



Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Intelligence 31 (2003) 331–337



Discussion

Issues in the theory and measurement of successful intelligence: A reply to Brody

Robert J. Sternberg*

Department of Psychology, Yale University, P.O. Box 208358, New Haven, CT 06520-8358, USA

Received 6 February 2002; received in revised form 23 July 2002; accepted 23 July 2002

Abstract

This article constitutes a reply to Brody's analysis of ability measurements based on the triarchic theory of successful intelligence. It discusses a number of issues raised in his critique, such as restriction of range, how much obtained variance can be attributed to *g*, and aptitude–treatment interactions.

© 2002 Elsevier Science Inc. All rights reserved.

1. A rebuttal

1.1. The Sternberg Triarchic Abilities Test (STAT)

Brody's analysis is of an earlier version of the STAT created roughly a decade ago (Sternberg, 1993) that is no longer in use by us. This test had short subtests measuring analytical, creative, and practical abilities in the verbal, quantitative, figural domains by multiple-choice items, as well as in the verbal domain by essay items. I agree with Brody that the test needed a lot of improvement and, indeed, it has been improved. Several concerns led us to improve the test, which overlap with, but are not limited to, Brody's. These concerns were (a) overly high correlations among subtests (Brody's main concern), (b) lack of performance-based measurements beyond essays, which are essentially verbal in nature, (c)

* Tel.: +1-203-432-4633; fax: +1-203-432-8317.

E-mail address: robert.sternberg@yale.edu (R.J. Sternberg).

existence of some items with unsatisfactory psychometric properties, and (d) lack of fit, overall, between the conceptualization of the theory of successful intelligence and the way the theory was operationalized by the earlier form of the test. I will not devote a great deal of space to replying, as the test is outdated (a fact of which Brody was aware when he undertook the analysis). However, I will reply, as I have many reservations about his analysis.

1.2. Corrections of correlation coefficients for restriction of range

I believe that my views on correction of correlation coefficients—for restriction of range—are not identical to those of Brody. Hence, it is worth saying what they are.

The principles behind such corrections are noncontroversial. But the practices are controversial. Many people correct correlations, others do not (see http://www.ucop.edu/sas/research/researchandplanning/pdf/sat_study.pdf; footnote8). Corrections engender assumptions that may seem nonproblematical, but they are not. Are his corrections of our data justified? Several issues arise in answering that question.

First, the corrections make a variety of assumptions. One is that the correlation of two variables is the same throughout the range being tested. But we know from the work of [Detterman and Daniel \(1989\)](#) and others that this assumption is false—that, for example, correlations among psychometric ability tests are, on average, higher in the lower ranges of IQ than in the higher ranges. This result makes sense on the theories of [Sternberg \(1985\)](#) and others (e.g., [Campione, Brown, & Ferrara, 1982](#)), that metacomponential functioning is a more critical differentiator of performance in the lower ranges than in the upper ranges. So, if we were to correct, we probably would need a lesser degree of correction in the upper than in the lower ranges to bring the observed correlation closer to the true correlation.

Second, the correction for restriction of range assumes that one is correcting to the correct population. But what is the correct population? The general assumption appears to be that “the” population is that with an IQ of 100 and a standard deviation of 15. But is it? If, say, one is testing people for managerial jobs or Army leadership jobs or college professorial jobs, as we have done in some of our research, then the range of people being considered for those jobs—the population of interest—is not a population with a mean IQ of 100 and a standard deviation of 15 or 16. The mean is higher and the standard deviation is lower relative to the overall population. People with low IQs typically are not applicants for such jobs and, if they were, they would be quickly screened out, probably without the need for the IQ scores. So, the true validity of the test for predicting performance among applicants to these jobs should not be based on correcting for restriction of range to the average (US) population. Indeed, one does not typically see corrections of SAT scores, despite the fact that the mean of 500 and the standard deviation of 100 do not correspond to the IQ mean of 100 and standard deviation of 15. Why? Because the average IQ of people applying to colleges requiring the SAT is not 100 nor is the standard deviation of the IQs likely to be 15 or 16. Correcting for restriction of range to fit an inappropriate population is, I believe, inappropriate, as it does not reflect the ecological situation for which the test was created.

