

sis of the predictor variable is heavily biased against that predictor; to the extent that admissions decision-makers made accurate use of the GRE scores, the resultant sample should be relatively free of performance differences associated with GRE-measured ability. Variance in graduate school performance should be attributable to qualities other than the academic aptitude presumably measured by GRE scores. Perhaps a larger problem is the testing of misleadingly simple correlational hypotheses (such as, "Do GRE scores correlate with graduate school performance?"), which do not reflect the multivariate complexity of real-world predictions and performance.

REFERENCE

- 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.

Correspondence concerning this comment should be addressed to Dewey G. Cornell, Programs in Clinical and School Psychology, Curry School of Education, University of Virginia, 405 Emmet Street, Charlottesville, VA 22903-2495. Electronic mail may be sent to dcornell@virginia.edu.

You Proved Our Point Better Than We Did: A Reply to Our Critics

Robert J. Sternberg
Yale University

Wendy M. Williams
Cornell University

We thank our critics for their thoughtful commentaries on our article (Sternberg & Williams, June 1997). Because of our word limit, we cannot answer commentators individually, but can only respond to the major issues they raise. First we reply to local statistical issues raised by these critics, and then to the global issues raised by their comments.

Statistical Issues

1. Restriction of Range

We acknowledged restriction of range in the article (Sternberg & Williams, 1997). But as

we pointed out, standard deviations were substantial, the predictor most restricted in range showed the best prediction, and some correlations (with grades) were actually quite substantial. In reality, virtually all graduate programs display some restriction of range. More selective programs tend to get students from the higher end of the distribution, whereas less selective programs tend to get students from the lower end. By design, we used the real range of students from a real program, not an artificially inflated range from a hypothetical program encompassing the full range of students.

We did an exploration whereby we increased standard deviations by including only (a) the top and bottom quartiles and (b) the top and bottom deciles of the distribution in our analyses (Sternberg & Williams, 1997). Correlations with grades shot up, but correlations with ratings remained meager. With regard to range in criteria, we agree that Yale shows grade inflation, but so do many other programs; our concern is with actual, not hypothetical, criterion measures of performance.

2. Studying Only Students Admitted in Part on the Basis of the Graduate Record Examination (GRE)

Our question (Sternberg & Williams, 1997) was with regard to admitted students, not with regard to all students who applied (which would be an unanswerable question, because we have no way of knowing how well rejected students would have done in our program). We agree that the ideal study would be in a graduate program that would forego the use of GRE scores for the years of the study, but such an ideal study would be extremely hard to do in the real world, largely because many faculty share the sentiments of our critics.

3. Unreliability of Criteria

Correcting for attenuation in criteria would give correlations not for this real world, but for a hypothetical world. To the extent there were halo effects in our ratings (Sternberg & Williams, 1997), so are there in real-world judgments of student and faculty performance.

In the real world of psychology—when candidates are judged for jobs, promotions, and awards—criteria do not show the reliability of standardized psychometric ability tests. Psychologists all work and live in this real world. We (Sternberg & Williams, 1997) looked at the kinds of evaluative criteria that psychologists use in real high-stakes judgments, which embrace implicit assessments of creative and practical as well as analytical

abilities. Just as different faculty members might evaluate different candidates for jobs or tenure depending on their specialty and familiarity with the candidates, so did different Yale faculty evaluate different students.

4. Students Who Speak English as a Second Language

All students in our graduate program do virtually all their reading and writing in English and are judged by the same criterion standards regardless of their native language. Had we omitted (the small numbers of) students speaking English as a second language, we presumably would have been criticized for artificially restricting our range, biasing our sample, and perhaps advocating a double standard in assessment of achievement. Of course, weights assigned to test scores in the admission process—in our program and many others—take into account a number of factors, including native language.

5. Compensation

We agree that admissions indices can be used in a compensatory way. But we can test with our data whether predictors must show negative correlations, because the GREs comprise four distinct, individually considered predictors (verbal, quantitative, analytical, subject matter), not just one. In fact, all intercorrelations were positive (Sternberg & Williams, 1997).

