51	.53	.48	.51
HSGPA	HS	.45	.45	.42	.34	.39	.44	.38	.37	.46	.48	.41
	HS+	.44	.45	.42	.41	.39	.49	.50	.50	.40	.44	.41
	B	.51	.50	.49	.44	.36	.57	.49	.49	.46	.49	.51
	G	.55	.54	.54	.51	.47	.58	.59	.53	.50	.53	.53
SAT I + HSGPA	HS	.49	.49	.49	.45	.49	.55	.49	.38	.51	.49	.45
	HS+	.48	.48	.48	.45	.49	.59	.58	.53	.49	.46	.45
	B	.55	.55	.56	.55	.42	.64	.61	.51	.57	.53	.56
	G	.60	.59	.60	.53	.51	.68	.66	.59	.59	.58	.59

[a] HS = High School diploma or less; HS+ = some college; B = Bachelor's Degree; G = Graduate Degree.

Figure 4, analyses run within family income categories revealed the same trends with some within-category variation but a tendency for correlations to be highest in the highest income category.

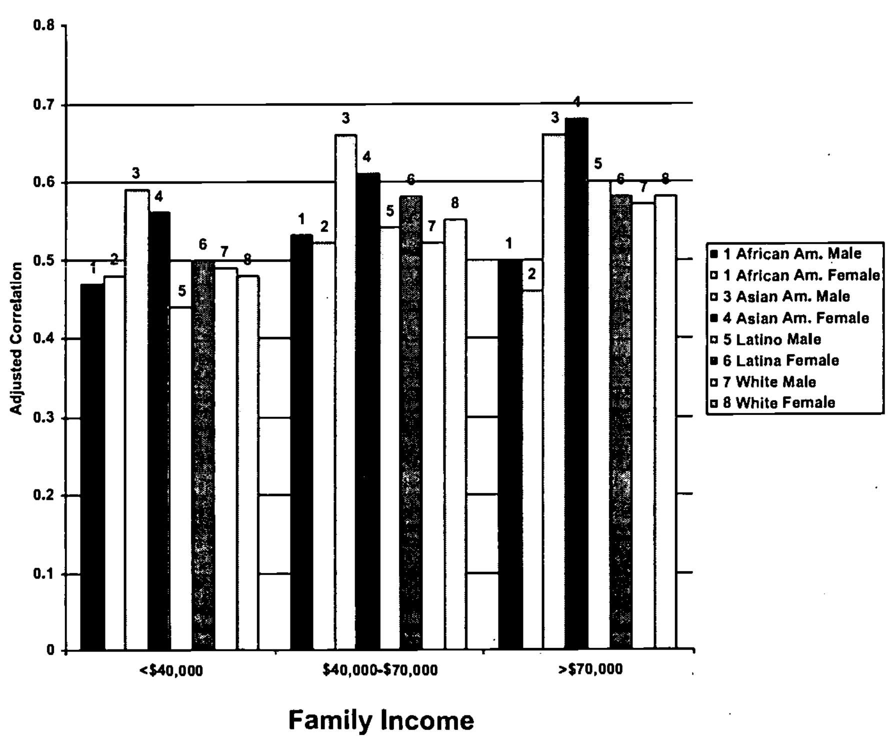

Figure 4. Adjusted correlations of the SAT I + HSGPA composite with FGPA by family income level for males and females in four ethnic groups.

Comparison of a Standardized HSGPA Report With SDQ HSGPA

One of the problems with using the high school grade-point average to predict FGPA is that high schools use different procedures to compute HSGPA. For example, some schools use all courses while other schools use only academic courses; some schools give extra credit for honors courses and others do not; schools that give extra credit for honors courses may differ in the way in which they define an honors course.

An attempt to address the problem of the noncomparability of high school grades has been made by the University of California (UC) system. UC applicants from California are required to submit high school grades in a uniform format (University of California, 1999). This special HSGPA includes only academic core courses as defined by the UC system; it explicitly excludes courses such as physical education, typing, and driver education. In addition, a bonus point is assigned for three categories of courses—honors courses that are approved by the UC, Advanced Placement Program® courses, and International Baccalaureate Higher Level courses. Thus, for example, if a student received a B in an approved honors course, this B would be worth 4 grade points rather than the 3 used in nonhonors courses. Applicants to the UC system fill out a special, machine-readable form that contains detailed instructions on how to enter this uniform grade information. Although these grades are self-reports, students know that the colleges receiving these reports frequently confirm them with similar information submitted directly by the high schools.

