Available online at www.sciencedirect.com

Intelligence 32 (2004) 95–107

INTELLIGENCE

Just one *g*: consistent results from three test batteries

Wendy Johnson*, Thomas J. Bouchard Jr., Robert F. Krueger,
Matt McGue, Irving I. Gottesman

Department of Psychology, University of Minnesota—Twin Cities, 75 East River Road, Minneapolis, MN 55455, USA

Received 8 April 2003; received in revised form 27 May 2003; accepted 15 June 2003

Abstract

The concept of a general intelligence factor or *g* is controversial in psychology. Although the controversy swirls at many levels, one of the most important involves *g*'s identification and measurement in a group of individuals. If *g* is actually predictive of a range of intellectual performances, the factor identified in one battery of mental ability tests should be closely related to that identified in another dissimilar aggregation of abilities. We addressed the extent to which this prediction was true using three mental ability batteries administered to a heterogeneous sample of 436 adults. Though the particular tasks used in the batteries reflected varying conceptions of the range of human intellectual performance, the *g* factors identified by the batteries were completely correlated (correlations were .99, .99, and 1.00). This provides further evidence for the existence of a higher-level *g* factor and suggests that its measurement is not dependent on the use of specific mental ability tasks.

Keywords: *g* factor; Mental ability battery; Intelligence assessment; General intelligence

1. Introduction

That performances of individuals on tests of different mental abilities are positively intercorrelated is a well-established fact. The most frequently offered explanation for these intercorrelations is the existence of an underlying general intelligence factor, commonly known as Spearman's *g* (Jensen, 1998). There is considerable empirical evidence for the existence of such a factor, and for a hierarchical structure of mental abilities (Carroll, 1993; Gustafsson & Undheim, 1996) with the *g* factor at the top, yet arguably no other concept in psychology has generated more controversy. Broadly generalized, the controversy

---

* Corresponding author. Tel.: +1-952-473-1673; fax: +1-952-473-1998.
E-mail address: john4350@tc.umn.edu (W. Johnson).

0160-2896/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/S0160-2896(03)00062-X

stems from the social implications of the existence of measurable individual differences in a personal characteristic that is highly predictive of a broad range of life outcomes (Gottfredson, 1997). The controversy takes place both at the level of social policies regarding measurement and intervention (e.g., Baumeister & Bacharach, 1996; Blair, 1999; Scarr, 1996; Sternberg, 2000; Williams, 2000) and at the level of theories of genetics, biochemistry, neurology, cognitive science, evolution, intelligence, and latent variable measurement (e.g., Elman et al., 1996; Eysenck, 1994; Haier, 1993; Horn, 1998; Huttenlocher & Dabholkar, 1997). The questions related to measurement have focused on two fundamental issues: the accuracy and equity with respect to the individual of the measurement instruments that have been developed, and the relevance of our measurements to the structure and, ultimately, biology of human intellectual abilities.

To the extent that there is a *g* factor, *g* factor measurements among individuals should be independent of the specific mental ability tests used to define the factor. That is, the nature of the general factor should be uniform from test battery to test battery. If the nature of the *g* factor changes with the composition of the test battery, the factor analytic approach to identifying structure is arbitrary, and so are the factors identified through its use (Horn, 1989). Put simply, the question is whether there is only one *g*. This question has been investigated from two perspectives.

The first perspective concerns the degree to which the *g*-loading of a particular test is inherent in the nature of the test and thus stable as the test is inserted in different sets of other tests. For example, Thorndike (1987) created six different test batteries, each consisting of eight different tests developed to assess vocational aptitude among military recruits. He separately embedded each of 17 additional tests in each of the six batteries, and then correlated the factor loadings from the inserted tests across the six groups. The correlations ranged from .52 to .94 with a median of .83, which he interpreted as providing evidence that the *g*-loading of a test is relatively independent of the battery of tests in which it appears, and thus a characteristic of the test itself. Vernon (1989) obtained similar results using *g* factors from several intelligence tests and a battery of reaction time tests. Thorndike focused on the evaluative properties of the individual test, which can of course be important, especially in situations where practicalities dictate that only a single test will be used to evaluate individual differences in general cognitive ability and their associations with other variables. On the other hand, Vernon focused on demonstration of a consistent relationship between various psychometric measurements of *g* and a physiological property.

