

Available online at www.sciencedirect.com



Intelligence 31 (2003) 319-329



Construct validation of the Sternberg Triarchic Abilities Test Comment and reanalysis

Nathan Brody

Department of Psychology, Wesleyan University, Middletown, CT 06457, USA

Received 10 February 2001; received in revised form 6 July 2001; accepted 2 October 2001

Abstract

This paper presents an alternative theoretical analysis of several analyses presented by Sternberg and his colleagues of studies designed to validate the Sternberg Triarchic Abilities Test (STAT). The paper contrasts a triarchic theory analysis of the data with one that emphasizes the relevance of g to an understanding of the results obtained by Sternberg and his colleagues. Three relationships are considered: (1) Relationships between triarchic abilities and academic achievements; (3) Relationships among triarchic abilities. It is argued that the g theory is required to understand the relationships obtained by Sternberg and his colleagues.

© 2001 Elsevier Science Inc. All rights reserved.

Keywords: Intelligence; Construct validation; Sternberg triarchic abilities test

1. Introduction

Sternberg and his colleagues published several analyses of two studies designed to assess the construct validity of the Sternberg Triarchic Abilities Test (STAT) (Sternberg, Castejon, Prieto, Hautamaki, & Grigorenko, in press; Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996; Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999). This paper critically evaluates Sternberg's interpretation of the data obtained in these studies. Three issues are considered. (1) What is the relationship between the abilities assessed by STAT and those measured by

E-mail address: nbrody@wesleyan.edu (N. Brody).

other tests of intelligence? What is the relationship between the three abilities assessed by STAT (analytical, creative, and practical) and g? (2) What is the relationship between STAT abilities and measures of academic achievement? (3) How are the STAT abilities related to each other?

The analyses considered here pertain to the corpus of published studies dealing with STAT. Sternberg and his colleagues have developed a new test designed to measure triarchic abilities. The new test differs considerably in format from the first version of STAT. It includes, inter alia, items that involve interpretations of movie clips, pictures, and cartoons and is less tied to traditional multiple-choice formats. While the new test has been developed, no studies using it have been published. Sternberg and his colleagues are using the new instrument in a large-scale validation study. The interpretation presented here of the results obtained using the original version of STAT may or may not be applicable to the revised test.

Sternberg assumes that conventional measures of intelligence are primarily measures of analytical abilities—they fail to assess creative and practical abilities. He believes that the ubiquitous relationship between g and measures of academic achievement is partially attributable to a narrow focus of formal schooling on analytical achievements and the relative neglect of practical and creative intellectual achievements. Several hypotheses may be derived from these assumptions that are relevant to the construct validation of STAT. These assumptions provide one set of predictions about each of the three relationships considered in this paper. The predictions for each of the three relationships are as follows.

Scores on the analytical subtest of STAT should be more substantially related to conventional measures of intelligence than scores on the creative and practical subtests.

Analytical ability is assumed to be predictive of academic achievement in conventional academic settings; it ought to be less predictive of academic achievement for individuals who are exposed to an educational experience that attempts to assess creative and practical achievements as well as analytical achievements. In such an academic setting, each of the abilities assessed by STAT should be predictive of relevant academic achievements. In a multitrait–multimethod analysis of the relationship of triarchic abilities and achievements, abilities and achievements with the same name ought to exhibit higher correlations than abilities and achievements with different names.

Analytical, creative, and practical abilities ought to be relatively independent.

Sternberg and his colleagues obtained STAT scores for a sample of 326 high school students who were nominated as gifted students by their high schools. The version of STAT used consisted of 36 multiple-choice items designed to assess analytical, creative, and practical abilities in each of three content domains—verbal, quantitative, and figural. Four multiple-choice questions were used to assess each of the abilities in each content domain. In addition, each of the abilities was assessed by a single essay question. Scores on STAT consisted of a composite based on the essay and multiple-choice components of the test.

A subset of 199 of these students participated in a summer school program at Yale consisting of a 4-week intensive college-level Psychology course based on Sternberg's textbook that was designed to include emphasis on creative and practical knowledge as well as analytical knowledge. Student achievements were assessed for analytical, creative, and practical knowledge on assignments, exams, and final projects.