The uniform California HSGPA (Cal-GPA) is a step in the direction of standardization, but it makes no attempt to make adjustments across high schools in terms of grading standards. There is no assurance that an A in one school represents the same level of accomplishment as an A in the other school.

Data for the current analyses came from freshmen at six UC campuses. Because of the extra credit granted for honors courses, the mean for the uniform California HSGPA (3.86) was higher than the GPA from the Student Descriptive Questionnaire (3.77) obtained from the same students. Standard deviations were comparable (0.37 for Cal-GPA and 0.39 SDQ-GPA). Because the standard deviation of the Cal-GPA in an unselected sample was not available, the range restriction correction used the SDQ-GPA as an explicit selection variable and used the Cal-GPA as an implicit selection variable. The Cal-GPA correlated .72 with SDQ-GPA, .26 with SAT I–V, and .30 with SAT I–M. In this sample, SDQ-GPA correlated .17 with SAT I–V and .23 with SAT I–M. Correlations with FGPA, adjusted for range restriction, for both high school grade-point averages are presented in Table 8. Correlations were generally higher for the Cal-GPA than for the HSGPA obtained from the SDQ-GPA. Thus, the uniform procedures for collecting high school GPAs in California did succeed in improving predictions. Both the use of SAT I scores and the use of the uniform California HSGPA are intended to even out some of the inconsistencies in predictions from nonuniform HSGPAs. It then might be expected

TABLE 8

Adjusted Correlations of Two Measures of HSGPA With FGPA

Score	Total	*Gender Total*		*African American*		*Asian American*		*Hispanic/Latino*		*White*	
		M	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*
Math/Science Intended Major											
Number of Students	6,183	3,274	2,909	61	146	1,563	1,288	363	396	1,001	863
SDQ-GPA	.59	.60	.58	.69	.47	.62	.62	.51	.42	.54	.61
Cal-GPA	.64	.64	.63	.69	.49	.64	.66	.55	.48	.61	.67
SAT I	.59	.58	.63	.58	.54	.58	.64	.57	.51	.57	.68
SAT I + SDQ-GPA	.67	.67	.69	.77	.66	.69	.72	.63	.56	.63	.74
SAT I + Cal-GPA	.69	.69	.70	.79	.69	.70	.73	.64	.57	.66	.74
Other Intended Major											
Number of Students	7,535	3,068	4,467	87	192	1,208	1,807	430	712	1,111	1,427
SDQ-GPA	.53	.53	.52	.38	.52	.50	.49	.45	.48	.53	.49
Cal-GPA	.57	.57	.56	.32	.51	.54	.53	.49	.50	.56	.56
SAT I	.54	.53	.58	.32	.57	.55	.56	.33	.51	.50	.57
SAT I + SDQ-GPA	.61	.61	.64	.50	.64	.60	.61	.47	.57	.59	.62
SAT I + Cal-GPA	.63	.62	.65	.47	.66	.62	.62	.51	.59	.61	.64

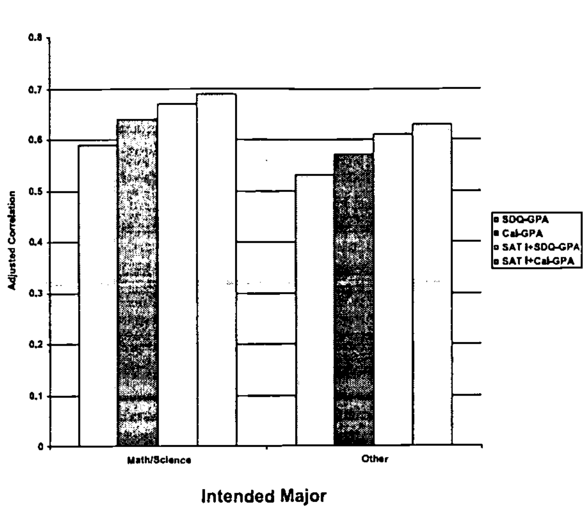

Figure 5. Adjusted correlations of admission measures (SDQ-GPA, Cal-GPA, and composites including SAT I) with FGPA for math/science and other intended majors.