Second, when a test is created for a specific purpose, it is also created for a certain population with certain characteristics. As psychologists, we do not know how the test would perform in a different population with different characteristics, nor even whether the scores would mean the same thing in a different population. For example, Brody (p. 6) readily applies corrections for restriction of range and attenuation to produce much larger correlations than those we originally obtained. Are the assumptions of his correction valid? In our study (Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999), teachers were asked to nominate students if the teachers believed the students were gifted in any way, not necessarily with respect to *g* or even necessarily academic abilities. Brody suggests a one-third correction for restriction of range, but has no apparent basis for choosing this figure. Participants were selected as gifted in a variety of ways, and it is not clear that any one correction would even apply to all the tests. What is he correcting for restriction in—analytical ability, creative ability, practical ability, or some combination of the three? How would he, or anyone, know what level of correction to apply? The correction is arbitrary, and does not follow in any way from any of our procedures. Hence, I do not believe the correction is credible. I will, therefore, not consider further his data corrected for restriction of range.

Brody proposes that a “crude estimate of the reliability of the STAT measures may be obtained by averaging the two reliabilities” (p. 6), that is, internal-consistency reliabilities with inter-rater reliabilities. He then proceeds to perform various corrections for attenuation based on this crude estimate. But I believe averaging internal-consistency with inter-rater reliabilities is not defensible, because they are measures of different things. One measures whether items consistently assess the same construct, the other whether people consistently assess the same construct. Inter-person and inter-stimulus correlations have no necessary relation to each other at all. For example, people could be perfectly consistent in their responses (person reliability of 1) to items that have no correlation with each other (item internal-consistency reliability of 0). Person and item data are independent. It does not make sense to average person data with item data, even if one calls the average, crude. For example, if in a given sample of data, the participant reliability is 1.00 (all participants consistently give the same responses) but the item reliability is 0.00 (each item measures a totally different construct), what sense does it make to say that the crude average reliability is .50?

The bottom line is that I accept neither Brody’s corrections for restrictions of range nor for attenuation, and hence will not discuss further his analyses based on what I am confident are invalid assumptions.

Brody notes that there were high correlations in the Sternberg et al.’s (1999) study among measures of achievement. I agree. As noted in that paper, there was a halo effect in these ratings data, and hence we did not take seriously convergent-discriminant analysis with the achievement measures. We needed to improve our instructions to scorers on how separately to rate analytical, creative, and practical skill, and since, have done so.

1.3. How much variance does each of the three abilities account for?

In his Table 3, Brody tabulates the percentage of variance accounted for by analytical ability alone and by the three measures of analytical, creative, and practical ability for each of

our dependent measures. He finds that analytical ability alone accounted for about 75% of the variance that was predicted by the three measures. It is possible to look at the results he tabulates in somewhat more detail.

For the assignments and the final examination, which were the more structured dependent measures and required the least independent initiative, analytical ability accounted for 85% of the predicted variance in both cases. For the final project, which was the least structured dependent measure and required the most independent initiative, analytical ability accounted for 61% of the predicted variance. We view these results as in line with the triarchic theory. The more independence and initiative a task requires, the better our newer measures predict. In any case, we do not necessarily expect that our measures will provide a vast improvement in prediction. What we got were predictions that we believe, for practical purposes, are of value. And the predictions were not due just to psychometric g .

1.4. Choosing to attribute results to psychometric g

Brody suggests that our “removing covariances among multiple-choice measures removes the g variance that is present in each of the measures” (p. 11). At this point, he dismisses our essay measures as “probably unreliable indexes” (p. 11).