The second perspective concerns the degree to which the *g*-loadings for tests depend on the particular method of factor analysis used to extract the *g* factor. For example, Ree and Earles (1991) extracted *g* factors using unrotated principal components, unrotated principal factors, and hierarchical analyses based on principal components and principal factors, extracting three to eight first-order factors for each. They then calculated scores on the *g* factors resulting from each extraction, and correlated them. The correlations ranged from .930 to .999, which they interpreted as evidence that the *g* loading of a test is independent of the method used to extract the *g* factor. Jensen and Weng (1994) obtained similar results using both simulated correlation matrices in which the true *g* was known, and empirical data from batteries of diverse mental tests. As Ree and Earles pointed out, however, these results do nothing more than provide instantiations of Wilks' (1938) theorem, which states that the correlation of two linear composites of multiple variables will approach 1.00 when the variables are all positively correlated and the weights in the composites are all positive.

Neither of these approaches directly addresses the nature of the *g* factor on which each test is loading, or the consistency of its nature from one battery of mental ability tests to another. If the *g* factors we extract from mental ability tests do in fact have some intrinsic meaning in the manifestation of intelligence, it is critical that these factors be identifiable with consistency in different contexts. There are a number of different batteries of mental ability tests in general use for the assessment of intelligence. These batteries were developed by different groups of individuals, but all were developed with the general goal of including tasks that assess a wide range of human abilities. They include different specific tasks and emphasize different types of tasks to varying degrees, reflecting their developers' differing conceptions of the structure of mental abilities. We suggest that an important test of the significance of the *g* factor is whether or not the *g* factors that can be extracted from scores on the tests in these batteries are the same across batteries. This test is important both to the conceptualization of human intelligence and to our ability to measure it. To carry out such a test, we made use of scores from three mental ability batteries in a heterogeneous sample of 436 adult individuals.

2. Method

2.1. Research participants

The 436 (188 males, 248 females) research participants for this analysis came from the Minnesota Study of Twins Reared Apart (MISTRA), a comprehensive sample of adult twins reared apart that also includes some adoptive and biological family members, friends, partners, and spouses of the twins. In most cases, the twins were separated early in life, reared in adoptive families, and not reunited until adulthood. They came from a broad range of occupations and socioeconomic backgrounds, several different countries, and ranged in age from 18 to 79 (mean = 42.7). Education levels varied from less than high school to postgraduate experience. The sample included 128 twin pairs, 2 sets of triplets, 117 spouses of twins, and 57 other biological and adoptive family members of the twins. MISTRA was initiated in 1979 and continued until 2000, with some participants returning for a second assessment 7 to 12 years after the initial one. The assessment consisted of a weeklong battery of psychological and medical tests. Details of recruitment and assessment are reported by Bouchard, Lykken, McGue, Segal, and Tellegen (1990) and Segal (2000). It included three batteries of cognitive ability tests, along with assessments of numerous psychological traits (personality, interests, attitudes, etc.) and medical and physical traits. We describe each of the three cognitive ability batteries in turn.

2.2. Measures

2.2.1. Comprehensive Ability Battery (CAB)

The CAB was developed by Hakstian and Cattell (1975). It consists of 20 primary ability tests developed with the goal of measuring a broad range of well-replicated primary abilities. To keep administration manageable, each test is short, requiring only 5 to 6 min. To avoid duplication of tasks in the extensive MISTRA assessment and make maximal use of available time, six of the tests in the CAB were not administered to the participants. In addition, for this analysis we eliminated the test of Esthetic Judgment as we judged it not directly relevant to cognitive ability. The tests included in our version are described briefly in Table 1. Hakstian and Cattell (1978) reported split-half and retest reliabilities from the tests ranging from .64 for Perceptual Speed and Accuracy to .96 for Memory Span. As the Verbal