that adding standardized test scores would show a greater validity boost for the SDQ-GPA than for the Cal-GPA. As can be seen in Figure 5, just such a differential boost was found. For the math/science intended majors, the incremental validity for the SAT over the SDQ-GPA was .08 and over the Cal-GPA it was .05; similarly for the other intended majors the SAT increment was .08 over the SDQ-GPA and .06 over the Cal-GPA. The SAT I significantly improved prediction even when the Cal-GPA was used. When combined with SAT I scores, there was only a .02 point difference between the multiple correlations that used SDQ-GPA and the multiple correlations that used Cal-GPA.

Adjustment for Course Difficulty

Because grading standards differ across courses, students in leniently graded courses may receive higher grades, on average, than students with the same academic background who take strictly graded courses. For some students, the FGPA may consist primarily of leniently graded courses while for other students the FGPA may consist primarily of strictly graded courses. Given that students with the highest scores on admission tests often select the scientific and quantitative courses that are graded most strictly, the correlation between admission test scores and FGPA can be attenuated (Elliott & Strenta, 1988; Goldman & Widawski, 1976; Ramist, Lewis, & McCamley-Jenkins, 1994).

For the current sample, adjustments were made in the seven colleges that provided grades in individual courses for 1995. A number of different adjustment methods have been proposed and evaluated (Stricker, Rock, Burton, Muraki, & Jirele, 1994). Three adjustment methods were used: a within-course predicted FGPA, a course-grade residual analysis, and an analysis within an intended college major.

The within-course predicted FGPA followed the procedure outlined by Ramist, Lewis, and McCamley-Jenkins (1994). In this adjustment method, admission scores are used to make linear regression grade predictions in each course containing at least seven freshmen. For each student, the predicted grade for each course taken is averaged over all of the courses taken by that student to form a predicted FGPA for that student. The predicted FGPA is then correlated with the actual FGPA. The within-course predictions were performed separately for each predictor (SAT I–V, SAT I–M, and HSGPA) as well as for the combinations of these predictors (SAT I combined; SAT I and HSGPA). If any equation contained negative regression weights, we removed the variable with the negative weight and recomputed the correlation using the remaining predictors. For courses with just a few students, the weight for a single predictor could be negative; in these cases we substituted the mean grade in the course for the regression estimate. Because regression estimates based on optimal weighting of multiple predictors in relatively small samples may inflate correlations by capitalizing on chance, we also computed the SAT I and HSGPA correlation based on uniform weights, that is the simple sum of SAT I–V and SAT I–M and (200 × HSGPA). The same uniform weight equation was used whether the course was predominantly verbal (such as English) or primarily quantitative (such as calculus), thus producing a very conservative estimate.

As shown in Table 9, the correction for course difficulty (with optimal weights) increased the SAT I and HSGPA correlation by about .06 (from .43 to .49) with an additional increase to .65 when also adjusted for range restriction. These corrections for grading differences are somewhat smaller than those found by Ramist, Lewis, and McCamley-Jenkins (1994), but are consistent with those computed by Stricker et al. (1994). The last six rows of Table 9, and Figure 6, show that the conservative uniform weight correlations were nearly as high as those with the optimal regression weights.

For the course-grade residual analysis, the overall SAT I and HSGPA prediction equation for a college to predict the average grade for all of the students in a given course was used. The course residual was the difference between the predicted grade of the students in that course and the actual mean grade of the students in that course. Thus, each course had a residual value associated with it, with positive residuals indicating a course with higher