Our confirmatory factor analysis revealed substantial method variance, which Brody chooses to interpret as construct (g) variance. We do not agree. We believe that multiple-choice tests tend to correlate with each other, above and beyond the constructs measured, because there are certain test-wiseness skills that are relevant to multiple-choice tests. Different test-wiseness skills are needed for other kinds of tests, such as essay tests. But at this point, we have nothing more than a disagreement on a matter of opinion, with no particular basis for his view other than his opinion that the shared variance is g -based variance rather than method variance.

1.5. The aptitude–treatment interaction analysis

Brody claims that “if the F tests are accepted as valid, the data reported in [his] Table 4 appear to be at variance with data obtained in the multitrait–multimethod analysis of the relationship between abilities and achievements” (p. 13). This conclusion is not correct. The data reported in his Table 4 are for different dependent variables *between* treatment conditions, whereas the data he reported earlier are for different dependent variables *within* those treatment conditions. His conflating the two kinds of analyses, which are independent, is responsible for his mistakenly seeing a contradiction. This analysis is a bit reminiscent of his earlier analysis conflating subject and item reliabilities, which also are independent. Because between-treatment and within-treatment effects are independent, the rest of Brody’s analysis of treatment effects will not be considered further. The data show, through analyses of various kinds reported in the articles, aptitude–treatment interactions. His comments are irrelevant to the data, and his interpretations merely speculations consistent with his g -based theoretical position.

1.6. The data

As Brody's analyses are largely based on confluations (of subject and item variance and of between- and within-treatment variance), it is, I believe, better simply to look at the data as reported in his tables, which are based on our tables. Table 1 shows a reasonable, although not perfect, pattern of data. For the Concept Mastery Test, a test of crystallized abilities, the respective correlations of the analytical, creative, and practical ability measures are .49, .43, and .21, in the order that would be predicted. For the Watson-Glaser, the respective correlations are .50, .53, and .32. Here, the analytical and practical correlations are as would be expected, the creative is not. For the Cattell Culture-Fair, which is relatively novel, the correlations are, respectively, .50, .55, and .36, suggesting that the Cattell measures analytical ability but in a more novel way than some tests, but does not measure practical ability as effectively. And the correlations for the Creative Insight Test are .47, .59, and .21, respectively, showing the expected pattern. These results certainly are not perfect, but neither are they much off the mark.

Table 2, as I noted earlier, shows nondifferentiated correlations with overall achievement measures, consistent with the halo effect I noted earlier with regard to our achievement measures.

Table 3, discussed earlier, shows that the creative and practical subtests did add predictive variance, and more so for the relatively novel task (the independent project) than for the more structured tasks (the assignments and final exam). This result also makes sense in terms of the triarchic theory.

Finally, Table 4 shows the aptitude-treatment interaction we predicted.

In sum, if one looks at the data, rather than Brody's reanalyses, which I view as highly questionable for the reasons described above, the data are largely, although certainly not fully, supportive of the triarchic theory, which is why we are working to improve our tests and our instructional procedures.

2. Recent developments

As I have mentioned, validation of the triarchic theory is an ongoing enterprise, not a completed one. We have two new projects that I believe are of some relevance to the ongoing discussion. I mention these new developments not as a rebuttal to Brody's critique, but to show ways in which we are attempting to deal with some of the issues he addresses.

One project is a study we have done at the University of Michigan Business School (UMBS), supported by UMBS, in which we have tested all MBA students in one entering class, and almost all in a second, for the incremental validity of our test of tacit knowledge over and above the GMAT. We found that a test of practical intelligence (acquisition and utilization of tacit knowledge) significantly and substantially increased prediction of first-year MBA grades, second-year MBA grades, and project grades beyond the prediction yielded by

the GMAT. The results of this more recent study are inconsistent with Brody's main theoretical and empirical claims.