Table 1
Tests included in the three batteries

Test	Assessment activity
Comprehensive Ability Battery	
1. Numerical Ability	Computations including fractions, decimal divisions, square roots, etc.
2. Spatial Ability	Interpretation of two-dimensional figural rotation or reversal.
3. Memory Span	Recall of digits presented aurally.
4. Flexibility of Closure	Identification of embedded figures.
5. Mechanical Ability	Identification of mechanical principles and tools.
6. Speed of Closure	Completion of gestalt.
7. Perceptual Speed	Evaluation of symbol pairs.
8. Word Fluency	Production of anagrams.
9. Inductive Reasoning	Identification of pattern in sequences of letter sets.
10. Associative Memory	Rote memorization of meaningless pairings.
11. Meaningful Memory	Rote memorization of meaningful pairings.
12. Verbal—Vocabulary	Multiple choice among possible synonyms.
13. Verbal—Proverbs	Interpretation of proverbs.
14. Spelling	Multiple-choice identification of misspellings.
Hawaii Battery with Raven	
15. Card Rotations	Matching of rotated alternatives to probe.
16. Mental Rotation	Identification of rotated versions of two-dimensional prepresentation of three-dimensional objects.
17. Paper Form Board	Outline of cutting instructions to form the target figure.
18. Hidden Patterns	Identification of probe figures in more complex patterns.
19. Cubes	Identification of matched figures after rotation.
20. Paper Folding	Identification of unfolded version of a folded probe.
21. Raven	Identification of analogous figure to follow a sequence of figures.
22. Vocabulary	Multiple choice among possible meanings.
23. Subtraction/Multiplication	Completion of two-digit subtractions and two-digit by one-digit multiplications.
24. Word Beginnings/Endings	Generation of words beginning and ending with specified letters.
25. Pedigrees	Identification of familial relationships within a family tree.
26. Things Categories	Generation of things that share assigned characteristics.
27. Different Uses	Generation of novel uses for specified objects.
28. Immediate Visual Memory	Recall of illustrations of common objects immediately following presentation.
29. Delayed Visual Memory	Recall of illustrations of same common objects after delay.
30. Lines and Dots	Trace of a path through a grid of dots.
31. Identical Pictures	Identification of alternative identical to probe.
Weschler Adult Intelligence Scale	
32. Information	Recall of factual knowledge.
33. Comprehension	Explanation of practical circumstances.
34. Vocabulary	Free definition.
35. Coding	Identification of symbol–number pairings.
36. Arithmetic	Mental calculation of problems presented verbally.
37. Similarities	Explanation of likenesses between objects or concepts.
38. Digit Span	Recall of spans of digits presented aurally, both forwards and backwards.

Table 1 (*continued*)

Test	Assessment activity
Weschler Adult Intelligence Scale	
39. Picture Completion	Identification of parts missing in pictures of common objects.
40. Block Design	Reproduction of two-dimensional designs using three-dimensional blocks.
41. Picture Arrangement	Chronological sequencing of pictures.
42. Object Assembly	Reassembly of cut-up figures.

Ability test consists of two completely separable tasks, we tabulated the scores on the two parts separately, which meant that we had a total of 14 test scores; the tests omitted from the battery included Auditory Ability, Originality, Representational Drawing, Aiming, Spontaneous Flexibility, and Ideational Fluency.

2.2.2. The Hawaii Battery, including Raven's Progressive Matrices (HB)

The HB was developed to assess familial resemblance in cognitive ability in the Hawaii Family Study of Cognition (DeFries et al., 1974; Kuse, 1977). The HB consists of 15 tests of primary abilities; each test is short, requiring 3 to 10 min for administration. To avoid duplication of tasks and to make maximal use of available time, two tests in this battery were not administered. In MISTRA, the battery was supplemented with four tests from the Educational Testing Services in order to better identify likely factors, so there were 17 tests in the battery in total. The Hawaii study included a printed and shortened version of the Raven Progressive Matrices Test (Raven, 1941). MISTRA utilized an untimed version of the Raven presented via slides.

The tests in our version of the HB are described briefly in Table 1. The tests added were Cubes, Paper Folding, Identical Pictures, and Different Uses. The tests not administered to our sample were Number Comparison and Social Perception. Internal consistency and retest reliabilities for the tests ranged from .58 for Immediate Visual Memory to .96 for Vocabulary (Kuse, 1977).

2.2.3. The Weschler Adult Intelligence Scale (WAIS)

The WAIS (Weschler, 1955) is probably the best known and most widely used individually administered test of general intellectual ability. Weschler believed that intelligence involved both abstract reasoning and the ability to handle practical situations involving performance and manipulative skills; thus, the WAIS includes Verbal and Performance subcomponents. The subtests were also chosen to be suitable over a wide range of ages and for both sexes and to be appealing to examinees in the sense that they were not tedious or irrelevant. There are 11 subtests of the WAIS, and they are also described briefly in Table 1. Internal consistency reliabilities range from .79 for Comprehension to .94 for Vocabulary (Weschler, 1955). For this sample, average WAIS full-scale IQ was 118.5, normed at the 1955 level. The standard deviation was 19.8.

2.3. Procedure

The tests were largely administered in blocks lasting 60 to 90 min across the 6 days of assessments. Because the full assessment administered by MISTRA was so extensive, some participants did not complete all of the tests. The most common reason for this was that some participants required more time than average to understand test instructions and move from test to test. These participants were also more likely to receive lower than average scores on the cognitive ability tests they did take, which means that their data were not missing at random. For this reason, we did not apply maximum likelihood estimation to the incomplete data set (Little & Rubin, 1987). Instead, to maximize the sample size available to us, we imputed some of the missing data.

