Predicting Academic Achievement: The Validity of SAT II Science Tests

A new report from the Educational Testing Service reveals that SAT II subject tests vary dramatically in their ability to predict college achievement. According to the findings, science and math achievement tests are the strongest predictors of college performance, while some of the language tests are the weakest.

The ETS has offered SAT II subject tests (formerly called SAT achievement tests) for several decades, but few colleges and universities have required applicants to take them. As a result, research had focused primarily on the validity of their more widely used counterpart, the SAT I Reasoning Test. However, a proposal by University of California president Richard Atkinson to replace the SAT I with an achievement test developed by the state spurred interest in more information about SAT II validity.

Atkinson has proposed that UC rely on a battery of SAT II tests until the state can develop its own test. Currently, UC requires applicants to take three SAT II exams: the writing test, one of the two math tests, and a third exam of their choice.

In response to the proposal, ETS researchers Leonard Ramist, Charles Lewis, and Carolyn McCauley-Jenkins undertook a detailed analyses of SAT I, SAT II, and high school grades as predictors of freshman GPA and performance in specific disciplines. UC also conducted its own validity study, based on data from its student body.

Of the two reports, the UC study has the advantage of being based on more recent data. The ETS analysis, however, is by far the more extensive and includes findings regarding the validity of individual tests that the UC report does not.

Of the SAT II tests for which data are provided, three are in the natural sciences and two in math. After adjustment for range restriction and other statistical artifacts, correlations between these five and freshman grade point average (FGPA) are striking. All science tests show correlations exceeding .5--and chemistry and math II tests have the highest validities of the 14 subject tests studied.

Among the remaining SAT II subject tests, only English composition (now known as Writing) had a correlation with FGPA exceeding 0.5. Subject tests for German language and European history showed considerably less validity. The Spanish test proved the weakest predictor of all and was the only test where the correlation between scores and grades fell short of 0.2.

Science/Math versus Other SAT II Subject Tests
Correlations with Freshman GPA*

Science & Math		English & History		Foreign Languages	
Biology	.51	American History	.47	French	.35
Chemistry	.58	English Composition	.51	German	.23
Math I	.52	European History	.28	Hebrew	.31
Math II	.58	Literature	.48	Latin	.38
Physics	.52			Spanish	.17

**Corrected for range restriction and shrinkage. Adapted from ETS report, Table 21.*

The analysis also found that while science and math tests are sound predictors of future performance in these disciplines, their validity is enhanced by combining scores with results from the SAT I. When results on both SAT II science tests and SAT I were considered together, predictive validity for the same subject taken in college was the highest achieved for any subject for which an SAT II is available. Nonetheless, the incremental validity of adding results from SAT II science or math test tests to the SAT I alone was modest--increasing predictive validity by .05-08.

Subject Tests and College Performance
Correlations with Specific Courses*

Test	Courses	
Biology	Advanced Biology	.50
Chemistry	Advanced Chemistry	.60
Math I	Advanced Math	.50
Math I	Calculus	.51
Math II	Advanced Math	.52
Math II	Calculus	.52
Math II	Advanced Math	.52
Physics	Advanced Physics	.59

**Corrected for range restriction and shrinkage. Adapted from Tables 38 &46 of report.*

The report also considers performance of different demographic groups, breaking down results by sexes, ethnicity, and native language. Test-takers were also categorized by an academic composite based on grades and test scores.

Minority students, those whose best language is not English, and students in the low academic composite generally performed better on the SAT II science tests than would have been expected based on their SAT I scores alone. By contrast, students in the high academic composite performed less well on the chemistry and physics tests than predicted based on their SAT I scores.

The most dramatic difference between expected and actual performance was seen among the small number of females who took the SAT II physics test. In addition, black students and those whose best language is not English performed better than predicted on SAT II math tests. This suggests that questions on the SAT I math sections are more dependent on the English skills than those on the SAT II.

A key limitation of the report is the absence of data for a broader range of academic outcomes. In a previous work on the predictive value of admissions criteria, author Ramist and colleague Nancy Burton reviewed evidence that a combination of SAT score and high school grades predicts multiple measures of success. In addition to GPA, these include college or departmental honors, acceptance to graduate or professional school, and completion of a graduate or professional degree. Corresponding data for the SAT II remain to be presented.

ETS cautions that some changes have been made to SAT II tests since data for the report were collected; whether these might impact the results is unknown. Also, as a group, SAT II test-takers are somewhat stronger academically than the general college bound population, since most schools requiring the SAT II are highly selective ones. Whether the findings reported here would change if SAT II test-takers become more representative of the college-bound population remains to be determined.

The new report, entitled Using Achievement Tests/SAT II: Subject Tests to Demonstrate Achievement and Predict College Grades; Sex, Language, Ethnic, and Parental Education Groups, is available free of charge on-line. Hard copies of both the new report and its predecessor, Predicting Success in College: SAT Studies of Classes Graduating Since 1980, are available for $15 each plus $4 shipping.

Orders should be mailed to College Board Publications, Box 886, New York, NY 10101-0886 or placed by phone at 800-323-7155. Item numbers are #992620 for the new report and #990299 for the earlier work on predicting success in college.

