Available online at www.sciencedirect.com
ScienceDirect
Intelligence 36 (2008) 81 – 95
INTELLIGENCE

Still just 1 *g*: Consistent results from five test batteries

Wendy Johnson [a,b,*], Jan te Nijenhuis [c], Thomas J. Bouchard Jr. [b]

[a] *Department of Psychology, University of Edinburgh, United Kingdom*
[b] *Department of Psychology, University of Minnesota, Twin Cities, United States*
[c] *Work and Organizational Psychology, University of Amsterdam, the Netherlands*

Received 26 November 2006; received in revised form 16 May 2007; accepted 7 June 2007
Available online 17 July 2007

Abstract

In a recent paper, Johnson, Bouchard, Krueger, McGue, and Gottesman (2004) addressed a long-standing debate in psychology by demonstrating that the *g* factors derived from three test batteries administered to a single group of individuals were completely correlated. This finding provided evidence for the existence of a unitary higher-level general intelligence construct whose measurement is not dependent on the specific abilities assessed. In the current study we constructively replicated this finding utilizing five test batteries. The replication is important because there were substantial differences in both the sample and the batteries administered from those in the original study. The current sample consisted of 500 Dutch seamen of very similar age and somewhat truncated range of ability. The batteries they completed included many tests of perceptual ability and dexterity, and few verbally oriented tests. With the exception of the *g* correlations involving the Cattell Culture Fair Test, which consists of just four matrix reasoning tasks of very similar methodology, all of the *g* correlations were at least .95. The lowest *g* correlation was .77. We discuss the implications of this finding.

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

Progress in science relies heavily on the process of replication. This is particularly true in the human behavioral sciences that depend on correlational studies, due to the ethical difficulties involved in the use of experimental designs. Replication is, however, seldom a straightforward exercise. Though successful replication substantiates the hypothesis in question, it acts only to increase the probability that the hypothesis is true; it cannot establish its absolute truth. And though failure to replicate can refute a hypothesis, more often there are questions of lack of statistical power or the existence of circumstantial differences that render interpretation of the failure to replicate less than clear. Despite these difficulties, or perhaps even because of them, replication itself can be a process of learning and discovery, generating ways of refining hypotheses and identifying new possibilities for underlying mechanisms.

In a seminal article on the subject, Lykken (1968) outlined three levels of replication. Literal replication involves exact duplication of sampling procedure, experimental conditions, measuring techniques, and methods of analysis. It may not be possible in practice to duplicate circumstances to this degree, but this may be of limited importance because the only real purpose of literal replication is to verify that the first study was conducted correctly. Operational replication simply involves duplication of sampling and experimental procedures, and its scope remains very limited. Operational replications tests whether the conditions and procedures described in the report of the original experiment produce the same result

---

[*] Corresponding author. c/o Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455, United States. Tel.: +1 952 473 1673; fax: +1 952 626 2079.

E-mail address: wendy.johnson@ed.ac.uk (W. Johnson).

0160-2896/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2007.06.001

when operationalized in a different sample by someone else. Constructive replication provides the strongest and most general test of the proposition because it deliberately avoids any duplication of procedures. Its purpose is to test the validity of the proposition independent of the original methods used to test it.

The positive intercorrelation among performances of individuals on tests of different mental abilities (positive manifold) is well established. This intercorrelation has been most frequently explained by positing the existence of an underlying general intelligence factor, commonly known as Spearman's "*g*" (Jensen, 1998). There is considerable empirical evidence for the existence and real-world importance of such a factor (Gottfredson, 1997; Gustafsson & Undheim, 1996; Jensen, 1998), but these demonstrations alone offer little explanatory power. There are two reasons for this. First, to date, the biological or even psychological basis of *g* has not been identified (Ackerman, Beier, & Boyle, 2005; Luciano et al., 2005), despite much effort and the identification of many biological and psychological factors to which *g* is clearly related (Deary, 2002; Detterman, 2002; Gray & Thompson, 2004). Absent such an understanding of the basis of *g*, it remains possible that *g* does not result from a single quantitative latent psychological and/or biological factor at all, but rather is the result of some developmental process such as reciprocal causation that generates a general factor, but that itself is multifaceted (van der Maas et al., 2006).

Second, many different mental ability tasks and batteries of tasks have been developed. When different combinations of these tasks are compiled, all of them produce a *g* factor. Some have argued, however, that there is no reason to expect that the *g* factors from different batteries should be related (Cattell, 1971; Horn, 1989). If they in fact were not, the concept of *g* as a unitary construct would be undermined. This is a psychometric issue that takes some precedence over the biological or psychological issues: if there is no consistency in our measurement of the construct we call *g*, there is little reason to suppose the existence of some biological or psychological basis for it. Yet, until recently, there was no direct evidence for the existence of this kind of consistency. This was probably due in part to the logistical difficulties involved in compiling the results of multiple mental ability test batteries in a single sample.

Recently, however, Johnson et al. (2004) provided such direct evidence. They made use of a sample of 436 individuals heterogeneous for age, sex, social class, and educational background that completed 42 mental ability tests from three test batteries. All three batteries included diverse ranges of tests: each included verbal tests, spatial ability tests, tests of knowledge as well as reasoning, along with tests of perceptual speed and fluency. In this sample, the correlations between the *g* factors for the three batteries were .99, .99, and 1.00. To the extent this is typical, it indicates that the *g* factors we measure are actually highly consistent from test battery to test battery. This was, however, only one sample in a single setting, and constructive replication in other circumstances is important. The purpose of the current study was constructively to replicate this finding (Johnson et al., 2004). We carried out our replication in a sample very different from the original. In addition, the new sample completed five test batteries very different in content from the original. Our sample consisted of 16-year-old professional seamen of the Royal Dutch Navy, tested in 1961 with 46 tests. The group of tests included few with verbal content and a large number of tests involving perceptual speed and mechanical aptitude.

Based on the sex differences reported in the literature (Halpern, 2000; Jensen, 1998), we would expect that a sample such as this, consisting entirely of young men, would show relatively weak performance on the verbal tests and some of the tests involving perceptual speed, but relatively strong performance on the perceptual tests involving spatial ability and the tests involving mechanical aptitude. The relative homogeneity might be expected to decrease the *g* correlations across batteries because of restriction of range. This should be even more the case to the extent that *g* correlations are not truly consistent from battery to battery. In addition, some of the batteries in this sample included tests of dexterity and motor coordination that were not present either consistently throughout these batteries, or in the batteries in the previous study (Johnson et al., 2004). Again, we might expect that, to the extent that these tests were included in some batteries but not others, they would reduce the correlations among *g* factors. Moreover, the five test batteries in the current study included an average of less than 9 tests per battery, while the three batteries in the previous study included an average of 14 tests per battery. Though this does not guarantee that the five batteries did not span the full range of cognitive abilities as well as the three batteries, it does suggest that this might be the case. We argue therefore that our current study provided a particularly strong test of the consistency of measurement of *g* from battery to battery.

1. Method

1.1. Sample

We made use of the data matrix of 46 mental ability tests published by de Wolff and Buiten (1963). This matrix is reprinted here as Appendix A. We did not have access to individual participant data of any kind. The sample on which the matrix was based consisted of 500 professional seamen of the Royal Dutch Navy. Thus the sample consisted entirely of male naval volunteers. Virtually all the seamen were 16 years of age at time of testing, in the summer of 1961. They were so-called sailors third class, the lowest rank in the Navy, and included no officers. During this period, 60% of the Dutch population stopped education after primary school. Of the Dutch born around 1945, 68% had already stopped their education around the age of 15, with 36% finishing only primary school at age 12 or a low-level extension of primary school for one or two additional years (Dutch Central Bureau of Statistics, 1980, p. 120–121). The education of the majority of the sample consisted of only primary school, technical school, or lower general secondary education. The Navy sought the more technically oriented students, as many of its jobs were technical in nature. Passage of final secondary school examinations and level of pre-selection test results (discussed in more detail below) was also important in selection of Navy seamen. Thus it is likely that limited numbers of seamen had IQs below 80, though we did not have access to this information. The sample probably consisted of individuals who represented the upper end of the IQ distribution in these lower educational groups. Thus average IQ in the sample was probably slightly below the full population average, with much smaller standard deviation. Otherwise, these young men were quite representative of the male population of the Netherlands, with a probable oversampling of persons living in the vicinity of the Navy harbour of Den Helder in the province of North-Holland (Charl de Wolff and Bert Buiten, personal communication, January 2006).

