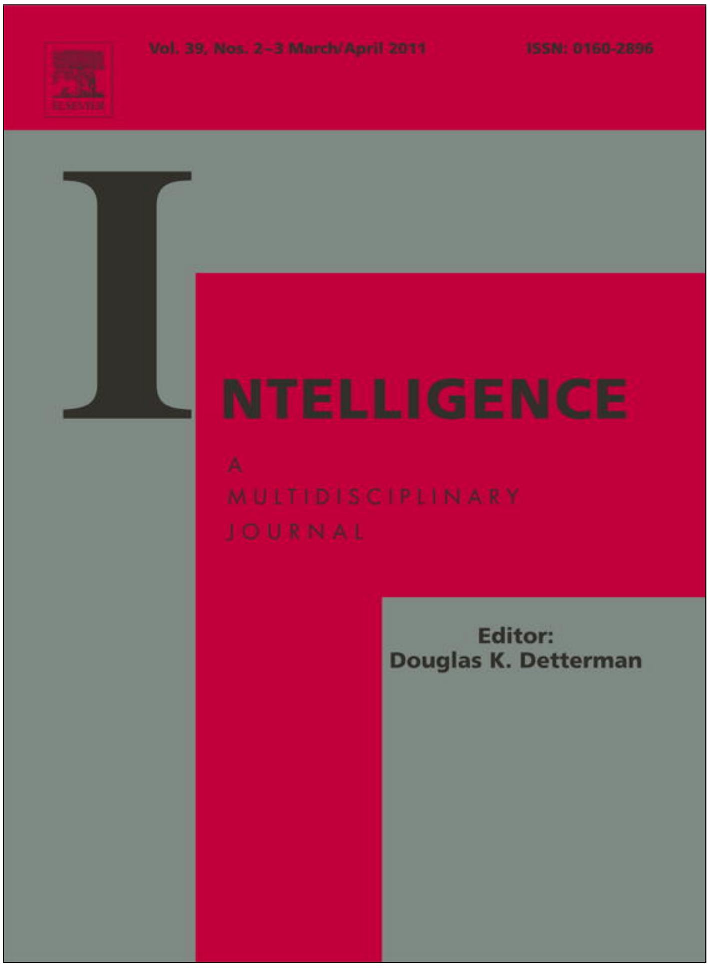
Vol. 39, Nos. 2–3 March/April 2011
ISSN: 0160-2896
INTELLIGENCE
A MULTIDISCIPLINARY JOURNAL
Editor:
Douglas K. Detterman

Intelligence 39 (2011) 82–88

Contents lists available at ScienceDirect

Intelligence

The MISTRA data: Forty-two mental ability tests in three batteries☆

Wendy Johnson[a,b,*], Thomas J. Bouchard Jr. [b]

[a] Centre for Cognitive Ageing and Cognitive Epidemiology and Department of Psychology, University of Edinburgh, UK
[b] Department of Psychology, University of Minnesota – Twin Cities, United States

ARTICLE INFO

Article history:
Received 24 February 2011
Received in revised form 25 February 2011
Accepted 25 February 2011
Available online 22 March 2011

Keywords:
Mental ability
Variance-covariance matrix
Correlations
Twins reared apart

ABSTRACT

The Minnesota Study of Twins Reared Apart (MISTRA) was initiated in 1979 and continued until 2000. It consisted of 139 pairs of twins who had been separated in early childhood and not re-united until adulthood, and members of their families. As part of a broader assessment, these participants completed 42 mental ability tests from three well-known test batteries. In this article, we present some background on the sample and tests and the correlation matrix of test scores, for the use of other researchers. As in all science, however, replication across samples of both tests and participants remains key to the development of ideas about mental ability.

Human mental abilities are multi-faceted, and many different kinds of tests have been developed to assess them. As Spearman (1904) noted, scores on these different kinds of tests tend to be correlated, but the correlations are always far from complete. The less-than-perfect correlations reflect more than just measurement error or assessment method: different kinds of mental ability tests tap different kinds of specific abilities that different people manifest to varying degrees. Thus comprehensive assessment of mental ability and investigation of its structure in the population requires measurement of some sample of the different kinds of abilities. As with all sampling, the larger and broader the sample, the more precise and accurate will be the estimates based on it.

