

SATs, ACHIEVEMENT TESTS, AND HIGH-SCHOOL CLASS RANK AS PREDICTORS OF COLLEGE PERFORMANCE

JONATHAN BARON AND M. FRANK NORMAN
University of Pennsylvania

This study asked how well students' grades at the University of Pennsylvania could be predicted from linear combinations of high-school class rank (CLR), total scholastic-aptitude-test score (SAT), and average achievement-test score (ACH), all of which are available in applications for admission to selective institutions. Underrepresented minorities were not included in the main analysis. Consideration of multiple regression coefficients revealed that CLR and ACH added significantly to overall prediction, whereas SAT did not. This asymmetry in incremental validity was not easily explained by restricted range, self-selection, or nonlinear effects. SAT did add significantly to prediction of grades in some individual courses. Of the three variables, only CLR added significantly to prediction of attrition or number of uncompleted courses.

THE usefulness of the Scholastic Aptitude Test (SAT) (Educational Testing Service, 1948–1990) for predicting success in college has recently been debated. Crouse and Trusheim (1988) argued that high-school grades are better than the SAT at predicting college performance and that, while the SAT improves prediction significantly over high-school grades alone, the improvement is too small to be worth the effort. In reply, Elliott and Strenta (1990) noted that the incremental value of the SAT over high-school grades is much greater for more selective institutions.

For crucial support and suggestions, we thank Norman Adler, Michael Aiken, Clarence Brest, James Crouse, Susan Shaman, Hugo Sonnenschein, and Willis Stetson. For assistance with obtaining data, and other advice, we thank Hua Pei Chen, Shari Lazarus and (especially) Dan Shapiro.

Copyright © 1992 Educational and Psychological Measurement, Inc.

Selective institutions typically ask applicants for achievement test scores in addition to high-school rank and SATs. Little detailed information is available relevant to possible redundancies between these three measures at such institutions. Elliott and Strenta do not report data concerning the value of the SATs when both high-school grades and achievement tests are taken into account. Crouse and Trusheim (1988, Table 8.1, p. 159) report average squared multiple correlations calculated from an ETS study of 22 "highly selective" institutions. These averages indicate that high-school grades and *either* SATs or achievement test scores were almost as good as all three predictors together ($R^2 = .227$ with all three, .213 with grades and achievements, .215 with grades and SATs), but many students in this sample took only a single achievement test. Crouse and Trusheim did not provide further information about this study, nor did they cite a publicly available source. Donlon (1984, Table 8.19, p. 166) summarized data suggesting that SATs have some value even when both achievement tests and high-school grades are known, but, again, the number of achievement tests taken is not reported, nor are many other relevant details. More recent sources, such as Willingham, Lewis, Morgan, and Ramist (1990), do not consider the role of achievement tests in prediction.

Detailed analyses of data from individual institutions may bring to light subtle relationships that are obscured in averages like those reported by Crouse and Trusheim (1988) and by Donlon (1984). Once discovered, the generality of such relationships can be investigated separately.

The purposes of the present study were (a) to examine the predictive utility of the SAT along with achievement tests and high-school class rank for the undergraduate classes entering the University of Pennsylvania in 1983 and 1984 and (b) to examine the role of restricted range, self-selection, and nonlinearity in the assessment of predictive utility.

Method

Sample

The sample consisted of a total of 4170 students (2781 in arts and sciences, 647 in business, 585 in engineering, and 157 in nursing), all students who entered the University as freshmen in the fall of 1983 and of 1984. The main analysis is based on 3816 students, because underrepresented minority students (African-Americans and Hispanics) were excluded.

Procedure

Admissions were coordinated through a single admissions office, which evaluated academic qualifications largely on the basis of a predictive index, or PI. The PI consisted of an equally weighted composite of (a) the mean SAT score (divided by 10, so that the maximum is 80); (b) the mean of *three* College Entrance Examination Board Achievement tests (divided by 10), the highest of English or English Composition and the highest two others; and (c) the student's high-school class-rank rescaled on the same 80 point scale. (High-school grade-point average was not recorded for most students.) These three measures were called SAT, ACH, and CLR, respectively. Other criteria were used subjectively for admission, so some students with low PIs were admitted.