Rudolf van der Giessen (1952), the first psychologist at the Royal Dutch Navy, described the Navy's two-step personnel selection procedure. The first requirement was passing the medical examinations that were specific to each job category. All of these included an assessment of mental stability using a Rorschach test and an assessment of intelligence using the Dutch adaptation of the Wechsler Intelligence Scale for Children (WISC), administered by medical doctors rather than by psychologists. Those who obtained WISC scores below an unknown cutoff were not allowed to participate in the second part of the selection procedure. The second part of the selection procedure consisted of attainment of sufficient scores (again, cut-off not known) on the series of intelligence tests of the Royal Dutch Navy. In general, this series of tests was administered two weeks before employment commenced; the tests from the 5 batteries used in this study were administered during a period of two weeks, about two months after employment commenced in order to determine specific job assignment. All tests were administered by qualified instructors. In addition to the 5 batteries used in this study, the series included two tests of instrument and table reading from the Royal Dutch Air Force. We did not make use of these two tests for this study. These tests were relatively similar to each other and different from the tests in the other batteries.

1.2. Measures

Table 1 gives a description of all the tests administered after the first selection procedure, their time limits, the numbers of test items, and the test means and standard deviations. Visualization, mechanical, perceptual speed, spatial relations, and computation tasks were well represented, and there were several tests of manual dexterity, and motor coordination, in keeping with the activities required of the seamen. There was, however, only one test of memory, and there were relatively few tests involving verbal abilities. The first eight tests constituted the battery of the Royal Dutch Navy. It included tests of comprehension of mechanical principles, verbal vocabulary and analogies, arithmetical computation and problem-solving, speed of clerical comparison, visual pattern analysis, and selective and continuous attention. The next 12 tests were from the Aptitude Intelligence Tests of Joseph E. King (see Buros, 1959, p. 667) adapted by the Twente Institute of Business Psychology [Twents Instituut voor Bedrijfspsychologie: TIB] for the Netherlands and Belgium. This battery is better known as the Factored Aptitude Series (see Buros, 1953, p. 681). It included tests of pattern identification and matching, spatial ability, arithmetical computation, series reasoning, verbal fluency, memory, and concentration and attention.

The following four tests came from the Test of *g*: Culture Fair, Scale 2 form A + B of R.B. Cattell and A.K.S. Cattell (Cattell Culture Fair Test [CCFT], see Buros, 1959, p. 439). This battery was particularly narrow in conception, as all four tests took the form of matrix reasoning tasks. Next came the 13 tests from the General Aptitude Test Battery (GATB) of the U.S. Department of Labour, adapted for the Netherlands by the Psychological Service of Royal Dutch Blast-furnaces and Steelworks factories [Koninklijke Nederlandse Hoogovens en Staalfabrieken N.V.] , the largest Dutch steelworks (van der Giessen, 1960). It included tests of pattern comparison, arithmetic computation and problem solving, vocabulary, spatial ability, and manual mark-making and fingerboard dexterity. The final battery consisted of the 7 tests from the Groningse Intelligentie Test (Groningen Intelligence Test [GIT]; Snijders & Verhage, 1962), adapted for group administration. It included tests of synonyms, closure, object assembly, and verbal fluency. The Mutilated Words Test, was an experimental closure test that was not used in the final version of the GIT (see Buiten [1964] for details).

Table 1
Tests used by de Wolff and Buiten (1963)

Test	Assessment activity	Time allowed	No. of items	Mean	Std. dev.
Test Battery of Royal Dutch Navy					
1. Mechanical comprehension	Make decisions using mechanical principles.	30′	50	49.5	11.0
2. Form perception	Identify analogous ribs in different drawings of bodies	20′	60	51.2	10.0
3. Verbal	Solve analogies, find antonyms, complete sentences.	30′	65	49.6	10.0
4. Computation part 1	Carry out simple arithmetical computations.	12′	25	49.7	9.6
5. Computation part 2	Solve arithmetic word problems.	30′	20	51.1	10.4
6. Four Letter Words	Find four-letter words in lines of 46 letters.	8′	110	31.4	10.3
7. Administrative ability	Compare two names or numbers — same or different?	15′	275	49.2	10.6
8. Administrative speed (Bourdon)	Identify dot patterns consisting of four or five dots.	6′	900	49.7	11.6
TIB Battery (Factored Aptitude Series)					
9. Parts	Match given parts with a single composite figure.	5′	48	31.2	6.1
10. Dimension	Identify pictures rotated and drawn in reverse.	5′	48	25.8	8.6
11. Blocks	Count blocks in piles in which not all are visible.	5′	32	17	4.2
12. Judgment	Identify next in a series of numbers or letters.	5′	54	19.2	6.2
13. Numbers	Carry out simple arithmetical computations.	5′	54	35.3	7.0
14. Fluency 1 and 2	Generate words meeting specified conditions.	2×3′	N/A	4.8	1.9
15. Perception	Identify identical patterns of numbers and letters.	5′	54	19.8	6.3
16. Memory	Recall of presented names and faces.	2′+3′	36	18.5	9.2
17. Tools	Identify tools.	5′	48	28.4	5.9
18. Precision	Identify identical pictures.	5′	48	29.6	6.2
19. Maze	Trace a line through a maze.	1′	90	46.5	14.3
20. Checks	Place two check marks in each box.	1′	120	53.4	9.3
21. Dots	Place one dot in each triangle.	1′	180	99.7	21.1
Cattell Culture Fair Test					
22. Test 1: series	Identify next matrix in a series.	2×3′	2×12	16.6	3.4
23. Test 2: classification	Identify the matrix in each group that does not belong.	2×4′	2×14	15.9	3.5
24. Test 3: matrices	Identify analogous matrix.	2×3′	2×12	17.7	4.0
25. Test 4: conditions (topology)	Identify the topologically equivalent matrix.	2×2.5′	2×8	11.7	3.4
General Aptitude Test Battery					
26. Name comparison	Compare two names — same or different?	6′	150	49.7	11.6
27. Computation	Carry out simple arithmetical computations.	6′	50	21.1	4.9
28. Vocabulary	Given 4 words, select pairs of ant- and synonyms.	6′	60	19.3	6.3
29. Arithmetic reason	Solve arithmetic word problems.	7′	25	12.3	3.1
30. Tool matching	Identify identical pictures of tools.	5′	49	27.6	5.6
31. Form matching	Identify identical figures.	6′	60	26.1	5.9
32. Three-dimensional space	Reconstruct 3-D figure from 2-D representation.	6′	40	19.1	5.5
33. Finger dexterity board: assemble	Assemble rivet and washer, place in assigned hole.	90″	50	26	3.9
34. Mark making	Place an underlined quotation mark in each square.	1′	200	62.7	9.7
35. Pegboard manual dexterity: place	Move pegs from one set of holes to another.	3×15″	3×48	85.4	7.6
36. Pegboard manual dexterity: turn	Invert and replace pegs in holes.	3×30″	3×48	91.2	8.5
37: Finger dexterity board: disassemble	Remove rivet from assigned hole, detach washer.	60″	50	25.3	3.4
Groninger Intelligence Test					
38. Word list	Identify synonym for given word from 5 choices.	9′	20	10.8	3.0
39. Gestalt completion	Complete figure given vertices.	10′	20	12.4	3.0
40. Mutilated words	Identify words with letters distorted or parts removed.	3′	20	9.6	3.1
41. Figures	Assemble specified object from given pieces.	9′	20	11.1	3.0
42. Sorting	Sort lists of 8 symbols into 2 groups of 4 each.	3′	20	8.5	3.3
43. Naming animals	Name as many animals as possible in 1 min	1′	N/A	14.3	4.1
44. Naming professions	Name as many professions as possible in 1 min	1′	N/A	9.3	1.2
2 tests of the Royal Dutch Airforce					
45. Dial reading	Read and interpret pictured dials.	10′	57	11.5	8.9
46. Table reading part 1	Interpret provided tables.	7′	43	13.8	8.6

Note: N/A is not applicable. These tests require the test taker to generate words meeting specified conditions. The 2 tests of the Royal Dutch Airforce were not included in this analysis as they were similar to each other, different from the others, and too few to consider a separate battery.