Testing takes time and resources and the cooperation of participants. Researchers, even those specifically interested in the structure of mental ability, often must make do with single tests intended to tap different aspects of ability simultaneously, such as the Raven, or with just a few tests of more specific abilities that are selected with the hope of spanning the range of relevant abilities. Every once in a while, however, a truly broad-based assessment of mental ability tests is carried out in a single group of participants. Noteworthy examples of this include Thurstone and Thurstone's (1941) sample of 60 mental ability tests administered to 710 Chicago-area adolescents, de Wolff and Buiten's (1963) sample of 46 tests from five batteries administered to 500 Dutch seamen, Project Talent's sample of 22–63 (depending on degree to which scores are composited) aptitude and achievement tests to almost 400,000 high school students (Flanagan, Dailey, Shaycoft, Orr, & Goldberg, 1962), and the Minnesota Study of Twins Reared Apart's (MISTRA; Bouchard, Lykken, McGue, Segal, & Tellegen 1990; Segal 2000) sample of 42 mental ability tests.

These samples are important, for they provide near-unique opportunities to survey the structure of ability in great detail. For example, Carroll (1993) carried out a massive, systematic, and very detailed analysis of more than 460 datasets to put together a proposed model of cognitive ability, but he pieced his conclusions together by working across many rather small collections of tests administered to many different collections of participants. If recent work (Johnson & Bouchard 2005a; Johnson & Bouchard 2005b; Johnson, te Nijenhuis, & Bouchard 2007) with some of these larger collections of tests in single samples is any indication, the necessity of piecing together the

☆ Editors Note: I would like to thank Johnson and Bouchard for their generosity in making this data available to everyone. This gesture embodies the highest scientific altruism of sharing data that have, for some, taken a lifetime to collect. I encourage others to follow their example.

* Corresponding author at: Centre for Cognitive Ageing and Cognitive Epidemiology and Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ UK. Tel.: +44 131 651 1304; fax: +44 131 650 3461.
E-mail address: wendy.johnson@ed.ac.uk (W. Johnson).

0160-2896/$ – see front matter
doi:10.1016/j.intell.2011.02.010

full range of abilities from much smaller distinct samples may have blinded him to a conflation of verbal with crystallized and spatial with fluid abilities that continues to muddy rather than clarify the literature.

No single scientific study can be conclusive, due to measurement error, random sampling fluctuations (perhaps especially winners curse), violations of assumptions underlying methods of data analysis, and unforeseen effects of specific conditions of assessment (Schmidt 2010). This is especially true in an area such as investigation of the structure of mental ability. There are several reasons for the particular difficulty in the investigation of the structure of mental ability, all of which relate to factor analysis, the primary method of analysis that has been used, whether in its confirmatory or exploratory forms. First, for all their quantitative natures, both forms require many subjective judgments for implementation. Second, results from factor analysis are highly dependent on the contents of the variables. If some constructs are measured in great detail with many variables and others are measured only in summary fashion, the constructs measured in detail will generate strong factors that make rather subtle distinctions, while the constructs measured only in summary will generate weaker and more general factors. Third, results are also dependent on the participant sample. If the range of abilities in the sample is too narrow in some or all areas, factor analysis will not generate factors that may be important in the population at large. Finally and most importantly, factor analysis cannot help to resolve questions of direction of causation. Factors may represent either latent causal or emergent resultant variables equally well (Bartholomew 2004; Van der Maas et al., 2006).