Results

Statistical Analyses

The prediction of academic performance as represented by the student's cumulative grade point average (CUM) at the university between the time of matriculation and 1988 was examined. Table 1 shows the linear relation between each PI component and CUM, both as a raw correlation and as a standardized multiple regression weight. All three elements correlated with CUM. The regression weights indicate the contribution that each predictor makes when the others are known. It is apparent that the SAT was not useful once ACH and CLR were known. The reason for this result is that SAT correlated highly (.73) with ACH, and ACH is somewhat more highly correlated with CUM than is SAT. All the useful predictive power of SAT seems to be contained in the achievement tests and high-school rank. The correlation between CUM and each of the two SAT scores, math and verbal, was lower than between CUM and each of the six most commonly taken achievement tests (among those who took each test): American History, Biology, Chemistry, English, English Composition, Mathematics 1, and Mathematics 2. High-school rank was the best single predictor.

It may be helpful to the reader to point out that, although other significant predictors were found, they do not alter the conclusions presented here. For example, the admissions staff's ratings of essays that students wrote in their applications were significantly useful in some analyses, although no other subjective ratings were useful. Further, when underrepresented minorities were included in the

TABLE 1
Predictive Power of SAT, ACH, and CLR

Measure	SAT	ACH	CLR
Raw correlation with CUM	.199*	.261*	.305*
first year only	.264*	.328*	.344*
Delaware (first year)	.313*	.366*	.409*
Regression weight for CUM	-.013	.220*	.266*
first year only	.015	.262*	.291*
Delaware (first year)	-.001	.214*	.326*
Logistic weight for status	.020	-.005	.039*
Regression weight for grade	-.032	.018	.126*

Note. Regression weights are standardized except for the prediction of status which is from a logistic regression. "Grade" is the negative of the number of incompletes or withdrawals. The asterisk (*) indicates statistical significance at the .05 level (typically, probabilities were much lower).

sample, the coefficient for SAT was significant and positive, but when indicator variables for minority groups were included in the regression, or when the groups were analyzed separately, the coefficient was again nonsignificant. These groups had lower CUMs and lower SAT scores than other students, but SAT did not contribute to the prediction of CUM *within* any group. Sex was somewhat useful as a predictor, with females doing better than otherwise predicted, but inclusion of sex in the regression does not change any reported conclusions.

Table 2 shows the mean CUM as a function of SAT (here, the sum of both tests) and a new predictive index, PI2, consisting of the sum of ACH and CLR alone. It is apparent that, holding SAT constant, CUM increases with PI2, but, holding PI2 constant, CUM does not change as a function of SAT. Subject to the qualification that the measurement categories are crude, this analysis shows that SAT makes no contribution to prediction, linear *or nonlinear*, once PI2 is given.

Table 3 shows the variance in CUM explained by each combination of predictors, and the standardized regression weights for each equation (correlations for single variables). It is apparent that the SAT has some incremental validity when only CLR is available (as found in all previous studies), but it has no incremental validity when ACH is available, whether CLR is also available or not.

The prediction of two other relevant measures available were also examined: (a) status (graduated vs. dropped out for any reason, with a few active students omitted) and (b) the number of uncompleted courses (incompletes, no-reports, and withdrawals). The raw regression weights for status, in a logistic regression (Dallal, 1988) are

TABLE 2
*Mean CUM as a Function of SAT (Sum of Both Tests) and PI2
 (Sum of CLR and ACH)*

		PI2					
		105	115	125	135	145	155
SAT	105	2.64	2.88	3.05			
	115	2.75	2.84	3.07	3.27		
	125	2.66	2.91	3.12	3.22	3.46	
	135		2.87	3.10	3.25	3.45	
	145			3.18	3.27	3.43	3.69
	155				3.28		

Note. Cells with fewer than 15 students were omitted. The label of each row and column indicates the interval from 5 below to 4 above.

shown in Table 1, as are the standardized weights for uncompleted courses. It is apparent that neither SAT nor ACH contributes significantly to prediction of these variables. As in prediction of CUM, inclusion of underrepresented minorities in the sample led to a significant weight for SAT, but the weight was not significant when indicator variables for the groups were included, or when the groups were analyzed individually.