2. Empirical outcomes

(1) What is the relationship between triarchic abilities and abilities assessed by other measures of intelligence? Sternberg et al. (1996) report that triarchic abilities are related to scores on four other tests—the Concept Mastery Test, the Watson–Glaser Critical Thinking Appraisal, the Cattell Culture-Fair test of g, and a test of creative insight constructed by Sternberg and his colleagues. Sternberg et al. obtained these correlations from an earlier sample of secondary school subjects attending the Yale summer school program. Table 1 presents the correlations they report.

Table 1 indicates that the STAT abilities are related to abilities assessed by conventional tests. The correlations reported in Table 1 underestimate the relationship between abilities assessed by STAT and those assessed by conventional measures. The STAT tests are not highly reliable and the disattenuated correlations between them and other tests are higher than those reported in Table 1. The sample, by virtue of its selection as a group of "gifted" nominees, is likely to be restricted in range of talent.

The Cattell test is a brief test that is assumed to be a good measure of g whose reliability is .83. Sternberg et al. reported reliabilities for the multiple-choice components of STAT of .63, .62, and .48 for the analytical, creative, and practical subtests, respectively. They reported interrater scoring reliabilities of .69, .58, and .68, for the analytical, creative, and practical essay components of the test, respectively. It should be noted that the former reliabilities are internal consistency measures and the latter are measures of scoring reliability for single items. A crude estimate of the reliability of STAT measures may be obtained by averaging the two reliabilities. The estimated disattenuated correlations between the Cattell test and the three triarchic abilities are .68, .78, and .51, for the analytical, creative, and practical subtests, respectively.

Corrections of the disattenuated correlations for restrictions in range of talent would increase the correlations. The subjects were nominated by their high schools as gifted students. It is unlikely that many, if any, would have IQs below the mean. The actual level of restriction in range of talent is not indicated in the Sternberg et al. papers. Assume that there is a one-third restriction in range of talent in the sample (i.e., the sample has a standard deviation in IQ of 10 rather than the unrestricted value of 15). The disattenuated range corrected correlations between the Cattell test and STAT abilities are .81, .93, and .61 for analytical, creative, and practical abilities, respectively. Two conclusions may be derived from this analysis. The abilities assessed by STAT are substantially related to conventional

Table 1 Correlations between STAT abilities and other measures of intelligence

Other measures	STAT abilities			
	Analytical	Creative	Practical	
Concept Mastery	.49	.43	.21	
Watson-Glaser	.50	.53	.32	
Cattell Culture-Fair	.50	.55	.36	
Creative Insight	.47	.59	.21	

measures of g. Conventional measures of g are not predominantly measures of analytical ability as assessed by STAT. Creative ability as assessed by STAT exhibits a marginally stronger relationship with g than analytical ability as assessed by STAT.

(2) What is the relationship between STAT and academic achievement? An analysis of these relationships is contained in Table 2. Table 2 indicates that the abilities assessed by STAT do not exhibit the requisite pattern of relationships with measures of academic achievement required for evidence of construct validity as assessed by a multitrait–multi-method analysis. Note that abilities and achievements with the same name are not more substantially related to each other than abilities and achievements with different names.

The three measures of achievement were positively correlated (mean r = .72, and the disattenuated mean r = .84). The substantial correlation of the diverse measures of achievement provides an explanation of the results of the multitrait–multimethod analysis of the relationship between triarchic abilities and achievements. If triarchic achievements are substantially related to each other, it is difficult to obtain differential predictive validity for different measures of ability. Although the Introductory Psychology course that the students were exposed to was based on the theoretical assumption that analytical, creative, and practical knowledge were substantially independent, the tests used to assess these types of knowledge led to scores that were substantially related to each other.

The data in Table 2 indicate that analytical ability is substantially related to academic achievement even where measures of academic achievement are obtained that represent an expanded definition of achievement derived from triarchic theory. The relationship between analytical ability as assessed by STAT and the overall assessments of analytical, creative, and practical achievements may be assessed by considering the estimated disattenuated correlations between these measures. Sternberg et al. (1996) reported an estimated reliability of .86 for the overall indices of achievement for scoring reliabilities for all of the essay assessments included in their evaluation of achievement. Using this estimate as a measure of reliability for achievement, the disattenuated correlations between analytical ability and achievements are .57, .57, and .60 for overall analytical, creative, and practical achievements, respectively. These correlations are not corrected for restrictions in range of talent. This analysis indicates that analytical ability is substantially related to all of the academic achievement indices.