Because no single scientific study can be conclusive, replication is crucial. Unfortunately, psychology is riddled with reports of effects that do not replicate (e.g., Pietschnig, Voracek, & Forman 2010). The problem is not limited to psychology, but pervades many other sciences as well (Fanelli 2010; Ionnidis 2005; Vinels 2009). Though there are many social reasons for these failures, one major implication of it is that replication carries most weight when it is carried out constructively (Lykken 1968). That is, we should be most ready to reassess our prior beliefs on a subject when several different studies using different samples and different procedures and relying upon different sets of assumptions all suggest in the same general way that our prior beliefs may have been unfounded. Thus, the work with the large datasets showing what may be inaccuracies in Carroll's (1993) summary model stands not on the details of the analysis in any one dataset, but on the consistency of results across the datasets in which the question has been investigated. Moreover, even failures to replicate are important, as they can generate understanding of the conditions limiting the original observations and thus suggest the necessary modifications to theoretical interpretations. For example, using the MISTRA mental ability tests, Johnson, Bouchard, Krueger, McGue, and Gottesman (2004) showed that correlations among *g* factors from the three test batteries completed by the participants were .99, .99, and 1.00. Replication in the five test batteries completed by the Dutch seamen (Johnson, te Nijenhuis, & Bouchard 2008) suggested the limitations of this observation: the one battery that consisted only of four tests that had one single, very similar format generated a *g* factor that was correlated as low as .77 and .79 with the *g* factors from two of the other batteries, and .88 with another. The other *g*-correlations among the five batteries, however, ranged from .95 to 1.00.

The purpose of this paper is to make the MISTRA dataset available to other researchers so that it can be used both to generate new ways of thinking about mental ability and to attempt to replicate findings generated in other datasets.

1. Method

1.1. Research participants

The MISTRA participants were gathered through a variety of sources over a period of many years. They came from a broad range of socioeconomic backgrounds and occupations, and most lived in North America, Great Britain, and Australia. They ranged in age from 18 to 79 years, and in education from less than high school to post-graduate experience. The pairs of reared-apart twins formed the heart of the sample. In most cases, they had been separated early in life, reared in adoptive families, and not re-united until adulthood. In addition to the twins, the sample included some of their spouses, adoptive and biological family members, partners, and friends. In total, 127 twin pairs, 2 sets of triplets, 116 spouses of twins, and 57 other family members contributed mental ability scores (186 males, 247 females).[1] They did so while participating in a week-long assessment of medical, physical, and psychological traits in addition to mental abilities such as personality, interests, and attitudes. Most of the mental ability tests were administered in sessions lasting 60 to 90 min spread throughout the assessment week. Some individuals were tested twice, 8–10 years after the first assessment. When this was the case, we used scores from the first assessment.

1.2. Measures

The participants completed 3 well-known mental ability test batteries. They are summarized in Table 1, and further details on each follow.

1.2.1. Comprehensive Ability Battery (CAB)

The CAB (Hakstian & Cattell 1975) consists of 20 brief (5–6 min each) tests developed to measure a broad range of generally accepted specific abilities. To make maximal use of available time, avoid task duplication, and keep the focus on mental ability, 6 of the 20 tests in the CAB were omitted (Auditory Ability, Originality, Representational Drawing, Aiming, Spontaneous Flexibility, and Ideational Fluency). In addition, because we deemed it not directly related to mental ability, we did not include the test of Esthetic Judgment in the data presented here. As the Verbal Ability Test consists of two completely different tasks, we considered scores on these two parts separately, producing a total of 14 scores. Hakstian and Cattell (1978) reported split-half and test-retest reliabilities for

[1] This differs by one twin pair from participant information reported previously because one twin pair, age 11, and one spouse duplication were included in those analyses. These data have been excluded here. They did not affect prior results and we wished to limit the data for general analysis to adults.

Table 1
Tests included in the 3 batteries.

Test	Assessment activity
Comprehensive Ability Battery	
1. Numerical ability	Computations including fractions, decimal divisions, square roots, etc.
2. Spatial ability	Interpretation of 2-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 2-dimensional prepresentation of 3-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 2-digit subtractions and 2-digit by 1-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.
Wechsler 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.
39. Picture completion	Identification of parts missing in pictures of common objects.
40. Block design	Reproduction of 2-dimensional designs using 3-dimensional blocks.
41. Picture arrangement	Chronological sequencing of pictures.
42. Object assembly	Reassembly of cut-up figures.

the tests ranging from .64 for Perceptual Speed and Accuracy to .96 for Memory Span.