The initial results suggested that the SAT is redundant when good measures of past performance are available, ACH and, especially, CLR (which is useful in predicting other outcomes aside from grades). Several reasons why the SAT might still contribute to prediction despite these findings were considered: (a) restriction of range, (b) self-selection, and (c) nonlinearity. Each topic is addressed in the following sections.

Restriction of Range

Students were selected on the basis of the very predictors that were examined. Such selection will reduce the range of variation on

TABLE 3
*R² and Standardized Regression Weight (Correlations for Single Variables) for
 each Combination of Predictors*

Predictors	Regression weights			R ²
	SAT	ACH	CLR	
SAT	.199	—	—	.040
ACH	—	.261	—	.068
CLR	—	—	.305	.093
SAT & ACH	.019	.247	—	.068
SAT & CLR	.145	—	.277	.113
ACH & CLR	—	.211	.265	.136
SAT & ACH & CLR	-.013	.220	.266	.136

each predictor and reduce its correlation with CUM and its regression weight (Dawes, 1975). The SAT might be valuable in prediction if the entire applicant pool were admitted.

To test this hypothesis, James Crouse and Dana Keller (of the Department of Educational Studies, University of Delaware) were asked to predict first-year grades for a sample from the University of Delaware, a less selective institution. The Delaware sample used only the 547 students who reported three achievement scores. Despite this limitation, the standard deviation (SD) of SAT was 6.9, which is closer to the 7.4 SD of our applicant pool than to the 5.4 SD of our sample. It was assumed that if restriction of range was a problem, the weight of SAT for this sample should have been higher. It was not. The weight of SAT was slightly, but not significantly, negative, despite the increased variance in SAT. It was concluded that restriction of range does not seem to explain the nonsignificant weight of the SAT.

Self-selection

Students who do well on the SAT might select themselves into more difficult courses. Strenta and Elliott (1987) provided evidence for self-selection by SAT, but they did not examine self-selection by ACH or CLR. Students might think that the SAT is a particularly valid indicator of their ability to master a challenge. To test this possibility, the grades were examined in the 83 courses that were most often selected by students admitted in 1983 (minimum enrollment 175 for the two years combined). A negative correlation across courses between *mean* grades in a course and *mean* SAT of those who take it would support this self-selection hypothesis. The same argument applies to ACH and CLR. In fact, mean grade was correlated negatively with all three predictors: raw correlations were $-.39$, $-.37$, and $-.36$ for SAT, ACH, and CLR, respectively, all significant at $p = .001$ or better. All three predictors seem to be equally important in self-selection.

The same kind of analysis was conducted by departments instead of courses, including all 45 departments with more than 100 students in the 2-year sample. The dependent measure in the regression was each department's mean, over all students taking courses in the department, of each student's mean grade in courses in the department. The predictors were the mean SAT, ACH, and CLR of students taking courses in the department. The raw correlations across the 45 departments for mean grade with mean SAT, ACH, and CLR, respectively, were $-.40$, $-.38$, and $-.39$ (all significant at

$p \leq .01$), indicating that better students tended to pick courses in harder fields. Again, the degree of selection on the three predictors was essentially identical, so self-selection alone cannot explain the nonsignificant regression weight of the SAT in the main analysis.