The Validity of GRE Subject Tests

The validity of science achievement tests is not limited to those taken in high school. A little-noticed paper published last year shows that Graduate Record Examination subject tests are also impressive predictors of future performance.

Like SAT II tests, GRE subject tests assess knowledge of a specific discipline. By contrast, the GRE-general measures reasoning ability in verbal, quantitative, and analytic domains. And like the SAT, the GREs have been under fire. According to three researchers at the University of Minnesota, criticisms of the GRE run the gamut from claims that the test is "situationally specific" to "useless."

To examine the soundness of these charges, the researchers—Nathan Kuncel, Sarah Hezlett and Dennis Ones—conducted a meta-analysis of studies containing validity data for GRE tests. Eight indicators of graduate school performance were considered: first year and overall GPA; scores on comprehensive exams; faculty ratings; number of publications; citations; degree attainment, and time to degree. More than 1500 studies contained relevant data, most often pertaining to the GRE-General.

Part of the analysis involved the calculation of operational validities that adjust for statistical artifacts that can obscure the true relationship between a predictor and future outcome. After adjusting for range restriction and criterion unreliability, the authors found substantial correlations between GRE scores and academic outcomes. "Lower correlations and much of the variability in previous research," they observe, "are likely to have been the result of range restriction and sampling error, respectively."

Comparing the operational validity of subject tests, the three sections of the GRE-General, and undergraduate GPA revealed that the subject tests averaged noticeably higher values. The best predictor for each of the eight performance indicators is shown in boldface in the chart that follows. As can be seen, subject test scores best predicted all but one.

When separate analyses were performed on the subject tests by category (humanities, social sciences, life sciences, and math-physical sciences), little deviation from this pattern emerged. Subject tests showed similar or better validity relative to other predictors. In the life sciences, subject GREs had an operational validity that was .05 validity points greater than the score on the quantitative section of the GRE-general, and .14 validity points higher than UGPA. In math and physical sciences, UGPA led the subject GRE by .01 validity point. However, the subject GRE bested the GRE-verbal by .10 validity points.

Operational Validity of Graduate Admissions Criteria for Predicting Graduate School Performance[a]

	SU	VE	QU	AN	UGPA
Performance Measure					
Graduate GPA	**.41**	.34	.32	.36	.30
1st Year GPA	**.45**	.34	.38	.36	.33
Comprehensive Exam Scores[b]	**.51**	.44	.26	NA	.12
Faculty Ratings	**.50**	.42	.47	.35	.35
Degree Attainment[c]	**.39**	.18	.20	.11	.12
Time to Degree Completion [b]	**.**02	**.28**	- .12	NA	-.08
Research Productivity[b]	**.21**	.09	.11	NA	NA
Publication Citation Count[b]	**.24**	.17	.23	NA	NA

SU=GRE Subject Test *VE=GRE Verbal* *QU= GRE Quantitative*
AN=GRE Analytical *UGPA= Undergrad GPA*
[a] bold=best predictor for criterion *[b]not corrected for criterion unreliability* *[c]not corrected for range restriction*

The authors explore possible explanations for the higher predictive validities of the subject tests. The most likely, in their view, is that subject tests better reflect the impact of motivation and interest on learning than more general tests. Nonetheless, they note that general tests play a valuable role in graduate admissions by providing a basis for predicting success among those who apply to programs in areas that they did not study as undergraduates.

While there has been much discussion about alternative assessments, the authors point to the need for such indicators to show validity across a broad range of outcome measures. They comment:

> *Proof of the incremental validity of an alternative predictor would also need to address the whole battery of predictors in use, including the GRE Subject Tests and UGPA in addition to the GRE General Tests. This demonstration of a measure's validity should be especially rigorous and comprehensive when the proposal is to replace a predictor such as the GRE, which has strong validity demonstrated through massive validation efforts. To the best of our knowledge, no alternative predictor of graduate school performance has met all of these rigorous yet important requirements (with perhaps the exception of other standardized cognitive ability measures such as the Miller Analogies Test.*

They are also unapologetic about their support for admissions criteria that are valid predictors of graduate school performance. In contrast to the charges that these standards are unfair, they argue that the use of valid criteria is essential to sound management of human and public capital. In their words:

> *Effective selection and training of graduate students is of critical importance for all fields requiring graduate training. Admission of poorly qualified students misuses the resources of students, faculty, and schools. Failure to admit and retain outstanding candidates ultimately weakens a field.*

The paper is: Kuncel, Nathan R.; Hezlett, Sarah A., and Ones, Deniz S. A Comprehensive Meta-Analysis of the Predictive Validity of the Graduate Record Examinations Implications for Graduate Student Selection and Performance *Psychological Bulletin* 127 (1), 162-181, 2001.

Correspondence may be addressed to Nathan R. Kuncel, Department of Psychology, University of Minnesota, Twin Cities Campus, 75 East River Road, Minneapolis, Minnesota, 55455-0344. Email: kunce001@tc.umn.edu. Those with access to PsycInfo Fulltext may retrieve the entire paper online.

The authors have also completed a meta-analysis of more than 1700 studies on SAT validity. The paper is not yet published, but a University news release provides a brief summary of the findings.