To do this, we utilized the hot deck multiple imputation procedure in the computer program LISREL 8.53 (Joreskog & Sorbom, 2002). This procedure compares the designated scores for the case to be imputed to those for the others available, and assigns to the missing data point the score received by the case that has the smallest sum of squared differences from the case to be imputed for the designated scores. As with maximum likelihood estimation, the procedure relies on the assumption that data are missing at random, but imputations are unbiased when the probability of missingness is related only to data that have been supplied (Collins, Schafer, & Kam, 2001). It seemed reasonable in our situation to make this assumption; we therefore imputed missing data for some tests using the five most highly correlated scores from other tests.

In order to minimize potential bias resulting from imputation, we developed three rules to guide this process: (1) Each time we imputed a data point, we deleted all family members of the case to be imputed from the data file and then reinstated them to the file for the next imputation. We thus carried out a separate imputation procedure for each data point imputed. We did this because we did not want our procedure to increase overall correlations among the test scores due to common influences on the scores of family members. (2) For the same reason, we did not impute data for both members of a twin pair—if both were missing, we left them missing. (3) We did not impute missing values for any case missing five or more test scores within a battery. In total, 69 data points were imputed out of the total possible of 18,312. This resulted in 307 individuals with complete data for all batteries. To remove arbitrary variance relationships due to age and sex effects, we carried out all our analyses using data corrected for their effects.

We next conducted exploratory factor analyses using the software program CEFA (Browne, Cudeck, Tateneni, & Mels, 2001) in order to develop second-order factor models independently for each of the three test batteries. In doing this, we made no adjustment for the correlated nature of the observations for the twin pairs within our sample. This should have the effect of inflating the model fit statistics so that the model appears to fit more closely, but should have little effect on parameter estimates (McGue, Wette, & Rao, 1984; Neale, 2003, personal communication). We used ordinary least squares factor extraction, with Infomax rotation and Kaiser row weights. For the CAB and the HB, we carried out these analyses in a completely exploratory manner, meaning that we evaluated the number of factors to be extracted using the Root Mean Square Error of Approximation (RMSEA) (Browne & Cudeck, 1992), and made the assignments of tests to factors on the basis of factor loadings rather than according to any theoretically based criteria. For the WAIS, because more work has been done to develop an accepted factor structure, we imposed the structure of three correlated factors originally identified by Cohen (1957). We used the models we developed to carry out maximum likelihood confirmatory factor analyses of each model separately using LISREL 8.53, fixing the first factor loading to 1.00 for each factor in order to identify the models. We then combined the models for each battery, maintaining the separate structure developed for each battery including extracting separate second-order *g* factors for each. The key result in our analysis was then the correlations among the three *g* factors.

3. Results

3.1. Individual battery models

3.1.1. Comprehensive Ability Battery

For the CAB, we extracted five correlated first-order factors, which we named Numerical Reasoning, Figural Reasoning, Perceptual Fluency, Memory, and Verbal. We chose five factors by examining several possible numbers of factors and choosing the solution that caused RMSEA to be less than .05 (indicating a close fit according to Browne & Cudeck, 1992) and provided the most clearly interpretable solution. There was no factor that seemed to identify spatial ability specifically. The one clearly spatial ability test, Spatial, loaded with Mechanical and Flexibility of Closure on the Figural Reasoning factor. The correlations among the factors ranged from .26 for Figural Reasoning and Memory to .88 for Numerical Reasoning and Perceptual Fluenct. The first-order factor loadings on the second-order *g* factor ranged from .50 for Memory to .98 for Numerical Reasoning. RMSEA for the model was .031 ($\chi^2 = 86.2$, $df = 62$, $P = .023$).

Our factor solution can be compared with that of Hakstian and Cattell (1978). They extracted six factors, though the tests loading on the sixth factor were primarily those not administered to our sample. Thus, it is reasonable to conclude that we agreed on the number of factors comprising the battery. The contents of the two sets of five factors, however, differed somewhat. The biggest differences surrounded the Induction test, which loaded with Number in Hakstian and Cattell's sample but not in ours, and the Verbal Ability test, which loaded with Mechanical Ability alone in Hakstian and Cattell's sample. We separated the Verbal Ability test in our sample into its component parts Vocabulary and Proverbs, and these subtests, along with Spelling, defined our Verbal factor. There was no factor in Hakstian and Cattell's solution that specifically identified verbal ability. We note that their sample consisted of 280 high school students. Thus, our solution, based on a sample both larger in size and more heterogeneous in age and background, is probably more representative than theirs.

3.1.2. Hawaii Battery

We also extracted five correlated first-order factors for the HB, using the same process as for the CAB. We named the factors Logical Reasoning, Spatial, Fluency, Visual Memory, and Patterns. There was no factor that seemed to identify specifically verbal knowledge. Vocabulary and Word Beginnings and Endings were the only verbally oriented tests in the battery. They loaded on the Logical Reasoning factor, along with the Raven, Subtraction and Multiplication, and Pedigrees. The correlations among the factors ranged from .33 between Spatial and Visual Memory to .74 between Logical Reasoning and Patterns. The first-order factor loadings on the second-order *g* factor ranged from .46 for Visual Memory to .88 for Patterns. RMSEA for the model was .050, indicating a close fit ($\chi^2 = 208.1$, $df = 109$, $P < .001$).