6. Failure Rates

We believe it invalid to conclude, as does one commentator, that low failure rates from our graduate program indicate the GRE's "hit rate of 98%" (Melchert, 1998, p. 574, this issue). These rates could as easily indicate a high hit rate for any other predictor, or for none at all, as the philosophy of the Yale psychology faculty is to be selective in admissions, but then to get admitted students through the program if at all possible.

The Global Issue

The problem is not really with the GRE. It is with a system that, during the years students are in school, emphasizes primarily memory and analytical skills in (a) ability testing, (b) instruction, and (c) assessment, thereby giving measures such as the GRE and related tests higher predictive validity for students in school than these tests will have once the students are out of school, when creative and practical abilities become more important (Sternberg, 1997a; Sternberg, Wagner, Williams, & Horvath, 1995). Consider, for example, how commentators might respond to our article (Sternberg & Will-

iams, 1997), or any article, really, and how these responses would be related to learning and assessment activities based on the triarchic theory of human intelligence (see Sternberg, 1997b; Sternberg & Spear-Swerling, 1996).

Commentators might choose primarily to (a) analyze the GRE, our assessment of the GRE, or our proposed triarchic alternative (Sternberg & Williams, 1997); or to (b) show how we might create a new, appropriate test that is better than either the GRE or our triarchic alternative; or to (c) describe how one might implement these or other tests in a more practical way, for example, better reaching students who speak English as a second language. All of the commentators chose the first option (analytical), admirably demonstrating themselves the skills required for the GRE and for which the educational system selected them (and us) for entrance into graduate study and later the field.

The risk is that such a system may develop students and later professionals with good memory, analytical, and even attack skills, but that the profession will end up with individuals who have the analytical skills metaphorically either to maintain or to tear down existing buildings, but perhaps not the creative and practical skills to build exciting new ones. Perhaps when psychologists start emphasizing creative and practical skills more in ability testing, instruction, and assessment of achievement, and make the typical educational environments more responsive to and nurturing of the range of abilities prized in the world after graduation,

they will become better able to predict real-world success. At the same time, perhaps the second and third options above will appear more attractive to commentators, and all will feel less bound to tests based on turn-of-the-century theories of abilities and how to measure them.

No technology of which we are aware—computers, telecommunications, televisions, and so on—has shown the kind of ideational stagnation that has characterized the testing industry. Why? Because in other industries, those who do not innovate do not survive. In the testing industry, the opposite appears to be the case. Like Rocky I, Rocky II, Rocky III, and so on, the testing industry provides minor cosmetic successive variants of the same product where only the numbers after the names substantially change. These variants survive because psychologists buy the tests and then loyally defend them (see preceding nine commentaries, this issue).

The existing tests and use of tests have value, but they are not the best they can be. When a commentator says that “it will never be possible to improve much on the current admissions policies of Yale and its direct competitors” (Darlington, 1998, p. 572, this issue), that is analogous to what some said about the Model T automobile and the UNIVAC computer. Comments such as this one prove our point (Sternberg & Williams, 1997) better than we ever could.

REFERENCES

- Darlington, R. B. (1998). Range restriction and the Graduate Record Examination. *American Psychologist, 53*, 572–573.
- Melchert, T. P. (1998). Support for the validity for the Graduate Record Examination. *American Psychologist, 53*, 573–574.
- Sternberg, R. J. (1997a). *Successful intelligence*. New York: Plume.
- Sternberg, R. J. (1997b). What does it mean to be smart? *Educational Leadership, 54*, 20–24.
- Sternberg, R. J., & Spear-Swerling, L. (1996). *Teaching for thinking*. Washington, DC: American Psychological Association.
- Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath, J. A. (1995). Testing common sense. *American Psychologist, 50*, 912–927.
- 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.

Preparation of this comment was supported under the Javits Act Program (Grant R206R50001) as administered by the Office of Educational Research and Improvement, U.S. Department of Education. Grantees undertaking such projects are encouraged to express freely their professional judgments. This comment, therefore, does not necessarily represent the positions or policies of the federal government, and no official endorsement should be inferred.

Correspondence concerning this comment should be addressed to Robert J. Sternberg, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205. Electronic mail may be sent to robert.sternberg@yale.edu or to WMW5@cornell.edu.