1.2.2. The Hawaii Battery, including Raven's progressive matrices (HB)

The HB (DeFries et al., 1974; Kuse 1977) consists of 15 tests of specific abilities that each require 3–10 min to administer. The battery was developed to assess familial resemblance in mental ability in the Hawaii Family Study of Cognition. To avoid test duplication and maintain focus on cognitive abilities, two tests (Number Comparison and Social Perception) in the battery were omitted. To provide clearer articulation of abilities thought likely to be important, four tests from the Educational Testing Service (Ekstrom, French, Harman, & Dermen 1976) were added (Cubes and Paper Folding for spatial ability, Identical Pictures for perceptual speed and accuracy, and Different Uses for verbal fluency), for a total of 17 tests. This battery included a shortened version of the Raven (Raven 1941), administered via slides without time restriction (Lykken 1982). Internal consistency and test-retest reliabilities for the tests ranged from .58 for Immediate Visual Memory to .96 for Vocabulary (DeFries et al., 1974).

1.2.3. The Wechsler Adult Intelligence Scale (WAIS)

The WAIS (Wechsler 1955) consists of 11 tests involving both abstract reasoning and the ability to handle practical situations requiring verbal articulation of reasoning based on factual knowledge. Internal consistency reliabilities range from .79 for Comprehension to .94 for Vocabulary (Wechsler 1955). In the MISTRA sample, average WAIS full-scale IQ was 109.7 (range 79–140), with standard deviation of 11.8, normed at the 1955 level. Jensen (1998, page 319) summarized the average rate of secular change in IQ since 1955; adjusted for this change, the average WAIS full-scale IQ for the sample was 101.3 (range