All of these tests were used for large-scale assessment in the Netherlands for many years.

1.3. Statistical analyses

We used the same statistical procedures in this study as were used to develop the model presented in Johnson et al. (2004), though of course results related to particular battery structures and correlations among residual variances differed because the test batteries differed. In addition, the procedures were considerably more straightforward because missing data were not an issue in this study. We assumed that all participants provided data for all tests, and believe that this assumption was reasonable given the method of recruitment and circumstances of administration. Moreover, if in fact some participants were missing some data, we had no basis for determining how many or making any adjustment for the missing data as we did not have access to the individual-level data.

We conducted exploratory factor analyses for each of the five test batteries using the software program CEFA (Browne, Cudeck, Tateneni, & Mels, 2001), in order to develop second-order factor models independently for each battery. We used ordinary least squares factor extraction, with Geomin oblique rotation and Kaiser row weights. We carried out these analyses in a completely exploratory manner. Specifically, 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. We used the models we developed in this exploratory manner to carry out maximum likelihood confirmatory factor analyses for the five models combined using LISREL 8.54 (Joreskog & Sorbom, 2003). In doing so, we maintained the five separate structures we had developed, fixing the first loading to 1.00 for each factor in order to identify the model and extracting separate second-order *g* factors for each battery. The key results of our analysis were thus the correlations among the 5 *g* factors.

2. Results

2.1. Models for individual batteries

2.1.1. Test Battery of the Royal Dutch Navy

For this battery, we extracted three factors, which we labeled Mechanical Ability, Problem Solving, and Perceptual Speed. We chose 3 factors by examining several possible numbers of factors and choosing the solution that caused the RMSEA to be less than .08 (indicating a reasonable fit according to Browne & Cudeck, 1992) and which provided the most clearly interpretable solution. The RMSEA for the model was .071 (chi-square = 24.75, 7 *df*, *p* = .001). We would have preferred a model with an RMSEA less than .05 (indicating close fit according to Browne & Cudeck, 1992), but the model with four factors, for which the RMSEA was .040, contained one unique variable variance that was effectively 0, and two of the factors contained important loadings from only single tests, so we selected the three-factor solution that was more clearly interpretable and contained no numerical anomalies. The correlations among the three factors ranged from .11 for Mechanical Ability and Perception to .59 for Mechanical Ability and Solution. The first-order factor loadings on the second-order *g* factor ranged from .61 for Perception .92 for Mechanical Ability. We diagram this model in Fig. 1.

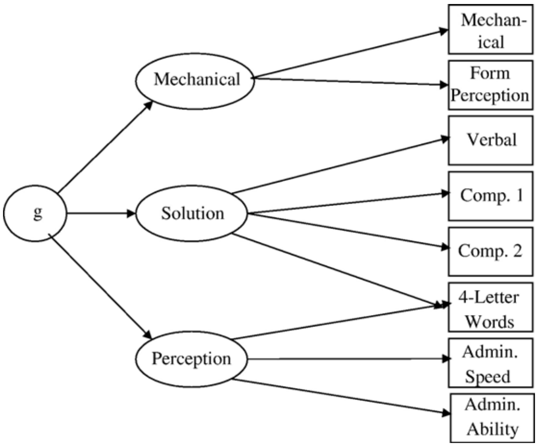

Fig. 1. Factor model for the Test Battery of the Royal Dutch Navy.

2.1.2. TIB

We extracted four factors for the TIB, labeled Pattern Identification, Reasoning, Memory, and Speed. For this model, the RMSEA was .035 (chi-square = 51.37, 32 *df*, *p* = .016). The correlations among the four factors ranged from .15 between Reasoning and Memory to .64 between Pattern Identification and Memory. The first-order factor loadings on the second-order *g* factor ranged from .46 for Clerical Speed to 1.00 for Pattern Identification. This model is diagrammed in Fig. 2.

2.1.3. CCFT

This battery consisted of only four rather similar matrix reasoning tests, so we extracted only a single factor for it. For this model, the RMSEA was .000 (chi-square = .51, 2 *df*, *p* = .776). The factor loadings ranged from .50 for Conditions to .73 for Classification. We diagram this model in Fig. 3.

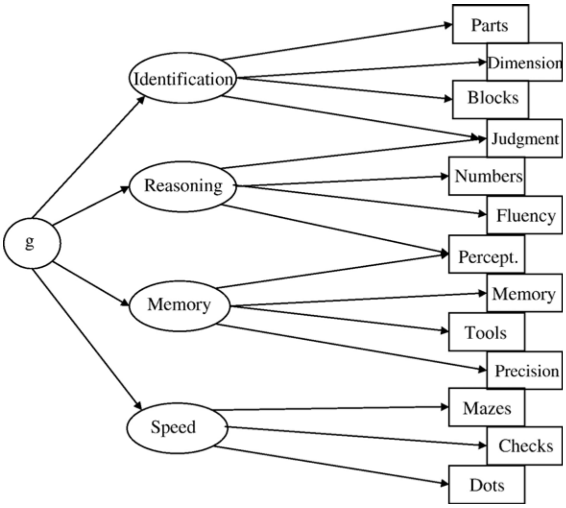

Fig. 2. Factor model for the Twente Institute of Business Psychology Battery.

2.1.4. GATB

We extracted four factors for this battery, and labeled them Verbal, Spatial, Building, and Dexterity. The RMSEA for the model was .046 (chi-square = 49.27, 24 *df*, *p* = .002). The correlations among the three factors ranged from −.05 between Verbal and Building to .64 between Verbal and Spatial. The first-order factor loadings on the second-order *g* factor ranged from .25 for Dexterity to .81 for Building. We diagram this model in Fig. 4.

2.1.5. GIT

For this battery we extracted three factors, labeled Closure, Organization, and Fluency. The RMSEA for this model was .040 (chi-square = 5.34, 3 *df*, *p* = .149). The correlations among the three factors ranged from .14 between Closure and Fluency to .34 between Closure and Organization. The first-order factor loadings on the second-order *g* factor ranged from .34 for Fluency to .90 for Organization. This model is diagrammed in Fig. 5.

2.2. Combined model

Fig. 6 diagrams the combined model we fit. It was based directly on the models for each of the batteries individually. The RMSEA indicated a reasonable fit (.073, chi-square = 3112.43, 847 *df*, *p* < .001), as did several other commonly used fit statistics (CFI = .95, SRMR = .09, NFI = .93). The top half of Table 2 shows the correlations among the *g* factors for the 5 batteries from the model in the figure. They ranged from .77 to 1.00. With the exception of the correlations involving the CCFT, all the correlations were at least .95.

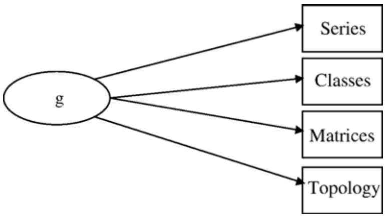

Fig. 3. Factor model for the Cattell Culture Fair Test.