This factor solution can be compared to that of Kuse (1977). He extracted only four factors, which he labeled Spatial, Verbal, Speed, and Memory. His Speed factor was defined primarily by Subtraction and Multiplication and Number Comparison; because we did not have Number Comparison in our battery, this factor did not emerge in our solution. On the other hand, we had Different Uses, an additional fluency measure, so we did get a factor defined primarily by the two fluency tests. The other additional factor in our solution was Patterns, which was defined primarily by Identical Pictures, along with Lines and Dots and Hidden Patterns. As Identical Pictures was one of the tests with which we supplemented the original HB, it seemed reasonable that this factor emerged. In addition, the five-factor solution fit our data significantly better than did the four-factor solution.

3.1.3. Weschler Adult Intelligence Scale

We extracted the generally accepted three factors—Verbal Comprehension, Freedom from Distraction, and Perceptual Organization—for the WAIS. The correlations among the factors ranged from .64 between Freedom from Distraction and Perceptual Organization to .72 between Verbal Comprehension and Perceptual Organization. The first-order factor loadings on the second-order *g* factor ranged from .78 for Freedom from Distraction to .88 for Verbal Comprehension. RMSEA for the model was .061 ($\chi^2 = 96.6$, $df = 40$, $P < .001$).

3.2. Combined model

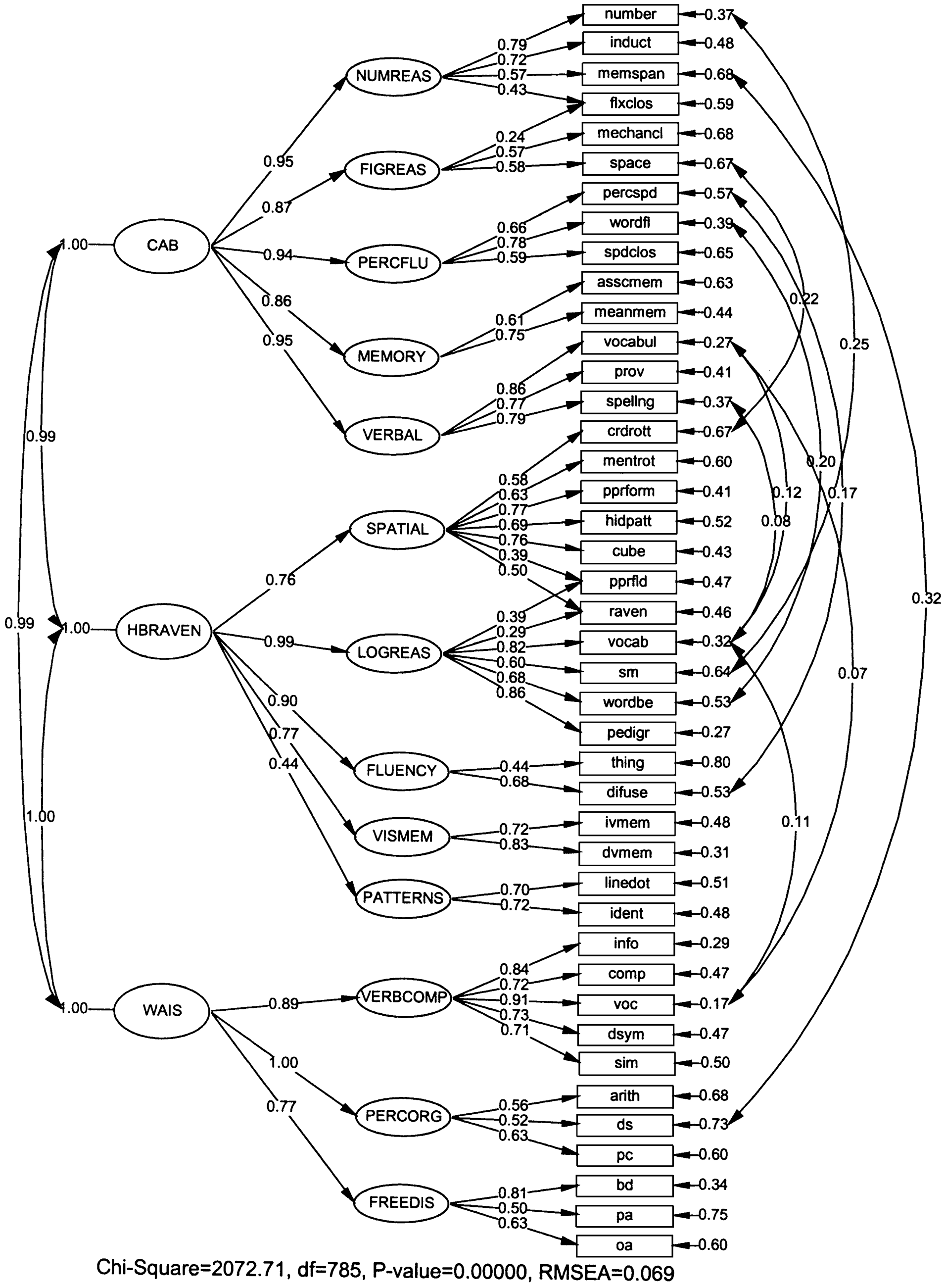
Fig. 1 diagrams the model we fit. RMSEA indicated a reasonable fit (.069, $\chi^2 = 2072.7$, $df = 785$, $P < .001$). As the figure shows, the correlations among the *g* factors for the three batteries ranged from .99 to 1.00.