Sternberg et al. (1999) reported the results of a set of multiple regression analyses relating triarchic abilities to measures of academic achievement. The measures of academic achievement were based on assessments of analytical, creative, and practical achievements on tests, assignments, and final projects—yielding nine different measures of achievement. In

Table 2 Correlations between triarchic abilities and achievements

Abilities	Overall achievements			
	Analytical	Creative	Practical	
Analytical	.43	.43	.45	
Creative	.38	.38	.45	
Practical	.31	.28	.30	

addition, they obtained measures of overall performance on tests, assignments, and final projects. They performed 12 separate multiple regressions relating analytical, creative, and practical ability scores derived from STAT to each of these measures of achievement. The multiple regression analyses were used to ascertain the independent relationship of triarchic abilities to each of the measures of academic achievement. Creative and analytical abilities had significant independent contributions in 10 of the 12 multiple regression analyses. Practical ability had a significant independent contribution in only 1 of the 12 analyses with beta weights ranging from -0.05 to 0.14. The independent contribution of practical intelligence to academic achievement is relatively small—its aggregate independent contribution accounts for less than 1% of the variance in the various measures of achievement. The significant independent construed by Sternberg and his colleagues as evidence of the predictive validity of STAT.

A comparison can be made between the predictive value of analytical ability as a single variable and the combined predictive value of the three abilities assessed by STAT. Table 3 presents the relevant data. Analytical ability considered by itself accounts for over 75% of the total predictive variance in these measures obtained by a consideration of the three triarchic ability scores. Sternberg et al. (1999) do not indicate whether the R^2 values they report are shrunken multiple correlations adjusted to take account of the fact that they are based on three variables. The shrunken values for the multiple correlations are .01 less than the values that are tabled. The prediction of academic achievement is only marginally improved by a consideration of the combined influence of each of the triarchic abilities as opposed to a prediction derived solely from analytical ability scores.

The regression analyses summarized in Table 3 are not fully informative about the relative magnitude of the contributions of general and specific components of intellect to the

<i>x</i> for an tracence abilities and <i>r</i> for analytical ability				
R^2	r^2			
.13	.12			
.11	.10			
.10	.07			
.12	.08			
.09	.04			
.15	.10			
.11	.09			
.13	.13			
.15	.11			
.121	.093			

Table 3 R^2 for all triarchic abilities and r^2 for analytical ability

prediction of academic achievement. Triarchic abilities are related to conventional measures of g. In addition, subsequent analyses to be reported here indicate that the triarchic abilities are related to each other—a finding that is implied by the suggestion that each is related to general intellectual ability. It would be possible to obtain a comprehensive measure of g based on scores on a battery of conventional tests designed to sample diverse intellectual abilities as well as the STAT abilities. A score on g could be entered as the initial term in a regression equation. Triarchic ability scores might then be added to the regression.

The independent contributions of creative and analytical abilities to the prediction of academic achievement may, in part, be attributable to the possibility that they jointly provide a better estimate of g than either does when considered by itself. A comprehensive regression analysis including a measure of g as an initial term in a regression model would enable one to ascertain the degree to which the contribution of each of the triarchic abilities to the prediction of academic achievement is truly independent of g.

(3) Are the STAT abilities related to each other? Sternberg et al. obtained correlations between analytical and creative abilities of .47, between analytical and practical abilities of .41, and between creative and practical abilities of .37. The disattenuated correlation of analytical and creative abilities is .75, of analytical and practical .66, and of creative and practical .62. The correlations would be increased by corrections for restrictions in range of talent. These data indicate that there is substantial overlap among the triarchic abilities as assessed by STAT. The overlap among the abilities is different for the essay and multiplechoice components of the exam. Multiple-choice items exhibit a median correlation of .52 (disattenuated r = .88). The median correlation for the essay measures is .21. Averaging corrected indices of scoring reliability for these three essays and correcting for attenuation increases the value of the median correlation to .32. Differences in the disattenuated relationships among the essay and multiple-choice components of STAT may be partially attributable to differences in the numbers of items used to obtain an ability score. It is impossible to disattenuate correlations among essay measures of triarchic abilities using indices of internal consistency reliability. The aggregate relationship among triarchic abilities is based on two different relationships-those among multiple-choice measures and those based on essay measures. The theoretical relationship among the triarchic abilities may be slightly higher than that indicated by the disattenuated correlations of the composite scores. If it were possible to correct for internal consistency unreliability for the essay portion of the test, the overall relationship between triarchic abilities might well be higher than that reported here.