Nonlinearity

Addition of quadratic terms to all predictors did not change the conclusion that the total variance in CUM accounted for by SAT (effect of linear and quadratic terms together on R^2) was nonsignificant. (Quadratic terms were individually nonsignificant for all three predictors, and all three quadratic terms together were nonsignificant.) Similar analysis of the prediction of status using ogival functions also yielded no statistically significant SAT effect.

Although the SAT had no *overall* incremental validity when other predictors were available, its regression weight was still significantly positive ($p \leq .05$ —two-tailed, but uncorrected for the number of tests) for predicting grades in 19 courses in the sample of 83 (in Economics, English, Mathematics, Psychology, and several business fields). (In the analyses described in this paragraph, indicator variables were included for female, African-American, and Hispanic classifications.) The regression weight of SAT was significantly negative in two courses (one in Mathematics, one in Art). In addition, the SAT contributed significantly to the prediction of grades in 14 out of 45 departments (Anthropology, Communications, Economics, Psychology, English, Nursing, and all 9 business departments except Statistics). The regression weight of SAT was (nonsignificantly) negative in other departments, including all foreign language and literature departments, Chemistry, Electrical Engineering, Mathematics, and Philosophy. The regression weight of SAT was higher in departments that are unrelated to high-school courses in which achievement tests are given (point biserial $r = 0.31$, $p = .04$), including business courses. In such fields, ACH and CLR may be less informative about specific knowledge. These fields constitute only a small proportion of courses taken, however. In most large courses and fields, the multiple regression weight of SAT is indistinguishable from zero.

Conclusion

It cannot be concluded that the incremental predictive value of the SAT is zero for all purposes. Rather, these results tend to strengthen the conclusion of Crouse and Trusheim (1988) that the SAT makes

a relatively small contribution to prediction. When achievement tests are known as well as high-school class rank, the SAT has even less incremental validity than their results suggested, even at selective institutions. Of course, results may vary from institution to institution, so the kind of analyses reported here should be repeated elsewhere.

When an institution considers whether to adopt a test with such questionable incremental value, other considerations become relevant aside from predictive power. Use of achievement tests and high-school class rank for college admission could provide high-school students with incentive to achieve academically, but the SATs are not designed to provide such incentive. If SATs are more "fair" because they apply the same nominal standard to all, this gain in fairness must be balanced against the cost in dilution of incentives for high-school performance.

Although the results of this study may call into question routine use of the SAT at some selective institutions, they ought to encourage more widespread reliance on current achievement tests and further development of such tests. It is interesting to note that achievement tests scores are complementary to class rank, in the sense that they are affected in compensatory ways by secondary school quality. A student at a better school may pay a CLR "penalty" due to competition from especially able or well-educated classmates but will reap an ACH "reward" from the same competition and from superior instruction. Consequently, a "predictive index" that sums CLR and ACH seems reasonable, as well as predictively effective.

REFERENCES

- Crouse, J. and Trusheim, D. (1988). *The case against the SAT*. Chicago: University of Chicago Press.
- Dallal, G. E. (1988). LOGISTIC: A logistic regression program for the IBM PC. *The American Statistician*, 42, 272.
- Dawes, R. M. (1975). Graduate admission variables and future success. *Science*, 187, 721-723.
- Donlon, T. F. (Ed.). (1984). *The College Board technical handbook for the Scholastic Aptitude Test and Achievement Tests*. New York: College Entrance Examination Board.
- Educational Testing Service. (1948-1990). *College Board Scholastic Aptitude Test*. Princeton, NJ: Author.
- Elliott, R. and Strenta, A. C. (1990). *Is the SAT redundant with high school record in college selection? Research and Development Update*. New York: The College Entrance Examination Board.

- Strenta, A. C. and Elliott, R. (1987). Differential grading standards revisited. *Journal of Educational Measurement*, 24, 281–291.
- Willingham, W. W., Lewis, C., Morgan, R., and Ramist, L. (1990). *Predicting college grades: An analysis of institutional trends over two decades*. New York and Princeton: The College Board and Educational Testing Service.