

Table 1

Observed and Operational Validities of the Graduate Record Examination (GRE): Reanalysis of Sternberg and Williams's (1997) Data

Criterion and correlation	GRE subscale			
	Verbal	Quantitative	Analytical	Psychology
Research ratings				
r_{obs}	.12	.07	.12	.14
r_{RR}	.18	.15	.17	.21
$\hat{\rho}$.24	.20	.22	.27
Dissertation ratings				
r_{obs}	.08	.07	.24	.15
r_{RR}	.12	.15	.33	.21
$\hat{\rho}$.14	.17	.37	.24
Teaching ratings				
r_{obs}	.15	-.04	.14	.01
r_{RR}	.23	-.08	.19	.02
$\hat{\rho}$.30	-.10	.25	.03
First-year grades				
r_{obs}	.18	.14	.17	.37
r_{RR}	.27	.28	.23	.51

Note. Grades were not corrected for unreliability because no reliability values were reported for this criterion. The computations took into account the differential numbers of raters for different criteria. r_{obs} = observed correlation; r_{RR} = correlation corrected for range restriction only; $\hat{\rho}$ = correlation corrected for both unreliability and range restriction (i.e., operational validity).

116 and 140, respectively. (On the basis of mean GRE scores, Yale University and the University of Minnesota appear to have similar admission requirements for psychology.)

The standard deviations for the Verbal and Quantitative scores from the University of Minnesota sample (native English speakers only) and the standard deviations of the Analytical and Psychology scores from the full Yale University sample were used to correct for range restriction. Standard deviations for the Analytical and Psychology scales for University of Minnesota admissions were not available. This step resulted in corrected correlations of .18, .15, .17, and .21 for prediction of the research performance rating by the GRE Verbal, Quantitative, Analytical, and Psychology scores, respectively.

Finally, on the basis of the data for multiple raters provided by Sternberg and Williams (1997), the interrater reliability of the ratings was .59 for a single rater, which is relatively high for criterion measures of this kind. In applying the corrections for unreliability, we took into account the differential numbers of raters for different criteria. The correlations in the rows marked $\hat{\rho}$ in Table 1 were corrected for both criterion unreliability and range restriction. For the research performance rating criterion, the corrected values were .24, .20, .22, and .27.

These estimates as well as those for the other three criterion variables shown in Table 1 do not support Sternberg and Williams's (1997) contention that "only the GRE Analytical test score was found to predict more consequential evaluations of student performance" (p. 637). Instead, they are reasonably consistent with the well-established cognitive ability, validity generalization literature and would be considered moderately large in selection contexts. The GRE does predict graduate school performance in psychology.

Questions about the relationships between standardized ability measures and graduate school performance are best answered through large-sample studies or properly conducted meta-analyses and not with single studies using relatively small sample sizes. Once before, psychology was confused about how to interpret population parameter estimates from single studies (Schmidt & Hunter, 1977). Sternberg and Williams's (1997) analysis incorporated a number of methodological flaws in parameter estimation and data interpretation that should not be duplicated by others. It would represent a step backward. Failure to correct for restriction of range and unreliability, faith in small-sample research, and reliance on statistical significance testing have led another study astray.

REFERENCES

- Briel, J. B., O'Neill, K. A., & Scheuneman, J. D. (Eds.). (1993). *GRE technical manual: Test development, score interpretation, and research for the Graduate Record Examinations Program*. Princeton, NJ: Educational Testing Service.
- Educational Testing Service. (1996). *Interpreting your GRE general test and subject test scores—1996–1997*. Princeton, NJ: Author.
- Goldberg, E. L., & Alliger, G. M. (1992). Assessing the validity of the GRE for students in psychology: A validity generalization approach. *Educational and Psychological Measurement*, 52, 1019–1027.
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 65, 643–661.
- Sternberg, R. J., & Williams, W. M. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training of psychologists? A case study. *American Psychologist*, 52, 630–641.
- Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. New York: Wiley.

Correspondence concerning this comment should be addressed to John P. Campbell, Department of Psychology, University of Minnesota, Twin Cities Campus, 75 East River Road, Elliott Hall, Minneapolis, MN 55455-0311. Electronic mail may be sent to campb006@tc.umn.edu.

Gatekeeping, Compensation, and Fallibility

John Ruscio
Brandeis University

Sternberg and Williams (June 1997) wrestled with an important question when they considered the extent to which Graduate Record Examination (GRE) scores predict meaningful outcomes in the graduate training of psychologists. Their discussion of the reasons why admissions committees rely on the GRE is informative, and their call for evaluations of its predictive validity may stimulate much-needed research. However, none of the conclusions regarding the utility of GRE scores or alternative assessments of ability in admissions decision making are supported by their study; the impacts of gatekeeping, compensation, and fallible criteria pose insurmountable interpretational problems. Therefore, the empirical contribution of their article to an understanding of the GRE is negligible, and a purely theoretical article would

have been more appropriate and compelling.