Sternberg et al. (1999) used a structural equation model to estimate the relationships among triarchic abilities. The model was based on the assumption that correlations among multiple-choice measures of the triarchic abilities were attributable to method variance. Similarly, correlations among essay measures of the triarchic abilities were also attributable to method variance. The assumptions of this analysis are problematic. Owing to the use of single-item assessments, essay measures are probably unreliable indices. Therefore, they would not be expected to exhibit substantial correlations with each other. Multiple-choice measures of triarchic abilities are relatively reliable and can, in principle, be correlated with each other. Triarchic ability measures are related to conventional measures of intelligence and to measures of g. If each of the triarchic ability measures contains g variance, they should be correlated with each other. Removing covariances among multiple-choice measures removes the g variance that is present in each of the measures.

Sternberg and his colleagues assess the relationship between triarchic abilities measured in two different ways with the covariances among the measures attributable to the method of measurement removed. Construct relevant variance would be demonstrated in this analysis by a relationship between triarchic abilities with the same name assessed in two different ways (essay and multiple choice) with covariances attributable to methods of measurement removed. Their formal model fitting indicated that there is substantial method variance present in the multiple-choice measures. The estimated relationships for analytical, creative, and practical measures are .77, .73, and .70, respectively, indicating that multiple-choice measures of different triarchic abilities are substantially related to each other. If this source of variance is removed from the multiple-choice measures, the relationships between them and essay measures of abilities with the same name with corresponding removal of method variance are .57, .05, and .07, for the analytical, creative, and practical multiple-choice measures, respectively. The relationships among triarchic abilities with covariances attributable to method variance removed are near zero (-.07 for analytical and creative, .00 for analytical and practical, and .06 for creative and practical).

The formal analysis fails to support the construct validity of STAT. For the multiple-choice components of STAT, method variance is a larger source of variance than trait variance for each of the triarchic abilities. When method variance is removed, two of the three triarchic abilities exhibit near-zero relationships with the latent abilities assessed by essay methods. A more direct test of the importance of method and trait variance in the two components of the STAT test could be obtained by using a multitrait–multimethod analysis. In such an analysis, the multiple-choice components of STAT would form the column variables of the matrix and the essay components of STAT would form the row variables. Traits with the same name ought to exhibit high correlations relative to the correlations for different traits assessed in the same way. Sternberg and his colleagues do not provide the relevant data for such an analysis. The results of the structural equation analysis of the relationships among triarchic abilities suggest that the more direct test of the construct validity of STAT as assessed by the multitrait–multimethod matrix analysis based on essay and multiple-choice measures would not provide strong evidence of construct validity.

I interpret the analyses reported by Sternberg and his colleagues as evidence that triarchic abilities are substantially related to each other. Sternberg et al. concluded that triarchic abilities are independent. These different conclusions derive principally from different interpretations of the covariances among multiple-choice measures of triarchic abilities. I believe that the covariances represent g, and given this interpretation, I find the independence of triarchic abilities a direct consequence of the removal of g. There is nothing extraordinary about the finding that abilities are relatively independent if the principal source of relationship among abilities—g—is removed.

There are two additional analyses that provide evidence for the relative independence of the three triarchic abilities. Students were assigned to a group whose scores were considered high on a particular ability if they met both of the following criteria: (1) Their scores were more than one-half a standard deviation above the group average for that ability; (2) Their scores on that ability were more than half a standard deviation higher than their scores on the two other triarchic abilities. There were 112 students who were classified as being high on one of three triarchic abilities. Each of the students was assigned to one of three types of discussion sections for the Psychology course. The sections were based on analytical, creative, or practical modes of instruction. Table 4 presents mean achievement scores for students who were assigned to a section that either matched or failed to match the ability in which they excelled. These data indicate that subjects assigned to a discussion group that did not match their strongest ability.