This model was highly restricted because we constrained each test to load only on the factors extracted for the battery to which it belonged. There was considerable unique variable variance within each battery. Examination of modification indices revealed that some of this was common variance not included in the individual factor models for any of the batteries. We allowed several residual correlations in order to clarify the sources of variance not included in any of the batteries, and the residual correlations are shown in Fig. 1. The sources of common variance across batteries can be summarized as numerical facility, shape perception, memory span, processing speed, verbal facility, and word knowledge. These sources of variance may have defined additional factors had we relaxed the model constraints imposed by the battery structure. We plan to develop such a model as part of a future study of the genetic and environmental influences on special and general mental abilities.

Some of the unique variable variance within the model for each battery was common variance in the models for other batteries. Though the model we report does not include any factor loadings across batteries, we reviewed the modification indices in order to clarify the sources of variance included in some batteries but not others. Similarities from the WAIS had the largest modification indices. This test shared variance with all five CAB factors and with the Logical Reasoning and Fluency factors from the

---

Fig. 1. Second-order model fitted. CAB = Comprehensive Ability Battery, HBRAVEN = Hawaii Battery with Raven, WAIS = Weschler Adult Intelligence Scale, NUMREAS = Numerical Reasoning, FIGREAS = Figural Reasoning, PERCFLU = Perceptual Fluency, LOGREAS = Logical Reasoning, VISMEM = Visual Memory, VERBCOMP = Verbal Comprehension, FREEDIS = Freedom from Distraction, PERCORG = Perceptual Organization, number = Numerical Ability, space = Spatial Ability, memspan = Memory Span, flxclos = Flexibility of Closure, mechancl = Mechanical Ability, spdclos = Speed of Closure, percspd = Perceptual Speed, wordfl = Word Fluency, induct = Inductive Reasoning, asscmem = Associative Memory, meanmem = Meaningful Memory, vocabul = Verbal—Vocabulary, prov = Verbal—Proverbs, spellng = Spelling, crdrott = Card Rotations, mentrot = Mental Rotation, pprform = Paper Form Board, hidpatt = Hidden Patterns, cube = Cubes, pprfld = Paper Folding, raven = Raven's Progressive Matrices, vocab = Vocabulary, sm = Subtraction/Multiplication, wordbe = Word Beginnings and Endings, pedigr = Pedigrees, things = Things Categories, difuse = Different Uses, ivmem = Immediate Visual Memory, dvmem = Delayed Visual Memory, linedot = Lines and Dots, ident = Identical Pictures, info = Information, comp = Comprehension, voc = Vocabulary, dsym = Coding, arith = Arithmetic, sim = Similarities, ds = Digit Span, pc = Picture Completion, bd = Block Design, pa = Picture Arrangement, oa = Object Assembly.



HB. There were interesting interrelationships among the numerical tests in some batteries and the spatially oriented factors in others. Arithmetic from the WAIS and Subtraction−Multiplication from the HB shared variance with the Figural Reasoning factor from the CAB (but not with the Numerical Reasoning factor from the CAB). Object Assembly from the WAIS also shared variance with this factor, and Arithmetic also shared variance with the Spatial factor from the HB. Vocabulary from the HB shared variance with both Verbal from the CAB and Verbal Comprehension from the WAIS: This made sense as there was no strong verbal knowledge factor on the HB. When we did relax these constraints across batteries, the correlations among the *g* factors did not decrease, though the overall model fit improved.