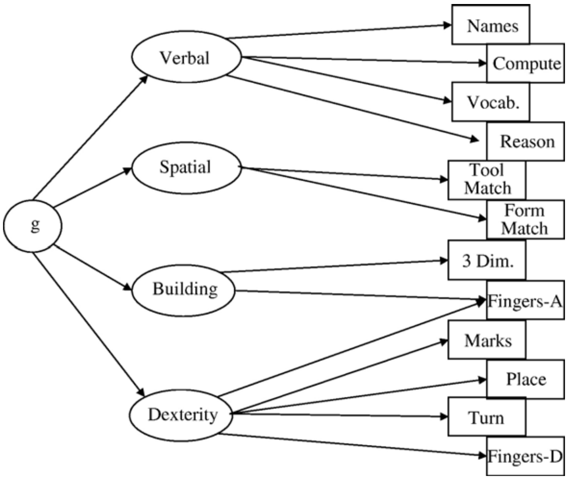

Fig. 4. Factor model for the General Aptitude Test Battery.

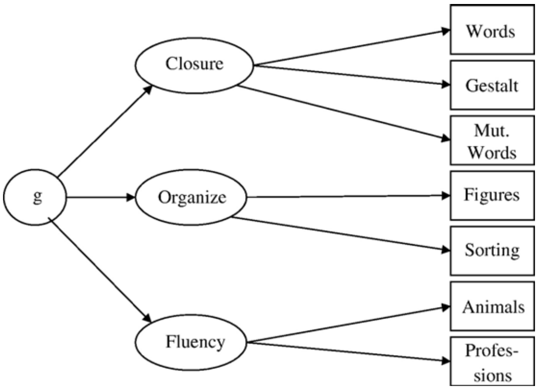

Fig. 5. Factor model for the Groninger Intelligentie Test.

This model was highly restricted because we constrained each test to load only on the factors extracted from the battery to which it belonged. There is no question that this was not the best model to describe the associations among the abilities assessed by this group of test batteries in this sample, as evidenced by the relatively poor, though still reasonable, model fit.[1] As we have already demonstrated,

---

[1] In fact, it was necessary to constrain the factor variance for the Identification factor of the TIB to 0 to prevent it from being negative. This is often an indication that the model is not optimal for the data, but we knew this to be the case, as we describe, so we allowed the constraint in this situation.

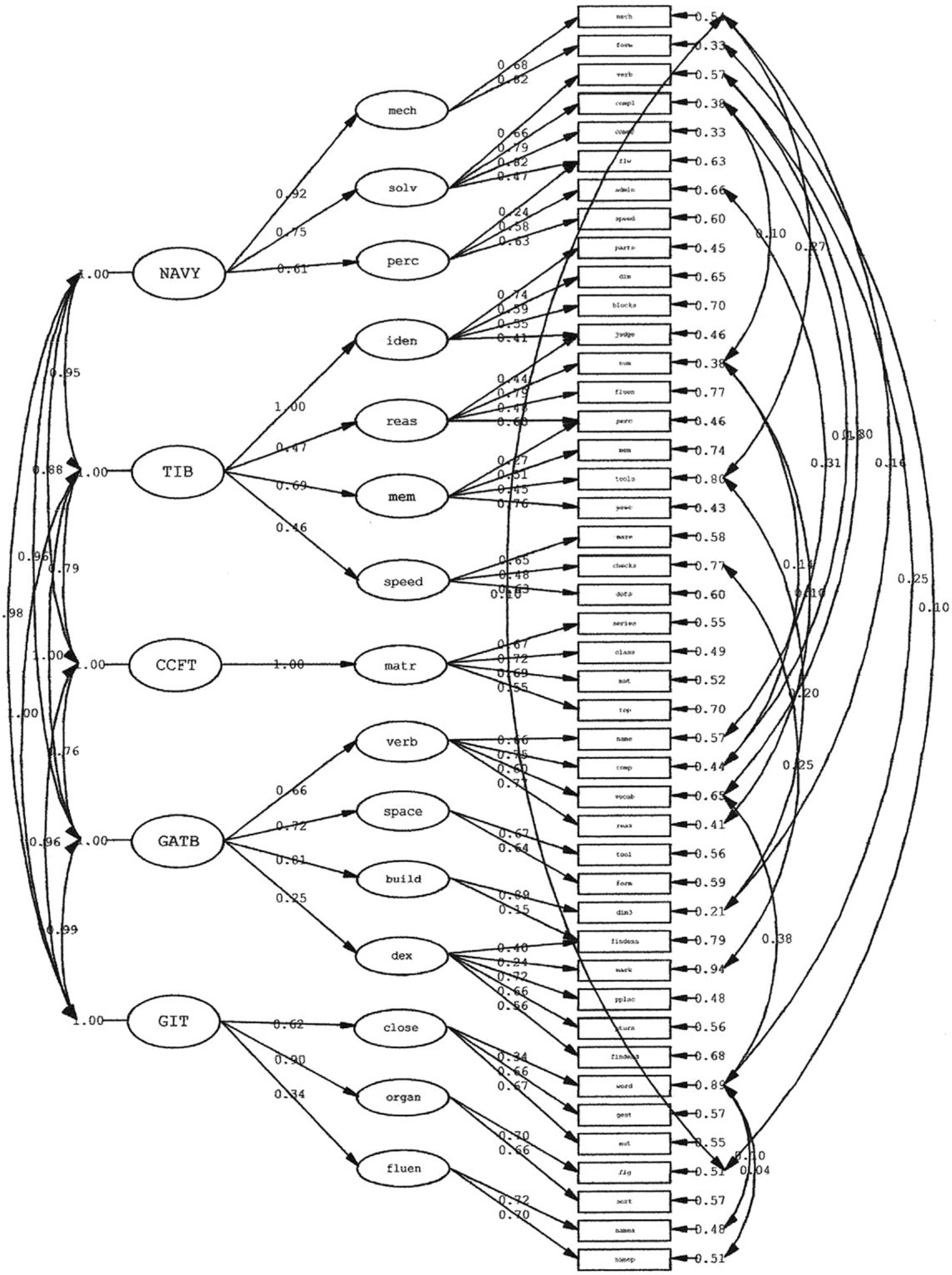

Fig. 6. Confirmatory factor analytic model estimating correlations among g factors in five test batteries. NAVY = Test Battery of the Royal Dutch Navy, TIB = Twente Institute of Business Psychology Battery, CCFT = Cattell Culture Fair Test, GATB = General Aptitude Test Battery, GIT = Groninger Intelligentie Test, mech = Mechanical Ability, solv = Problem Solving, perc = Perceptual Speed, reas = Reasoning, mem = Memory, speed = Clerical Speed, matr = Matrices, verb = Verbal, space = Spatial, build = Building, dex = Dexterity, close = Closure, organ = Organization, fluen = Fluency. The tests are listed in the order given in Table 1.

Table 2
Correlations among the *g* factors

Test Battery	1	2	3	4	5
From Full Model					
1. Royal Dutch Navy	1.00				
2. TIB (Factored Aptitude Series)	.95	1.00			
3. Cattell Culture Fair Test	.88	.79	1.00		
4. General Aptitude Test Battery	.96	1.00	.77	1.00	
5. Groninger Intelligence Test	.98	1.00	.96	.99	1.00
From Models of Pairs of Batteries					
1. Royal Dutch Navy	1.00				
2. TIB (Factored Aptitude Series)	1.20 [a]	1.00			
3. Cattell Culture Fair Test	.89	.81	1.00		
4. General Aptitude Test Battery	1.05 [b]	1.17 [d]	.73	1.00	
5. Groninger Intelligence Test	1.14 [c]	1.10 [e]	.93	1.19 [f]	1.00

[a] The *g*-factor correlation could be reduced to 1.00 by allowing correlations between the residuals for the Mechanical comprehension and Tools tests, the Mechanical and Identification factors, and the Reasoning and Perception factors; see Fig. 1.