Two of the problems in interpreting the results of Sternberg and Williams's (1997) study stem from admissions procedures used at Yale University (and at many other institutions, as noted by Sternberg and Williams). The first problem is that admissions committees fulfill a gatekeeping role, offering admission only to qualified applicants. However, the only way to determine the true efficacy of the GRE as a decision-making tool is through an investigation in which the GRE itself is not a basis for acceptance or rejection. Therefore, Sternberg and Williams can potentially answer only one restricted question: Among students with GRE scores high enough to merit acceptance, do remaining differences in GRE scores predict meaningful outcomes? This question does not reflect the intended purpose of the test nor the reason that admissions committees use it. Instead, the appropriate question is as follows: Does having a GRE score above the cutoff used to offer acceptance, rather than below the cutoff, predict meaningful outcomes?

A second problem stemming from the use of GRE scores in admissions procedures is that of compensation. Describing the admissions criteria by which their sample was selected, Sternberg and Williams (1997) stated, "There are no explicit cutoffs, although the lower the level of the scores, the more an applicant needs compensating factors to gain admission" (p. 631). Such compensatory decision models can result in negative correlations between predictors among accepted applicants (Dawes, 1975). At Yale University, low GRE scores must necessarily be associated with high values on other factors, such as grade point average, letters of reference, and research experience. The prevalence of positive attributes among low GRE scorers not only confounds the results of this study but also stacks the deck squarely against the predictive validity of the GRE.

A third interpretational problem is the fallible nature of the criteria that were used. Faculty ratings of students were made after many years, even decades, had passed. Although Sternberg and Williams (1997) did acknowledge the potential unreliability of faculty ratings because of subjective and retrospective biases, they provided little or no evidence indicating that these ratings were reliable or valid measures of any meaningful outcomes other than the faculty's global impression of each student. For example, the correlation of grades with these ratings should not be surprising, given that the same set of faculty members both assigned the grades and made the ratings. Similarly, the large intercorrelations among all of the ratings provide strong evidence for a halo effect

(Thorndike, 1920). If one views each rating as an imperfect measure of the global impression of each student, the internal consistency computed across the faculty ratings supports such a halo conjecture: Cronbach's alpha equals .92. Given the plausibility of this halo effect on faculty ratings and the fact that low GRE scorers were particularly well-liked by the faculty from the beginning (i.e., they had enough compensating factors to merit acceptance despite low test scores), it might not have been surprising had Sternberg and Williams (1997) obtained negative correlations between GRE scores and faculty ratings.

Although normatively appropriate evaluations were precluded in Sternberg and Williams's (1997) study, a comparison of the predictive validity of the GRE and alternatives within the population of admitted students could have been conducted. To seriously entertain the possibility that alternative assessments of ability, such as tests of creative or practical skills, predict meaningful graduate-training outcomes better than the GRE, these tests could have been given to applicants and their utility then compared with that of the GRE. Other less questionable criteria could have been used as well. For example, Sternberg and Williams could have reported the GRE scores of students who did and did not complete a dissertation, which is certainly one valuable criterion that is less susceptible to halo effects or retrospective distortions than are faculty ratings.

A skeptical attitude toward the utility of the GRE in admissions decision making is admirable, but a well-controlled examination of the predictive validity of the GRE would be extremely challenging to conduct. An "open admissions" policy—offering admission to all applicants—would allow one to answer the appropriate research question by comparing the outcomes of applicants who otherwise would or would not be offered admission on the basis of their GRE scores (see Gilovich, 1991, pp. 42–43). However, offering admission to all applicants is not an attractive option for a graduate program with limited space and a reputation to protect. Likewise, simply disregarding the GRE scores of applicants, a strategy affording a limited analysis of the test's utility, is scarcely more attractive. It would be interesting to know whether any graduate program in psychology is skeptical enough of the GRE to suspend its use in order to assess its true value.

REFERENCES

Dawes, R. M. (1975, February 28). Graduate admission variables and future success. *Science*, 187, 721–723.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York: Free Press.

Sternberg, R. J., & Williams, W. M. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training of psychologists? A case study. *American Psychologist*, 52, 630–641.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25–29.

Correspondence concerning this comment should be addressed to John Ruscio, Department of Psychology (MS 062), Brandeis University, Waltham, MA 02254-9110. Electronic mail may be sent to ruscio@brandeis.edu.

A Case Study Versus Science

Corey E. Miller, Gerald V. Barrett,
and Dennis Doverspike
University of Akron

Sternberg and Williams (June 1997) presented a validation study of the Graduate Record Examination (GRE) that was conducted with a small, highly restricted sample in a unique setting. The substantial literature on meta-analysis and the vast body of literature on the relationship between general ability and performance should lead researchers to be very cautious of overinterpreting any single study finding a lack of relationship between cognitive ability and occupational or academic performance. Sternberg and Williams's study suffers from severe range restriction in the predictor and the criteria and from unreliable criteria.

Considering only the GRE takers who did well enough to apply or be admitted to graduate school is a restriction of range. Yale University's Department of Psychology reports that, on average, slightly less than five percent of applicants are accepted. On the basis of the data presented in Sternberg and Williams's (1997) article and in the *GRE Technical Manual* (Briel, O'Neill, & Scheuneman, 1993), it appears that the average psychology graduate student at Yale University earns a GRE Psychology test score in the 93rd percentile. Although these calculations are admittedly crude, the conclusion that the average psychology graduate student at Yale University is far above average is inescapable.

Linn and Hastings (1984) found that more than half of the Law School Admission Test validity differences observed by law schools were due to differences between the schools in selectivity, with the least selec-