TABLE 9

Correlations With FGPA Adjusted for Course Difficulty and Range Restriction

Score	Correction Type*	Total	Gender Total M	Gender Total F	African American M	African American F	Asian American M	Asian American F	Hispanic/Latino M	Hispanic/Latino F	White M	White F
Number of Students		13,344	6,175	7,169	479	713	786	816	599	704	4,002	4,631
SAT I–V	U	.26	.24	.28	.25	.24	.20	.23	.18	.23	.24	.27
	CD	.34	.32	.35	.32	.31	.32	.37	.32	.44	.30	.32
	CD+R	.50	.48	.52	.42	.46	.55	.54	.47	.53	.47	.49
SAT I–M	U	.30	.32	.34	.31	.33	.39	.38	.17	.29	.29	.29
	CD	.37	.37	.39	.38	.38	.42	.40	.31	.47	.34	.34
	CD+R	.52	.52	.55	.48	.52	.62	.56	.46	.56	.51	.52
HSGPA	U	.35	.34	.34	.31	.32	.26	.32	.32	.28	.34	.37
	CD	.42	.41	.42	.41	.40	.37	.39	.42	.47	.41	.40
	CD+R	.59	.56	.60	.51	.54	.58	.59	.56	.58	.56	.59
SAT I	U	.33	.34	.37	.35	.35	.41	.41	.23	.34	.32	.33
	CD	.40	.39	.42	.39	.40	.44	.43	.33	.48	.36	.37
	CD+R	.56	.55	.58	.49	.54	.63	.60	.49	.57	.53	.56
Optimal Weights SAT I + HSGPA	U	.43	.43	.45	.43	.44	.46	.47	.41	.42	.41	.42
	CD	.49	.48	.50	.47	.48	.48	.47	.45	.54	.45	.46
	CD+R	.65	.63	.67	.56	.61	.67	.67	.59	.65	.62	.66
Uniform Weights SAT I + HSGPA	U	.39	.39	.41	.37	.38	.39	.41	.29	.35	.37	.39
	CD	.46	.45	.48	.46	.45	.44	.45	.43	.52	.43	.44
	CD+R	.63	.61	.66	.54	.60	.66	.66	.57	.63	.60	.65

*U is uncorrected correlation, CD is corrected for course difficulty, CD + R is corrected for course difficulty and range restriction.

grades than would be expected from the admission scores of the students in that course, that is, a course with lenient grading; negative residuals indicated strict grading. For a given student, these residuals were averaged over all of the courses taken by that student. This mean residual was then used as an additional predictor (along with SAT I–V, SAT I–M, and HSGPA) in predicting the FGPA for a student. Results of this procedure were nearly identical to those for the predicted FGPA procedure (mean correlation over colleges of .50 for mean grade-residual analysis compared to .49 for the predicted FGPA procedure).

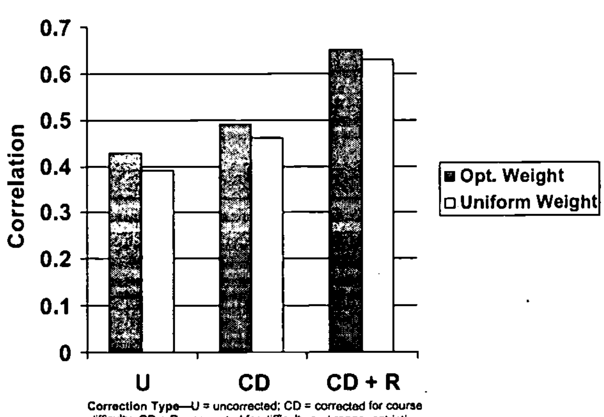

Figure 6. Effects of correction for course difficulty and range restriction on correlation of FGPA with SAT I + HSGPA composite (optimal weights and uniform weights).

The third procedure did not directly adjust for differences in course grading; it merely grouped students into more homogeneous categories based on their intended college majors as indicated by their responses to the Student Descriptive Questionnaire. This method has obvious drawbacks in that students frequently change their intended majors before or after enrolling in college, and even students with different majors can have a similar mix of courses during the freshman year. Nevertheless, this approach has the distinct advantage of not requiring colleges to supply any course-level information, so it could be used for all 23 of the colleges in the sample. All majors were grouped into two categories—"math/science" included majors in the physical and biological sciences, engineering, and mathematics; all other majors were put in the "other" category. As indicated in Table 10 and Figure 7, correlations were uniformly higher in the math/science group. Note that because these correlations were adjusted for range restriction the higher