[b] The *g* factor could be reduced to 1.00 by allowing correlations between the residuals for the Verbal and Vocabulary tests and between the Tool matching and Administration tests; see Fig. 1.

[c] The *g* factor could be reduced to 1.00 by allowing correlations between the Mechanical and Organization factors. See Fig. 1.

[d] The *g* factor could be reduced to 1.00 by allowing correlations between the residuals for the Three-dimensional space and Tools tests, the Three-dimensional space and Parts tests, and the Arithmetic Reasoning and Numbers tests, and between the Verbal and Reasoning factors; see Fig. 1.

[e] The *g* factor could be reduced to 1.00 by allowing correlations between the residuals for the Parts and Figures tests, and the Gestalt completion and Tools tests, and between the Identification and Organization factors; see Fig. 1.

[f] The *g* factor could be reduced to 1.00 by allowing correlations between the residuals for the Three-dimensional space and Figures tests and the Three-dimensional space and Parts tests, and the Computation and Gestalt completion tests; see Fig. 1.

Johnson and Bouchard's (2005) Verbal-Perceptual-Image Rotation (VPR) model provided a much better fit to these data (chi-square=1837.45, 929 *df*, RMSEA=.044), and even the Fluid-Crystallized model (Hakstian & Cattell, 1978) provided a much better fit than the current model (chi-square=2301.78, 932 *df*, RMSEA=.054 (Johnson, te Nijenhuis, & Bouchard, in press). It was, however, the model necessary to estimate the correlations among the *g* factors generated by each test battery. The arbitrary restrictions on the associations among the tests imposed by the battery boundaries complicated the model fitting process at two levels. As in all factor models, the individual tests had variance that was not captured by the factors on which they load. In this case in which the factors were defined by common content only within each test battery, there were individual tests that had variance in common with tests from other batteries. The same situation existed for the first-order factors. Failure to acknowledge this common variance was indicated by correlations between second-order *g* factors in excess of 1.00, and large modification indices for correlations between residual test variances and between first-order factor variances.

To the extent that these correlations were reasonable based on large modification indexes and common test and factor content, we allowed their presence in the model we show in Fig. 6 until the involved correlations among the second-order *g* factors fell to 1.00 or less. The correlations among the residual test variances that we allowed are shown explicitly in the figure. In addition, we allowed correlations between the Problem Solving and Reasoning (.40), Problem Solving and Verbal (.39), Problem Solving and Closure (.08), Problem Solving and Organization (.08), Perceptual speed and Fluency (.17), Reasoning and Verbal (.60), Memory and Fluency (.18), Clerical Speed and Spatial (.21), Verbal and Dexterity (.05), Spatial and Closure (.16), Building and Organization (.05), and Building and Fluency (.05) factors. We thus did not directly measure or test the correlations among the batteries as we could always recognize further such covariances and likely would eventually reduce the correlations among the *g* factors substantially. These covariances arose, however, because of *excess* correlation among the *g* factors, and we recognized them only in order to *reduce* this *excess* correlation. Thus, we provide evidence for the very high correlations we present, and no evidence at all that the actual correlations were lower. This is all that is possible within the constraints of our full model and given the goal of this study, which was to estimate the correlations among *g* factors in test batteries.

It is of course possible that the sheer complexity of our full model including five test batteries with 44 individual tests on a sample of 500 strained the capacity of the maximum likelihood parameter estimation procedure we used to produce meaningful estimates for the correlations in which we were interested. To address this possibility, we broke the full models into 10 pairs of two batteries and estimated the correlations between the *g* factors involved. Because of the similarities in tests across batteries, we still encountered some of the difficulties we describe above, making correlations between residuals and cross-battery factors necessary in order to reduce the correlations between the *g* factors to 1.00, though of course these difficulties were not as numerous in the models of pairs of batteries. Moreover, the models did not generally fit well, for the same reasons that the full model did not. The fit statistics for the pairs of models are given in Table 3. Overall, the resulting correlations between the pairs of *g* factors were almost identical to those from the full model. We summarize

Table 3
Fit statistics for the models of pairs of batteries

	χ^2 (*df*)	RMSEA	CFI	SRMR	NFI
Royal Dutch Navy–TIB	841.93 (175)	.087	.934	.075	.917
Royal Dutch Navy–CCFT	303.16 (50)	.101	.945	.067	.935
Royal Dutch Navy–GATB	1025.33 (160)	.104	.914	.091	.898
Royal Dutch Navy–GIT	448.81 (82)	.095	.928	.072	.913
TIB–CCFT	318.90 (112)	.061	.957	.060	.935
TIB–GATB	1013.31 (258)	.077	.934	.073	.911
TIB–GIT	563.29 (157)	.072	.934	.067	.909
CCFT–GATB	355.22 (98)	.073	.940	.064	.918
CCFT–GIT	89.56 (40)	.050	.977	.043	.968
GATB–GIT	708.05 (142)	.089	.887	.078	.863

Note: TIB is Twente Institute of Business Psychology Battery; CCFT is Cattell Culture Fair Test; GATB is General Aptitude Test Battery; GIT is Groninger Intelligentie Test.

them in the bottom part of Table 2 for the models allowing no residual or cross-battery correlations, which explains the presence in the table of factor correlations in excess of 1.00. We also note the residual and cross-battery correlations necessary to reduce any correlations in excess of 1.00 to 1.00. In no case did we add residual or cross-battery correlations in any situation in which a *g* correlation was not in excess of 1.00.

Another way to make provision for this variance would have been to allow test factor loadings across batteries. We did not do this, however, because it would have destroyed the battery structure whose *g* factor correlations we were trying to measure. In order to clarify the sources of variance included in some batteries but not others, we report the tests with the largest modification indexes on factors from other batteries. They included Administrative Ability from the Test Battery of the Royal Dutch Navy (TBRDN) on the Reasoning factor from the TIB, the Word List test from the GIT on the Problem Solving factor from the TBRDN, the same test on the Reasoning factor from the TIB, the Four-Letter Word test from the TBRDN on the Verbal factor from the GATB, the Administrative Ability test from the TBRDN on the same factor, the World List test from the GIT on that factor as well, and the Mark Making test from the GATB on the Fluency factor from the GIT. Allowing these loadings did not cause substantive reductions in the correlations among the second-order *g* factors.

As a test of the robustness of our conclusions, we fit a model with only a single first-order *g* factor for each battery, thus eliminating the individual battery first-order factor structures but retaining the separations between batteries (results not shown). The correlations among the 5 *g* factors were nearly identical, but the model fit was substantially poorer (RMSEA = .102, chi-square = 5454.71, *df* = 876, $p < .001$, CFI = .91, SRMR = .092, NFI = .89). At the same time, a model with the same first-order factor structure as the one we present (3 first-order factors for the Battery of the Royal Dutch Navy, 4 for the TIB, 1 for the CCFT, 4 for the GATB, and 3 for the GIT) but with only a single *g* factor on which all the first-order factors loaded did not fit substantively differently than the one we present (RMSEA = .074, chi-square = 3215.00, *df* = 855, $p < .001$, CFI = .95, SRMR = .09, NFI = .92; results not shown[2]).

3. Discussion

Our goal in this study was constructively to replicate Johnson et al.'s (2004) observation that the *g* factors from three independently developed test batteries were completely correlated. We did this in a very demographically different sample from the one that generated the original observation. This sample completed five independently developed test batteries, none of which was included among the batteries used in the earlier study. For these reasons alone, our current results provide a very strong test of the robustness of the original study. In addition, however, each of the batteries in the current study was more narrow in scope. Verbal and memory content were represented only sporadically, yet there were several manual dexterity tests of a type not included among the batteries in the original study.