The other project is a collaborative study we have done called the Rainbow Project, supported by the College Board, in which we have gone well beyond the earlier STAT in developing measures of analytical, creative, and practical abilities. We have worked at 15 institutions in this study—high schools, community colleges, and 4-year colleges ranging from not very selective to highly selective. Our site collaborators in this study (Richard Duran, Ann Ewing, Edward Friedman, Elena Grigorenko, Jane Halonen, Diane Halpern, Charles Huffman, Linda Jarvin, Laura Maitland, Carol Rashotte, Jerry Rudmann, Brent Slife, Mary Spilis, Carlos Torre, and Richard Wagner), data-analyst collaborators (Karen Schmidt and Jack McArdle), and College Board collaborators (Wayne Camara, Howard Everson, and Amy Schmidt) have diverse points of view. Some have more traditional views of intelligence, others, less traditional views. The data, based on 1023 cases, have been and are continuing to be analyzed both by ourselves, consulting with Brent Bridgeman at the Educational Testing Service, but also independently by a team at the University of Virginia led by McArdle and K. Schmidt, who have no vested interest whatsoever in the theory of successful intelligence.

Recognizing our problems with method variance in the earlier work, we have developed what we believe are better tests of creative and practical abilities that are more faithful to the constructs we are trying to measure than are the earlier tests that Brody analyzed. For example, we measure creative abilities in the same ways as before, but also through captioning cartoons, writing creative short stories, telling creative short stories orally, and designing objects such as greeting cards and company logos using computer software. We measure practical intelligence as before, but also through use of movies in which a test-taker must say what a student confronting a problem in the movie should do, and measures of tacit knowledge for school and generalized job success.

The results are, we believe, very favorable. Factor analyses reveal separate analytical, creative, and practical factors. Hierarchical multiple-regression analyses reveal significant and substantial increments in prediction of freshman grades over and above socioeconomic status, sex, SAT-V, SAT-M, and high school GPA. The results, which are new and not yet published, have been reported at a meeting of the College Board and a meeting of the American Association for the Advancement of Science. The results of this study are inconsistent with Brody's main theoretical and empirical claims.

I predict that, with time, the data will continue to support the notion that there is quite a bit more to intelligence than *g*. There will be some die-hards whose faith in *g* as pretty much the whole story of intelligence will never be shaken. Perhaps, it will require one generation, maybe two, before this group becomes a vastly shrinking minority. Eventually, I predict, the large majority of researchers will be using innovative performance-based assessment devices, different from those that dominated the twentieth century, to complement traditional paper-and-pencil assessment devices, in the search for the full structure of intelligence, including but not limited to psychometric *g*. Of course, my prediction may be wrong. Science is self-correcting, so time will tell.

Acknowledgements

I thank Nathan Brody for the time and effort he has put into providing a thoughtful analysis of some of my work. Brody (2002) is critical of aspects of work regarding the triarchic theory of successful intelligence. This article constitutes a reply.

Preparation of this article was supported under the Javits Act Program (Grant No. R206R000001) as administered by the Office of Educational Research and Improvement, US Department of Education and grant REC-9979843 from the US National Science Foundation. Grantees undertaking such projects are encouraged to express freely their professional judgment. This article, therefore, does not necessarily represent the position or policies of the Office of Educational Research and Improvement or the US Department of Education, and no official endorsement should be inferred.

References

- Brody, N. (2002). Construct validation of the Sternberg Triarchic Abilities Test (STAT): comment and reanalysis. *Intelligence*.
- Campione, J. C., Brown, A. L., & Ferrara, R. (1982). Mental retardation and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 392–490). New York: Cambridge Univ. Press.
- Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence*, *13*, 349–359.
- Sternberg, R. J. (1985). *Beyond IQ: a triarchic theory of human intelligence*. New York: Cambridge Univ. Press.
- Sternberg, R. J. (1993). *Sternberg Triarchic Abilities Test*. Unpublished test.
- Sternberg, R. J., Grigorenko, E. L., Ferrari, M., & Clinkenbeard, P. (1999). A triarchic analysis of an aptitude–treatment interaction. *European Journal of Psychological Assessment*, *15*(1), 1–11.