TABLE 10

Correlations With FGPA by Intended Major, Adjusted for Range Restriction

Score	Intended Major	Total	*Gender Total* M	*Gender Total* F	*African American* M	*African American* F	*Asian American* M	*Asian American* F	*Hispanic/Latino* M	*Hispanic/Latino* F	*White* M	*White* F
Number of Students	M/S	16,742	9,446	7,296	496	631	2,039	1,613	626	624	5,810	4,053
	O	29,553	12,556	16,997	603	1,135	1,627	2,375	826	1,257	8,958	11,498
SAT I–V	M/S	.50	.48	.52	.43	.49	.51	.55	.46	.47	.46	.49
	O	.41	.45	.50	.32	.48	.49	.48	.34	.46	.44	.49
	T	.47	.46	.50	.37	.48	.51	.50	.40	.46	.44	.48
SAT I–M	M/S	.54	.54	.58	.50	.54	.59	.61	.47	.50	.51	.55
	O	.47	.47	.52	.40	.51	.53	.50	.35	.48	.45	.49
	T	.48	.49	.52	.44	.52	.57	.54	.41	.49	.46	.49
HSGPA	M/S	.58	.58	.57	.51	.53	.59	.58	.54	.48	.57	.55
	O	.54	.52	.54	.42	.45	.50	.49	.45	.49	.52	.54
	T	.54	.53	.53	.47	.49	.55	.52	.49	.48	.53	.53
SAT I	M/S	.57	.56	.60	.55	.58	.61	.65	.54	.57	.54	.57
	O	.51	.50	.56	.47	.55	.57	.55	.39	.53	.49	.53
	T	.52	.52	.56	.46	.55	.59	.58	.45	.54	.49	.53
SAT I + HSGPA	M/S	.66	.65	.67	.63	.67	.69	.72	.66	.64	.64	.65
	O	.60	.58	.63	.56	.59	.62	.61	.51	.60	.58	.62
	T	.61	.60	.63	.55	.61	.66	.64	.56	.61	.59	.61

Note: M/S = math/science; O = other; T = total.

correlations in the math/science category cannot be attributed merely to greater variability of scores in that category. The means, standard deviations, and standardized differences *(d)* in Table 11 indicate that the difference

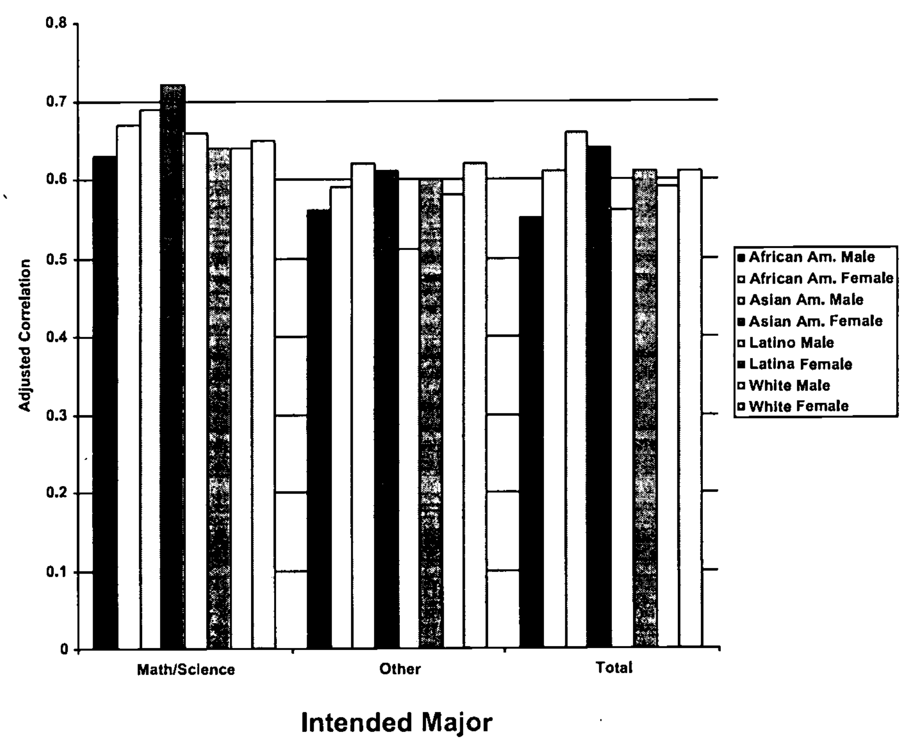

Figure 7. Adjusted correlations of the SAT I + HSGPA composite with FGPA by intended major for males and females in four ethnic groups.