As a test of the robustness of our conclusions, we fit a model with only a single *g* factor for each battery (results not shown). Again, the three *g* factors were completely correlated, but the model fit significantly more poorly (RMSEA=.104, $\chi^2 = 3783.7$, $df = 807$, $P < .001$). At the same time, a model with the same first-order structure as the one we present but with only a single *g* factor did not fit significantly differently (RMSEA=.069, $\chi^2 = 2078.5$, $df = 788$, $P < .001$, results not shown).

4. Discussion

Our analyses indicate that *g* factors from three independently developed batteries of mental ability tests are virtually interchangeable. This is in spite of the fact that the batteries emphasize somewhat different aspects of mental ability, and, though there are some similarities among them, none of the tasks we administered directly overlap. (The most similar tasks are the multiple-choice synonym Vocabulary tasks from the CAB and HB, and even there the words included differ in the two forms. WAIS Vocabulary is administered orally and requires free response definitions. Other similarly titled tasks differ in specific form or method of administration.) The CAB emphasizes inductive reasoning and verbal knowledge while the HB emphasizes nonverbal reasoning and pattern visualization, both in a multiple-choice format. The WAIS is relatively balanced in these areas, but generally elicits free responses. Thus, it seems unlikely that the correlations among the *g* factors arose either because of common task content or common examination methods, strictly construed. In addition, these correlations were rather impervious to the specific constraints on loadings of tests on the first-order factors within batteries. We relaxed several combinations of factor constraints and changed the factor loadings of several tests without altering the correlations among the second-order *g* factors (results not shown). It is possible, of course, as Garlick (2002) suggested, that these tests merely assess abilities for which there is little variance in exposure. If as he proposed, however, neural plasticity underlies intellectual performance, the use of tests that tap abilities to which exposure should be relatively uniform would be the best way to assess differences in neural plasticity, and neural plasticity would be a plausible biological mechanism to explain the *g* factor we observed.

These results provide the most substantive evidence of which we are aware that psychological assessments of mental ability are consistently identifying a common underlying component of general intelligence. This evidence addresses both the question of the existence of a general intelligence factor and the question of the accuracy of its measurement. It does not, of course, address the extent of associations between general intelligence and other life outcomes, nor does it shed light on the possible biological basis of the factor. In addition, our results make clear again that the general factor does not capture all aspects of mental ability, and in particular, that the general factor is an intrinsically higher-order concept drawing together distinctive primary facets. The aspects not measured by the general factor should not be considered simply “noise.” There are substantive correlations among these aspects from battery to battery, and different tests are able to measure them with reliability comparable to that associated with the general factor. Another way to look at this is to note that the high factor loadings of the first-order factors on the second-order *g* factors demonstrate that the general factor accounts for substantial variability in mental ability (perhaps half is a reasonable generalization, as Carroll, 1993,

Table 2

g loadings for the tests in the three batteries

Test	*g* loading
Comprehensive Ability Battery	
1. Numerical Ability	.75
2. Spatial Ability	.51
3. Memory Span	.54
4. Flexibility of Closure	.62
5. Mechanical Ability	.50
6. Speed of Closure	.56
7. Perceptual Speed	.62
8. Word Fluency	.73
9. Inductive Reasoning	.68
10. Associative Memory	.53
11. Meaningful Memory	.65
12. Verbal—Vocabulary	.82
13. Verbal—Proverbs	.73
14. Spelling	.75
Hawaii Battery with Raven	
15. Card Rotations	.44
16. Mental Rotation	.48
17. Paper Form Board	.59
18. Hidden Patterns	.52
19. Cubes	.58
20. Paper Folding	.68
21. Raven	.67
22. Vocabulary	.81
23. Subtraction/Multiplication	.59
24. Word Beginnings/Endings	.67
25. Pedigrees	.85
26. Things Categories	.40
27. Different Uses	.61
28. Immediate Visual Memory	.55
29. Delayed Visual Memory	.64
30. Lines and Dots	.31
31. Identical Pictures	.32
Weschler Adult Intelligence Scale	
32. Information	.75
33. Comprehension	.64
34. Vocabulary	.81
35. Coding	.65
36. Arithmetic	.56
37. Similarities	.63
38. Digit Span	.52
39. Picture Completion	.63
40. Block Design	.62
41. Picture Arrangement	.39
42. Object Assembly	.49

among others, also observed). Substantial variability is also both measurable and not accounted for by the general factor. Any system intended to describe the structure of human mental ability must make allowance for this fact.