The data in Table 4 are based on 72 of the 112 subjects considered high on one of three abilities. Sternberg et al. (1999) present the following rationale for the additional reduction in the size of the sample: "In such a small sample, random fluctuations in scores (which might have been due to the impact of nonacademic factors of the YSPP, such as staying up late in the dormitory, etc.) are especially noticeable. In order to control for the impact of the random variance, the data were screened for deviant scores, and these extreme scores were deleted from the analyses" (Sternberg et al., 1999, p. 10). Sternberg et al. used this rationale to remove data derived from over 35% of the subjects. It is not clear whether they deleted the same subjects for all of the analyses or eliminated data from different subjects for different analyses. The criterion used for the decision to eliminate a subject's data is not stated. The means and standard deviations of the data prior to the elimination of subjects in the analysis are not presented. Whether the elimination of subjects changed the magnitude of betweengroup differences in the analyses cannot be ascertained from the results reported by Sternberg et al. The procedure of eliminating subjects would reduce the within-group variance. Sternberg et al. obtained a number of significant F tests for comparisons between the performance of matched and nonmatched subjects. The tests indicate that subjects assigned to

Means for Aptitude \times Treatment interaction analysis					
Groups	Analytical	Creative	Practical		
Better matched					
Assignment 1	0.37	0.28	0.21		
Assignment 2	0.54	0.26	0.18		
Midterm examination	0.50	0.18	0.06		
Final examination	0.24	-0.04	0.05		
Independent project	0.15	0.12	0.30		
More poorly matched					
Assignment 1	0.00	0.09	-0.01		
Assignment 2	-0.09	0.00	-0.08		
Midterm examination	-0.15	-0.07	-0.12		
Final examination	-0.12	0.01	0.06		
Independent project	-0.14	-0.06	0.02		

Table 4 Means for Aptitude \times Treatment interaction analysis

The scores range from -1 to 1. S.D. values range from 0.93 to 1.13

a group in which the mode of instruction matched their strongest ability had higher academic achievement than subjects assigned to a group using a mode of instruction that did not match their strongest ability. The F tests comparing the between- and within-group variance are compromised by the procedure used to eliminate subjects from the analyses. The F tests are based on the assumption that the within-group variability represents random variations that occur among subjects assigned to the same experimental group. Eliminating subjects whose scores were deemed to be deviant creates a biased estimate of the within-group variability.

If the F tests are accepted as valid, the data reported in Table 4 appear to be at variance with data obtained in the multitrait-multimethod analysis of the relationship between abilities and achievements. It is possible to offer a tentative reconciliation of these apparently discrepant analyses. There is abundant evidence indicating that g relates to academic achievement in most settings. If the relationship between triarchic abilities and academic achievement is determined substantially by the g variance contained in STAT, then we would expect that various methods of assessing achievement would all exhibit positive relationships with any triarchic ability measure. Components of variance in STAT that are independent of gdo not directly influence performance on corresponding achievements. As seen in Table 2, there is not the slightest indication of a relationship between a particular triarchic ability and the corresponding academic achievement. Nor is this relationship obscured by an interaction in which individuals perform better on achievements that match their intellectual strengths if they are assigned to a discussion group that matches their strongest triarchic ability. The Aptitude×Instructional interaction that is observed is one in which individuals excel on each of the three types of academic achievements if they are assigned to an instructional group that corresponds to an ability in which they excel. This finding may be interpreted as a motivational effect rather than a conventional ability effect. Individuals might be more motivated by and interested in modes of instruction that are congruent with their intellectual strengths. The motivational interpretation of the interaction is speculative. No measures of motivation were obtained in this research. Whether this hypothesis is correct or not, a significant Aptitude×Treatment interaction would provide support for the assumption that components of variance that are independent of shared covariance among triarchic abilities can have a significant influence on academic achievement. Whether or not significant interaction effects were present in the data collected by Sternberg et al. cannot be determined. I suspect that an analysis of all of the data obtained from the 112 subjects who were high on one of the triarchic abilities might not have obtained convincing evidence for an Aptitude×Instructional interaction.