TABLE 11

Means and Standard Deviations of Scores and GPAs by Intended Major

	Math/Science		Other		
	M	SD	M	SD	d
SAT I–V	578	84	560	84	0.21
SAT I–M	628	76	564	83	0.81
HSGPA	3.70	0.43	3.54	0.46	0.36
FGPA	3.01	0.79	2.96	0.74	0.07

Note. The standardized difference between the Math/Science and Other categories, *d*, is the mean difference divided by the square root of the unweighted average of the squared standard deviations.

in FGPA between the math/science and other groups is considerably smaller than the difference in any of the admission measures, suggesting that grading standards were indeed more rigorous for students whose intended major was in a math/science field.

Over/underprediction adjusted for course difficulty

Because members of different gender and ethnic subgroups may differentially sort themselves into courses with relatively strict or lenient grading standards, adjusting for course difficulty can also have an impact on the extent to which grades are over- or underpredicted. For each subgroup, we used the FGPA predicted in the course grade residual analysis (SAT I and HSGPA and Residual) that accounts for course difficulty differences, and then found the difference between the predicted and actual FGPA. These differences are presented in Table 12; as before, positive differences indicate predictions for a group that were higher than the grades actually received (overprediction). The correction reduced the underprediction of women's grades from -0.07 to -0.05; for women in the two colleges in the highly selective category that provided course grades, underprediction was -0.04 before correction and -0.03 after correction.

For the full sample of colleges, not just those that supplied course grades, over/underprediction results by intended major are presented in Table 13. Grade predictions were made without regard to intended major, but differences between predicted and actual grades were computed within major (math/science or other). Results were consistent with the notion that grading standards are more strict in math/science fields. In every gender/ethnic category, grades for students with intended majors in math/science were not as high as predicted from the SAT I and HSGPA equation for the entire college. The generalization that grades of women are underpredicted was not true for these math/science students, though the overprediction for men was notably larger than the overprediction for women.

The results of all of the course-adjustment procedures underscore the importance of taking grading differences into account whenever possible for predictions of FGPA. When adjustments cannot be made, it should at least be acknowledged that the resulting correlations are underestimates of the ability of the admission measures to predict college grades.

Conclusions

The content changes in the SAT and the recentering of the score scale had little or no impact on the overall predictive validity of the test. The SAT I appears to predict about equally well across ethnic groups. At most colleges, grades of females are more predictable than grades of males, but at the most highly selective colleges, the grades of males and females are predicted equally well. Males generally perform slightly worse in their freshman year than predicted from test scores and high school grades; women perform slightly better than predicted. Within African-American and Hispanic/Latino groups, men perform worse than predicted and women perform about as predicted.

Across ethnic groups, grades are more predictable for higher socioeconomic status (SES) students than for lower SES students. This is true for both parent education and income definitions of SES.

TABLE 12

Over- (+) and Underprediction (-) of FGPA Adjusted for Course Difficulty

	Gender Total		African American		Asian American		Hispanic/Latino		White	
Score	M	F	M	F	M	F	M	F	M	F
Number of Students	6,175	7,169	479	713	786	816	599	704	4,002	4,631
SAT I	+.13	-.11	+.23	-.07	+.02	-.12	+.16	-.01	+.13	-.13
SAT I + Residual	+.11	-.09	+.24	-.07	-.03	-.11	+.18	+.02	+.11	-.11
SAT I + HSGPA	+.08	-.07	+.16	-.05	.00	-.07	+.12	+.02	+.09	-.09
SAT I + HSGPA + Residual	+.06	-.05	+.15	-.04	-.05	-.05	+.13	+.05	+.07	-.07

TABLE 13

Over- (+) and Underprediction (-) of FGPA for Students With Math/Science or Other Intended Majors