Our results also provide a "head-to-head" comparison of *g* loadings among many frequently administered intelligence tests. Intelligence tests are often compared on this basis. It is commonly believed that the Raven (1941) is most highly *g* loaded, but directly comparable data are rarely available. Table 2 shows the *g* loadings for each test in our sample. The highest loadings were for Pedigrees (.85) from the HB and the three Vocabulary tests, one from each battery (.82, .81, .81 for the Vocabulary tests from the CAB, the HB, and the WAIS, respectively). The loading for the Raven was .67. Its loading was also exceeded by the loadings for Numerical Ability, Word Fluency, Proverbs, and Spelling from the CAB; Paper Folding from the HB; and Information from the WAIS. Recall, however, that the Raven was administered in a somewhat modified fashion in our sample.

References

Baumeister, A. A., & Bacharach, V. R. (1996). A critical analysis of the Infant Health and Development Program. *Intelligence*, *23*, 79–104.

Blair, C. (1999). Science, policy, and early intervention. *Intelligence*, *27*, 93–110.

Bouchard, T. J., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota Study of Twins Reared Apart. *Science*, *250*, 223–228.

Browne, M. W., & Cudeck, R. (1992). Alternative methods of assessing model fit. *Sociological Methods and Research*, *21*, 230–258.

Browne, M. W., Cudeck, R., Tateneni, K., & Mels, G. (2001). *Comprehensive Exploratory Factor Analysis 1.03b*.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England: Cambridge University Press.

Cohen, J. (1957). The factor structure of the WAIS between early adulthood and old age. *Journal of Consulting Psychology*, *21*, 283–290.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6* (4), 330–351.

DeFries, J. C., Vandenberg, S. G., McClearn, G. E., Kuse, A. R., Wilson, J. R., Ashton, G. G., & Johnson, R. C. (1974). Near identity of cognitive structure in two ethnic groups. *Science*, *183*, 338–339.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Eysenck, H. J. (1994). A biological theory of intelligence. In D. K. Detterman (Ed.), *Current topics in human intelligence* (vol. 4, pp. 117–149). Norwood, NJ: Ablex.

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review*, *109*(1), 116–136.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*, 79–132.

Gustafsson, J. E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner, & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 186–242). New York: Simon and Schuster Macmillan.

Haier, R. J. (1993). Cerebral glucose metabolism and intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligence* (pp. 317–373). Norwood, NJ: Ablex.

Hakstian, A. R., & Cattell, R. B. (1975). *The Comprehensive Ability Battery*. Champaign: Institute for Personality and Ability Testing.

Hakstian, A. R., & Cattell, R. B. (1978). Higher stratum ability batteries on a basis of 20 primary mental abilities. *Journal of Educational Psychology*, *70*, 657–669.

Horn, J. L. (1989). Models of intelligence. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29–73). Urbana, IL: University of Illinois Press.

Horn, J. L. (1998). A basis for research on age differences in cognitive abilities. In J. J. McArdle, & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 57–91). Chicago: Riverside.

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cortex. *Journal of Comparative Neurology*, *387*, 167–178.

Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger.

Jensen, A. R., & Weng, L. (1994). What is a good *g*? *Intelligence*, *18*, 231–258.

Joreskog, K., & Sorbom, D. (2002). *LISREL 8.53: User's reference guide*. Chicago: Scientific Software International.

Kuse, A. R. (1977). Familial resemblances for cognitive abilities estimated from 2 test batteries in Hawaii. Unpublished dissertation from the University of Colorado at Boulder.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

McGue, M., Wette, R., & Rao, D. C. (1984). Evaluation of path analysis through computer simulation: Effect of incorrectly assuming independent distribution of familial correlations. *Genetic Epidemiology*, *1*, 255–269.

Raven, J. C. (1941). Standardization of progressive matrices, 1938. *British Journal of Medical Psychology*, *19*, 137–150.

Ree, M. J., & Earles, J. A. (1991). The stability of *g* across different methods of estimation. *Intelligence*, *15*, 271–278.

Scarr, S. (1996). How people make their own environments: Implications for parents and policy makers. *Psychology, Public Policy, and Law*, *2*, 204–228.

Segal, N. L. (2000). *Entwined lives: Twins and what they tell us about human behavior*. New York: Plume.

Sternberg, R. J. (2000). Implicit theories of intelligence as exemplar stories of success: Why intelligence test validity is in the eye of the beholder. *Psychology, Public Policy, and Law*, *6*, 159–167.

Thorndike, R. L. (1987). Stability of factor loadings. *Personality and Individual Differences*, *8* (4), 585–586.

Vernon, P. A. (1989). The generality of *g*. *Personality and Individual Differences*, *10*, 803–804.

Weschler, D. (1955). *Manual for the Weschler Adult Intelligence Scale*. New York: The Psychology Corporation.

Wilks, S. S. (1938). Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrika*, *3*, 23–40.

Williams, W. M. (2000). Perspectives on intelligence testing, affirmative action, and public policy. *Psychology, Public Policy, and Law*, *6*, 5–19.