The results of the current study were quite consistent with those of Johnson et al. (2004). Of the ten *g* factor correlations among the five batteries, seven were at least .95, indicating that the *g* factors were effectively interchangeable. There were three correlations that were not quite of this magnitude, however, though the lowest of these was .77. Even this correlation indicates a very high level of common measurement. For example, a re-test correlation of this magnitude is generally considered to indicate a test that can be considered to measure its construct reliably. Moreover, there is a viable explanation for the three correlations that fell below .95 (Table 4).

All of these correlations involved the Cattell Culture Fair Test, which consists of only four tests, all of which are matrix reasoning tests of highly similar format. Thus the *g*

---

[2] The degrees of freedom for this model differed from those of our full model by 8. From a purely conceptual perspective, the difference between the two models is the elimination of the 10 covariances among the five *g* factors for each battery. One of the two degrees of freedom was lost due to removal of the variance constraint on the variance of the CCFT factor. The other was lost as the constraint on the variance of the Identification factor of the TIB was not necessary in this model.

Table 4
g loadings for the tests in the five batteries

Test	
Test Battery of Royal Dutch Navy	
1. Mechanical comprehension	.52
2. Form perception	.63
3. Verbal	.67
4. Computation part 1	.69
5. Computation part 2	.70
6. Four Letter Words	.63
7. Administrative ability	.49
8. Administrative speed (Bourdon)	.41
TIB Battery (Factored Aptitude Series)	
9. Parts	.61
10. Dimension	.54
11. Blocks	.49
12. Judgment	.73
13. Numbers	.60
14. Fluency 1 and 2	.41
15. Perception	.60
16. Memory	.45
17. Tools	.22
18. Precision	.49
19. Maze	.27
20. Checks	.34
21. Dots	.34
Cattell Culture Fair Test	
22. Test 1: Series	.55
23. Test 2: Classification	.58
24. Test 3: Matrices	.57
25. Test 4: Conditions (topology)	.48
General Aptitude Test Battery	
26. Name comparison	.59
27. Computation	.64
28. Vocabulary	.65
29. Arithmetic reason	.66
30. Tool matching	.47
31. Form matching	.51
32. Three-dimensional space	.57
33. Finger dexterity board: assemble	.18
34. Mark making	.13
35. Pegboard manual dexterity: place	.09
36. Pegboard manual dexterity: turn	.21
37: Finger dexterity board: disassemble	.18
Groninger Intelligence Test	
38. Word list	.47
39. Gestalt completion	.42
40. Mutilated words	.46
41. Figures	.56
42. Sorting	.57
43. Naming animals	.34
44. Naming professions	.33

factor for this battery was very narrow in scope, both in absolute terms and in relation to the *g* factors from the other batteries. We believe that this acts to substantiate the robustness of the very high correlations observed among the other batteries and in the earlier study (Johnson et al., 2004), as it shows that these other very high correlations were not somehow artifacts of the method used to estimate them–that same method will produce lower correlations when the circumstances warrant. At the same time, even the *g* factor from a very narrow battery of tests is highly correlated with those from more broadly constructed batteries, attesting to the high level of uniformity in the *g* factor construct. Future research may make it possible to specify the number and content of tests necessary to develop batteries that will provide *g* factors that can be expected to correlate completely. Thus our results provide an example of the value of Lykken's (1968) concept of constructive replication. In essence, we reproduced the findings of the original Johnson et al. (2004) study, but the small differences between our current results and those of that study refine our understanding of the nature of measured *g* factors in ways that can be used both to improve future research that relies on measured *g* factors and to articulate more clearly the test properties necessary to measure *g* well.

Another way to think about these results involves the models we fit to test the robustness of our conclusions about our full model. As noted above, the model with only a single first-order *g* factor for each battery fit much worse than our full model, though the correlations among the *g* factors were very similar to those from the full model. This implies that the factor structure within each battery was important in explaining the pattern of relations within the data, but that our full model was describing the relations among the variance pools common to each of the batteries accurately. At the same time, the model retaining the factor structure for each battery but allowing only a single *g* factor fit very similarly to our full model. It is very reasonable to argue that this model did not fit as well as did our full model, but there is no question that it fit much better than the other competing model. In addition, the reasons that it did not fit as well as did our full model were exactly those given above: the *g* factor from the very narrowly defined Cattell Culture Fair Test did not correlate as highly with the others as they did with each other. This indicates not that there is more than one *g*, but that to measure that one *g* it is necessary to have a test or tests that have some breadth of content and format. In particular, it seems likely that both verbal and figural content are important, but this should be explored in future research.

Our current results also make possible a "head-to-head" comparison of *g* loadings among these frequently administered intelligence tests. The validity of intelligence tests is often compared on this basis. Table 3 shows the *g* loadings for each test in the current sample. Interestingly, though most of the tests had substantive *g* loadings, the highest was .73 for Judgment from the TIB, a series reasoning test. In the previous study (Johnson et al., 2004), the highest loadings were over .80 for vocabulary tests and Pedigrees, a test of familial relationships. In addition, tests in that study of the kinds well represented in the current batteries had *g* loadings in the .55–.65 range. These were very similar to the *g* loadings these kinds of tests had in the current study. The tests in the previous study that generated the *g* loadings in excess of .80 were represented poorly or not at all in the batteries in the current studies. Moreover, the manual dexterity tests in the current study that would appear to be less related to mental abilities generated quite low *g* loadings, even though there were quite a few of them. This indicates that, though far from absolute, tests of particular kinds of abilities may tap *g* to relatively consistent degrees. This acts to confirm Spearman's concept of the "indifference of the indicator" (Jensen, 1998, pp. 32–34).

A reviewer of this manuscript raised the possibility that the *g* factors we estimated correlated as highly as they did because all of the tests in the batteries were speeded, pointing out that de Wolff and Buiten (1963) characterized most of the tests in the batteries as speeded and that the correlation between test time limit in Table 1 and test *g* loading in Table 3 was .66. De Wolff and Buiten clearly were interested in comparing the performance of short tests and longer tests, and also were clearly interested in developing valid tests that could be administered in minimal amounts of time. They addressed this issue directly in their 1963 study, concluding on page 237 that the short tests (with administration times of 5–7 min) did not perform more poorly than the longer tests. Probably more importantly, they noted that the tests in the Battery of the Royal Dutch Navy all tended to be power tests with longer administration times, while those particularly from the GATB tended to be highly speeded. Thus, in our analysis, we estimated *g* factor correlations between batteries consisting primarily of power tests, batteries consisting primarily of speeded tests, and batteries consisting of combinations of speed and power tests. The only variable that appears to have made a difference in the level of correlation among the *g* factors is test homogeneity within batteries, as manifested by the CCFT. Jensen, too, has addressed the issue of test speededness, concluding (1980, page 134–135) that there is little difference in test performance between speeded and non-speeded tests.

The correlation between test time limit and test *g* loading that the reviewer noted appears to us to result because the tests in these batteries that involve overt reasoning exercises that would be expected to have high *g* loadings on a theoretical basis tend to have longer time limits, while the tests involving primarily manual dexterity and mark-making exercises that would be expected to have low *g* loadings on the same theoretical basis tend to be relatively brief. Moreover, test time limit is not a measure of test speededness. Test speededness reflects some ratio of typical time required for each item and number of items that is not easily captured by either test time limit or number of items. De Wolff and Buiten (1963) referred to this as well, noting that many of the tests in the TIB battery, though administered under strict time constraints, tended to be power tests because of the very limited numbers of items. For these reasons, we see no reason to suspect that our results were influenced in any particular way by test speededness.