Score	Gender Total		African American		Asian American		Hispanic/Latino		White	
	M	F	M	F	M	F	M	F	M	F
Math/Science Intended Major										
Number of Students	9,519	7,342	513	642	2,072	1,641	653	646	5,846	4,075
SAT I	.13	-.04	.22	-.02	.07	-.01	.20	.08	.13	-.07
SAT I + HSGPA	.12	.02	.17	.03	.08	.03	.20	.11	.13	.01
Other Intended Major										
Number of Students	12,644	17,138	626	1,157	1,653	2,404	847	1,277	9,022	11,597
SAT I	.11	-.13	.23	-.04	.10	-.09	.18	-.03	.09	-.16
SAT I + HSGPA	.04	-.11	.12	-.03	.05	-.08	.13	-.03	.02	-.13

Accounting for differences in course grading practices produces a noticeable improvement in predictions of FGPA. Even as simple a procedure as running correlations separately for students who indicate that they would like to be math/science majors has an impact on the size of validity coefficients. Validity coefficients would have been even larger had we adjusted for unreliability in the FGPA criterion as was done by Ramist et al. (1994). However, we had no way of adequately estimating the reliability of the FGPA for all of the schools in our sample. If grade reliability were about the same in our sample as in the Ramist et al. sample (a reasonable but unverifiable assumption), about .05 should be added to the adjusted correlations. Thus, for example, correcting for the unreliability of the FGPA would raise the correlation for the SAT I and HSGPA composite in math/science students from .66 to .71.

Many issues remain to be explored in future analyses of the data base created for this study, notably how SAT II: Subject Tests relate to the other measures in predicting FGPA. In addition, the database can also be used to simulate the effects of different selection models (such as using SAT II scores in the place of SAT I scores) on the quality and ethnic/gender composition of an admitted class. Additional data will be needed to explore longer-term validity issues.

References

Breland, H. (1979). *Population validity and college entrance measures.* Princeton, NJ: College Board Publication Orders.

Burton, N. W., & Ramist, L. (in preparation). *Predicting long term success in undergraduate school: A review of predictive validity studies.* Princeton, NJ: Educational Testing Service.

College Board/Educational Testing Service (1995). *1995 College Bound Seniors.* New York: College Entrance Examination Board.

Elliott, R., & Strenta, A. C. (1988). Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race particularly. *Journal of Educational Measurement, 25*, 333-347.

Freeberg, N. E., Rock, D. A., & Pollack, J. (1989). *Analysis of the revised Student Descriptive Questionnaire: Phase II, Predictive validity of academic self-report* (College Board Report No. 89-8). New York: College Entrance Examination Board.

Goldman, R. D., & Widawski, M. H. (1976). A within subjects technique for comparing grading standards: Importance in the validity of the evaluation of college achievement. *Educational and Psychological Measurement,* 36, 381-390.

Gulliksen, H. (1950). *Theory of Mental Tests.* New York: John Wiley and Sons.

Linn, R. L. (1978). Single-group validity, differential validity, and differential prediction. *Journal of Applied Psychology, 63*, 507-512.

Morgan, R. (1994). *Effects of score choice on predictive validity.* Princeton, NJ: Educational Testing Service.

Pennock-Román, M. (1994). *College major and gender differences in the prediction of college grades* (College Board Report No. 94-2). New York: College Entrance Examination Board.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). *Student group differences in predicting college grades: Sex, language, and ethnic groups* (College Board Report No. 93-1). New York: College Entrance Examination Board.

Sawyer, R. (1986). Using demographic subgroup and dummy variable equations to predict college freshman grade average. *Journal of Educational Measurement, 23*, 131-145.

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.

University of California (1999). *Admissions as a Freshman.* http://www.ucop.edu/pathways/infoctr/introuc/fresh.html

Willingham, W. W. (1985). *Success in college: The role of personal qualities and academic ability.* New York: College Entrance Examination Board.

Willingham, W. W., Lewis, C., Morgan, R., & Ramist, L. (1990). *Predicting college grades: An analysis of institutional trends over two decades.* Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1983). *A review of research on the prediction of academic performance after the freshman year* (College Board Report No. 83-2). New York: College Entrance Examination Board.

NOTICE

Reproduction Basis

- [x] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

- [ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").