Sternberg et al. (in press) used confirmatory factor analyses to compare different models of the structure of STAT. In their analyses, they used the multiple-choice data from the gifted sample of American high school students as well as samples of Finnish and Spanish subjects. They compared seven different structural models including three first-order factor models including a single g factor model, a three-factor model based on the three triarchic abilities, and a three-factor models based on the three content factors (verbal, quantitative, and figural). None of the first-order models provided satisfactory fit. They also studied second-order models based on a first-order analysis that included nine first-order factors consisting of each of the triarchic abilities as assessed in each of three ways. They tested three second-order

factor models including a g model in which all of the first-order factors were assumed to load on a single g factor. They also formed two other second-order factor models—one with three content factors and one with three triarchic ability factors. The best fitting model for these data was the one that postulated the existence of three second-order triarchic ability factors. This analysis provides evidence in support of the triarchic theory.

Although the confirmatory analysis reported by Sternberg et al. provides evidence in favor of the theory that generated the test, it should be noted that the second-order triarchic abilities are not independent of each other. Sternberg et al. obtained correlations between analytical and practical factors of .93, between analytical and creative factors of .85, and between practical and creative factors of .72. Thus, the superior fit of the triarchic model occurs only where the triarchic abilities are substantially related to each other. Clearly, a model in which triarchic abilities were constrained to have zero correlations or even moderate correlations would provide a poor fit. The presence of substantial correlations among the latent second-order factors implies that a model that assumed that the second-order factors were related to a single g factor on a third level would provide an adequate description of the structure of STAT.

The model fitting analysis does not rule out the present of g variance in STAT. Owing to the strong relationships among the second-order factors, the model is compatible with the assumption that the g variance in the test (i.e., the shared relationship among the independent latent factors) is larger than the components of variance in the latent traits that are independent of each other.

The confirmatory analysis used by Sternberg et al. does not provide an ideal method of ascertaining the presence of g variance in the test. An analysis of STAT and several conventional measures of intelligence would provide additional information about the locus of STAT abilities in the taxonomic structure of intelligence. Sternberg et al. (in press) assert that the ubiquitous evidence for g contained within Carroll's (1993) comprehensive investigation of the taxonomy of intellectual abilities is attributable to his failure to analyze creative and practical intellectual abilities. Ideally, this claim should be tested by confirmatory analyses of conventional measures found to support a g factor and STAT measures. If Sternberg is correct, analytical ability ought to have a different locus within the taxonomy of abilities than creative and practical abilities. The first ability should be highly g loaded, the latter two abilities should not.

3. A concluding comment

The presence of g variance in STAT is directly supported by three different analyses. Each of the STAT abilities is related to a conventional measure of g—the Cattell test. The structural equation model indicates that there are substantial correlations among the multiple-choice components of STAT and that when this source of variance is excluded from this part of the test, two of the three abilities have near-zero relationship with corresponding essay measures. Finally, the confirmatory analysis of data collected from three samples achieves a satisfactory fit only where the STAT abilities are substantially correlated with each other. The mean

correlation among the STAT abilities in this confirmatory analysis is .855, implying that the covariance among STAT abilities is larger than independent sources of variance in the STAT abilities. A principal component analysis of these correlations yields a first principal component that accounts for 89% of the explained variance.¹ The loadings of analytical, creative, and practical factors on this component are .984, .939, and .906, respectively. If this component is interpreted as g, then g constitutes the largest source of variance on STAT.

References

Carroll, J. B. (1993). Human cognitive abilities: a survey of factor-analytic studies. New York: Cambridge.

Sternberg, R. J., Castejon, J. L., Prieto, M. D., Hautamaki, J., & Grigorenko, E. L. (in press). Confirmatory factor analysis of the Sternberg Triarchic Ability Test (multiple-choice items) in three international samples: an empirical test of the triarchic theory of intelligence. *European Journal of Psychological Assessment*.

Sternberg, R. J., Ferrari, M., Clinkenbeard, P., & Grigorenko, E. L. (1996). Identification, instruction, and assessment of gifted children: a construct validation of a triarchic model. *Gifted Child Quarterly*, 40, 129–137.

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, 3–13.

¹ The principal component analysis was performed by Professor Arthur Jensen who reviewed this paper.