In combination with the findings of the earlier study (Johnson et al., 2004), our results provide the most substantive evidence of which we are aware that most psychological assessments of mental ability of any breadth are consistently identifying a common underlying component of general intelligence. These results provide evidence both for the existence of a general intelligence factor and for the consistency and accuracy of its measurement. At the same time, these results demonstrate 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. The specific abilities not captured by the general factor are systematic and thus remain important. There are substantive correlations among these specific abilities from battery to battery, and from first-order factor to first-order factor, and different tests measure them with reliability comparable to that associated with the general factor. The biological basis of the brain structures that underlie and drive these correlations and the manner in which they develop remain to be articulated.

Acknowledgement

We thank Arne Evers, Jacqueline van Vliet-Mulder (both of the Dutch Test Committee), Jacques Dane (Archives of Dutch Psychology, University of Groningen), and especially Jan van de Linde for their help in finding information about the older tests. We are grateful to Charl de Wolff and Bert Buiten for enthusiastically giving additional details on their valuable dataset. Wendy Johnson holds a Research Council of the United Kingdom Fellowship.

Appendix A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Test Battery of Royal Dutch Navy																
1. Mechanical comprehension	1.000															
2. Verbal	.463	1.000														
3. Computation part 1	.372	.511	1.000													
4. Computation part 2	.415	.634	.682	1.000												
5. Administrative ability	.039	.178	.403	.266	1.000											
6. Form perception	.592	.429	.412	.471	.181	1.000										
7. Administrative speed (Bourdon)	.155	.141	.224	.218	.396	.214	1.000									
8. Four Letter Words	.194	.409	.447	.402	.450	.245	.291	1.000								
TIB Battery (Factored Aptitude Series)																
9. Tools	.455	.146	.110	.108	.014	.308	.092	.002	1.000							
10. Parts	.533	.333	.387	.363	.217	.552	.323	.298	.340	1.000						
11. Precision	.241	.239	.219	.181	.375	.340	.286	.277	.323	.410	1.000					
12. Judgment	.297	.498	.471	.541	.391	.410	.279	.503	.151	.369	.334	1.000				
13. Dimension	.365	.374	.272	.317	.198	.427	.257	.308	.196	.405	.382	.407	1.000			
14. Blocks	.310	.254	.251	.324	.169	.379	.267	.265	.105	.404	.196	.373	.339	1.000		
15. Numbers	.076	.346	.581	.462	.486	.171	.298	.507	−.102	.192	.200	.486	.204	.281	1.000	
16. Perception	.109	.406	.428	.389	.511	.216	.244	.512	.020	.202	.398	.459	.258	.225	.551	1.000
17. Fluency 1 and 2	.163	.378	.297	.307	.199	.153	.039	.369	−.044	.137	.108	.335	.219	.136	.342	.346
18. Memory	.222	.403	.268	.316	.225	.306	.219	.338	.226	.216	.347	.317	.218	.123	.219	.353
19. Maze	.154	.035	.121	.145	.230	.178	.192	.206	.150	.212	.189	.153	.148	.208	.129	.170
20. Checks	.090	.139	.202	.151	.296	.075	.211	.297	.089	.153	.259	.247	.184	.117	.286	.301
21. Dots	.150	.113	.212	.203	.263	.196	.169	.205	.049	.223	.261	.256	.121	.213	.277	.310
Cattell Culture Fair Test																
22. Test 1: Series	.340	.383	.366	.376	.248	.417	.189	.209	.150	.359	.234	.391	.305	.280	.296	.261
23. Test 2: Classification	.390	.440	.358	.416	.153	.446	.190	.253	.138	.410	.285	.426	.339	.310	.242	.252
24. Test 3: Matrices	.304	.441	.383	.431	.197	.460	.204	.253	.120	.353	.202	.442	.304	.287	.284	.241
25. Test 4: Conditions (topology)	.394	.387	.293	.368	.130	.435	.193	.208	.196	.325	.215	.394	.336	.237	.165	.171
G.A.T.B.																
26. Name comparison	.142	.322	.453	.346	.627	.217	.264	.541	.069	.223	.351	.485	.235	.218	.575	.597
27. Computation	.181	.370	.625	.449	.531	.301	.288	.468	.014	.300	.255	.520	.215	.266	.704	.500
28. Three-dimensional space	.602	.379	.334	.357	.099	.602	.212	.246	.416	.647	.347	.338	.419	.429	.089	.134
29. Vocabulary	.353	.690	.498	.532	.284	.324	.202	.518	.117	.315	.212	.496	.354	.218	.425	.508
30. Tool matching	.141	.207	.260	.207	.394	.273	.262	.379	.093	.305	.460	.303	.273	.256	.288	.424
31. Arithmetic reason	.221	.473	.625	.608	.406	.316	.233	.504	−.030	.275	.207	.492	.251	.256	.676	.448
32. Form matching	.190	.179	.327	.269	.388	.319	.279	.375	.122	.379	.342	.330	.233	.282	.349	.336
33. Mark making	−.078	−.022	.086	−.041	.264	−.077	.225	.188	.018	.062	.232	.046	−.042	.067	.228	.204
34. Pegboard manual dexterity: place	.032	−.086	−.034	−.056	.172	.009	.214	.111	.051	.059	.113	−.006	.022	.122	.074	.105
35. Pegboard manual dexterity: turn	.135	.034	.091	.056	.125	.128	.226	.110	.102	.193	.108	.091	.136	.170	.103	.136
36. Finger dexterity board: assemble	.156	.043	.073	.106	.050	.159	.149	.058	.168	.171	.134	.167	.154	.194	.037	.015
37. Finger dexterity board: disassemble	.109	.024	.078	.050	.196	.105	.245	.129	.174	.148	.170	.093	.084	.181	.074	.065
Groninger Intelligentie Test																
38. Word list	.207	.561	.374	.429	.227	.184	.108	.432	−.007	.164	.110	.376	.245	.102	.363	.388
39. Figures	.498	.377	.375	.397	.152	.575	.245	.277	.161	.544	.219	.334	.353	.384	.166	.132
40. Gestalt completion	.319	.311	.229	.244	.123	.333	.085	.230	.304	.335	.285	.269	.377	.219	.074	.163
41. Sorting	.328	.429	.356	.420	.201	.470	.241	.287	.086	.397	.300	.424	.294	.281	.225	.250
42. Mutilated words	.245	.304	.288	.290	.169	.251	.084	.345	.057	.256	.164	.330	.313	.223	.259	.320
43. Naming animals	.079	.240	.251	.162	.222	.111	.202	.313	.118	.183	.215	.269	.120	.103	.259	.267
44. Naming professions	.067	.201	.187	.151	.137	.100	.186	.255	.119	.229	.236	.274	.132	.106	.255	.270
2 tests of the Royal Dutch Airforce																
45. Dial reading	.360	.439	.414	.461	.268	.447	.267	.337	.142	.379	.203	.465	.330	.347	.381	.287
46. Table reading part 1	.172	.288	.315	.336	.352	.292	.253	.298	.070	.240	.223	.441	.259	.277	.385	.310

