# The Validity of GRE Subject Tests

***By Patricia Hausman***
***Science Insights, January 2002***

A little-noticed paper published last year shows that Graduate Record Examination subject tests are 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.

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 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.

---

**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*

---