17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1.000														
.211	1.000													
.086	.119	1.000												
.151	.135	.355	1.000											
.126	.139	.424	.312	1.000										
.159	.212	.096	.153	.102	1.000									
.156	.254	.110	.156	.142	.479	1.000								
.157	.250	.115	.092	.098	.528	.490	1.000							
.153	.306	.080	.108	.055	.343	.349	.364	1.000						
.317	.250	.204	.347	.317	.277	.147	.212	.156	1.000					
.319	.233	.127	.242	.227	.360	.287	.329	.202	.611	1.000				
.111	.257	.142	.046	.179	.375	.407	.339	.353	.189	.214	1.000			
.386	.364	.028	.209	.126	.310	.379	.360	.302	.438	.423	.271	1.000		
.189	.205	.267	.205	.265	.158	.179	.190	.133	.400	.293	.237	.231	1.000	
.316	.259	.112	.228	.166	.340	.295	.364	.216	.515	.639	.200	.515	.287	1.000
.134	.231	.155	.272	.214	.215	.263	.229	.164	.434	.400	.324	.274	.428	.387
−.024	.122	.186	.372	.232	.102	.023	−.044	−.094	.226	.211	−.052	.048	.153	.096
−.022	.023	.234	.223	.150	.062	−.041	−.020	−.007	.141	.074	.009	−.048	.147	.044
.019	.081	.236	.309	.166	.145	.122	.070	.040	.128	.117	.148	.061	.140	.073
−.030	.017	.158	.111	.112	.066	.125	.082	.120	.077	.048	.208	.022	.106	.052
−.105	.074	.225	.120	.017	.115	.037	.027	.056	.198	.134	.121	.061	.150	.108
.365	.294	−.009	.119	.100	.203	.213	.204	.267	.324	.295	.108	.657	.193	.425
.138	.196	.157	.067	.173	.387	.431	.384	.328	.123	.241	.527	.301	.195	.302
.247	.198	.156	.149	.055	.214	.266	.197	.232	.164	.131	.381	.286	.286	.156
.171	.178	.060	.142	.163	.405	.547	.445	.314	.207	.291	.386	.366	.187	.310
.297	.174	.178	.190	.195	.215	.265	.237	.161	.229	.258	.281	.328	.334	.295
.226	.208	.027	.165	.150	.149	.171	.104	.103	.284	.275	.174	.275	.159	.219
.235	.204	.009	.147	.138	.172	.144	.106	.101	.242	.267	.177	.270	.109	.200
.236	.276	.226	.166	.144	.404	.351	.462	.335	.315	.438	.378	.427	.228	.482
.170	.205	.154	.224	.170	.294	.288	.306	.262	.320	.440	.229	.292	.167	.372

(continued on next page)

Appendix A (*continued*)

	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
Test Battery of Royal Dutch Navy															
1. Mechanical comprehension															
2. Verbal															
3. Computation part 1															
4. Computation part 2															
5. Administrative ability															
6. Form perception															
7. Administrative speed (Bourdon)															
8. Four Letter Words															
TIB Battery (Factored Aptitude Series)															
9. Tools															
10. Parts															
11. Precision															
12. Judgment															
13. Dimension															
14. Blocks															
15. Numbers															
16. Perception															
17. Fluency 1 and 2															
18. Memory															
19. Maze															
20. Checks															
21. Dots															
Cattell Culture Fair Test															
22. Test 1: Series															
23. Test 2: Classification															
24. Test 3: Matrices															
25. Test 4: Conditions (topology)															
G.A.T.B.															
26. Name comparison															
27. Computation															
28. Three-dimensional space															
29. Vocabulary															
30. Tool matching															
31. Arithmetic reason															
32. Form matching	1.000														
33. Mark making	.150	1.000													
34. Pegboard manual dexterity: place	.183	.252	1.000												
35. Pegboard manual dexterity: turn	.131	.206	.499	1.000											
36. Finger dexterity board: assemble	.097	−.006	.278	.301	1.000										
37. Finger dexterity board: disassemble	.184	.162	.405	.327	.319	1.000									
Groninger Intelligentie Test															
38. Word list	.140	−.019	−.077	-.010	−.029	−.018	1.000								
39. Figures	.317	−.068	.062	.129	.141	.095	.200	1.000							
40. Gestalt completion	.191	−.065	−.002	.134	.158	.046	.252	.249	1.000						
41. Sorting	.222	.059	.019	.098	.094	.062	.292	.466	.237	1.000					
42. Mutilated words	.202	−.053	−.012	.087	.111	.009	.296	.234	.457	.213	1.000				
43. Naming animals	.172	.340	.066	.118	−.021	.036	.215	.053	.145	.221	.031	1.000			
44. Naming professions	.187	.331	.114	.094	−.015	.002	.156	.111	.103	.202	.119	.508	1.000		
2 tests of the Royal Dutch Airforce															
45. Dial reading	.307	−.035	.049	.099	.079	.131	.332	.439	.264	.389	.314	.166	.130	1.000	
46. Table reading part 1	.314	.079	.117	.116	.041	.168	.221	.281	.160	.322	.160	.112	.149	.485	1.000

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: Same or different constructs. *Psychological Bulletin*, *131*, 30−60.

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

Browne, M. W., Cudeck, R., Tateneni, K., & Mels, G. (2001). Comprehensive exploratory factor analysis 1.03b.

Buiten, B. (1964). Gebroken Woorden Test: Open eind of meervoudige keuze? [Broken Words Test: Open end or multiple choice?]. *Nederlandst Tijdschrift voor de Psychologie*, *19*, 309−314.

Buros, O. K. (1953). The fourth mental measurements yearbook Highland Park, NJ: The Gryphon Press.

Buros, O. K. (1959). The fifth mental measurements yearbook Highlands Park, NJ: The Gryphon Press.

Cattell, R. B. (1971). Abilities: Their structure, growth, and action Boston: Houghton-Mifflin.

de Wolff, C. J., & Buiten, B. (1963). Een factoranalyse van vier testbatterijen [A factor analysis of four test batteries]. *Nederlands Tijdschrift voor de Psychologie*, *18*, 220−239.

Deary, I. J. (2002). g and cognitive elements of information processing: An agnostic view. In R. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence* (pp. 151–182). Mahwah, N.J: Erlbaum.

Detterman, D. K. (2002). General intelligence: Cognitive and biological explanations. In R. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence* (pp. 151–182). Mahwah, NJ: Erlbaum.

Dutch Central Bureau of Statistics (1980). Statistisch zakboek 1980 [Statistical pocket book 1980] The Netherlands, the Hague: Central Bureau of Statistics.

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

Gray, J. R., & Thompson, P. M. (2004). Neurobiology of intelligence: Science and ethics. *Nature Neuroscience*, *5*, 471−482.

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 & Schuster Macmillan.

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

Halpern, D. F. (2000). *Sex differences in cognitive abilities.* Mahwah, NJ: Lawrence Erlbaum, Publishers.

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.

Jensen, A. (1980). Bias in mental testing New York: Free Press.

Jensen, A. (1998). The g factor Westport, CN: Praeger.

Johnson, W., & Bouchard , T. J., Jr. (2005). The structure of human intelligence: it's verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*, 393−416.

Johnson, W., Bouchard, T. J., Jr., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, *32*, 95−107.

Johnson, W., te Nijenhuis, J., & Bouchard , T. J., Jr. (2007). Replication of the hierarchical Visual-Perceptual-Image Rotation (VPR) in de Wolffe & Buiten's (1963) Battery of 46 tests of mental ability. *Intelligence*, *35*, 69−81.

Joreskog, K., & Sorbom, D. (2003). LISREL 8.54: User's reference guide Chicago: Scientific Software International.

Luciano, M., Posthuma, D., Wright, M. J., de Geus, E. J. C., Smith, G. A., & Geffen, G. M. (2005). Perceptual speed does not cause intelligence, and intelligence does not cause perceptual speed. *Biological Psychology*, *70*, 1−8.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, *70*, 151−159.

Snijders, J. T., & Verhage, F. (1962). Voorlopige handleiding bij de Groninger Intelligentie Test: GIT [Provisional Manual of the Groningen Intelligence Test: GIT] Amsterdam: Swets.

van der Giessen, R. W. (1952). Enkele aspecten van het psychotechnisch onderzoek bij de Koninklijke Marine (Nederland), The Royal Navy (Engeland) en The United States Navy [Some aspects of the psychotechnical selection research at the Royal Dutch Navy (the Netherlands), The Royal Navy (Great Britain), and The United States Navy]. Unpublished master's thesis, Free University, Department of Art, Philosophy, Mathematics, and Physics, Amsterdam, the Netherlands.

van der Giessen, R. W. (1960). De GATB in de bedrijfspsychologische praktijk [The GATB in the business psychological practice. *Nederlands Tijdschrift voor de Psychologie*, *15*, 472−496.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizinga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842−861.