Table 2
Correlation matrix of 42 mental ability tests completed by participants in the Minnesota study of twins reared apart.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1. Verbal 1	1.000																							
2. Verbal 2	.729	1.000																						
3. Spelling	.711	.609	1.000																					
4. Vocab	.809	.711	.733	1.000																				
5. Info	.723	.624	.641	.701	1.000																			
6. Comp	.641	.596	.491	.616	.640	1.000																		
7. Voc	.823	.728	.693	.801	.767	.714	1.000																	
8. Sim	.582	.566	.484	.564	.634	.570	.663	1.000																
9. Pedigr	.664	.641	.626	.681	.636	.488	.632	.527	1.000															
10. Raven	.601	.464	.520	.488	.575	.486	.548	.472	.571	1.000														
11. Thing	.481	.481	.451	.532	.495	.450	.544	.403	.474	.334	1.000													
12. Difuse	.531	.524	.418	.529	.507	.545	.589	.508	.513	.409	.612	1.000												
13. PC	.432	.428	.362	.430	.481	.404	.466	.443	.455	.484	.347	.358	1.000											
14. PA	.357	.397	.265	.333	.405	.416	.387	.321	.403	.398	.316	.319	.445	1.000										
15. OA	.252	.246	.173	.259	.324	.300	.265	.276	.319	.393	.258	.286	.435	.339	1.000									
16. Memsp	.415	.378	.465	.430	.439	.305	.416	.362	.485	.398	.363	.366	.228	.185	.112	1.000								
17. Spdclos	.349	.365	.449	.409	.381	.253	.350	.372	.478	.306	.321	.282	.335	.197	.252	.329	1.000							
18. Wordfl	.590	.515	.652	.648	.545	.452	.581	.488	.616	.461	.553	.527	.368	.260	.202	.484	.493	1.000						
19. Wordbe	.598	.542	.622	.626	.539	.504	.618	.474	.569	.478	.499	.477	.377	.325	.220	.429	.410	.717	1.000					
20. DS	.388	.382	.461	.392	.338	.276	.404	.324	.451	.388	.320	.283	.258	.171	.117	.597	.356	.466	.484	1.000				
21. Asscmem	.342	.324	.367	.370	.389	.246	.335	.275	.436	.314	.206	.334	.237	.202	.168	.336	.336	.370	.233	.288	1.000			
22. Mnmem	.533	.523	.468	.503	.487	.450	.555	.461	.506	.444	.360	.461	.411	.345	.223	.295	.334	.446	.444	.345	.439	1.000		
23. Ivmem	.261	.225	.132	.247	.229	.186	.217	.179	.266	.258	.142	.166	.217	.154	.171	.092	.067	.195	.182	.110	.260	.309	1.000	
24. Dvmem	.254	.224	.177	.250	.232	.217	.256	.156	.269	.237	.196	.187	.195	.138	.176	.137	.132	.239	.259	.203	.209	.324	.531	1.000
25. Percspd	.405	.387	.487	.408	.346	.239	.341	.310	.562	.369	.277	.355	.282	.180	.167	.373	.448	.493	.365	.323	.334	.337	.158	.222
26. Ident	.388	.401	.391	.475	.422	.312	.384	.339	.560	.309	.408	.423	.415	.348	.418	.283	.449	.455	.324	.190	.292	.321	.178	.145
27. Crdrot	.129	.152	.184	.219	.222	.186	.171	.168	.411	.358	.165	.176	.284	.252	.353	.237	.257	.269	.229	.126	.233	.212	.176	.175
28. Cubes	.371	.366	.346	.373	.373	.363	.395	.314	.514	.473	.290	.340	.340	.301	.298	.349	.344	.426	.368	.296	.342	.358	.201	.205
29. Number	.581	.542	.661	.595	.585	.452	.570	.462	.658	.527	.445	.475	.333	.261	.212	.546	.471	.612	.516	.495	.403	.427	.153	.239
30. SM	.436	.432	.580	.551	.457	.317	.436	.329	.513	.329	.372	.353	.213	.146	.133	.442	.415	.582	.460	.376	.314	.350	.131	.187
31. Linedot	.255	.273	.208	.265	.265	.224	.247	.234	.354	.205	.216	.268	.231	.167	.239	.204	.226	.212	.143	.198	.260	.245	.167	.160
32. Dsym	.456	.401	.511	.501	.407	.314	.422	.357	.525	.418	.305	.394	.344	.211	.195	.430	.326	.532	.402	.404	.372	.372	.236	.267
33. Arith	.578	.561	.612	.572	.619	.483	.598	.517	.617	.522	.359	.402	.357	.352	.205	.457	.345	.581	.535	.454	.410	.475	.149	.225
34. Space	.301	.275	.269	.302	.294	.248	.288	.257	.494	.438	.233	.316	.300	.239	.301	.308	.271	.323	.290	.248	.284	.281	.218	.173
35. Flxclos	.507	.416	.448	.466	.447	.432	.450	.374	.471	.427	.313	.431	.367	.310	.296	.357	.362	.442	.402	.298	.316	.372	.102	.166
36. Mchancl	.397	.416	.270	.342	.386	.438	.416	.354	.374	.418	.288	.367	.391	.388	.415	.177	.217	.269	.315	.248	.137	.373	.146	.171
37. Induct	.528	.491	.513	.494	.485	.358	.505	.416	.607	.533	.389	.395	.385	.320	.260	.400	.455	.503	.470	.378	.343	.423	.203	.214
38. Mentrot	.257	.283	.241	.283	.290	.267	.260	.229	.379	.389	.161	.229	.348	.341	.369	.239	.285	.249	.248	.202	.240	.268	.167	.141
39. Pprform	.405	.395	.315	.405	.447	.375	.394	.367	.517	.478	.385	.396	.458	.363	.500	.313	.327	.393	.346	.282	.311	.336	.197	.267
40. Hidpat	.465	.399	.454	.472	.472	.334	.428	.362	.573	.528	.368	.409	.427	.384	.444	.376	.359	.464	.409	.303	.348	.350	.241	.242
41. Pprfld	.410	.388	.325	.397	.449	.349	.421	.349	.531	.539	.290	.366	.396	.354	.439	.283	.359	.384	.374	.276	.314	.392	.247	.265
42. BD	.428	.405	.360	.398	.426	.381	.430	.361	.551	.587	.342	.378	.436	.376	.501	.335	.308	.405	.401	.303	.289	.376	.207	.233

Table 2 (*continued*)

	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
1. Verbal 1																		
2. Verbal 2																		
3. Spelling																		
4. Vocab																		
5. Info																		
6. Comp																		
7. Voc																		
8. Sim																		
9. Pedigr																		
10. Raven																		
11. Thing																		
12. Difuse																		
13. PC																		
14. PA																		
15. OA																		
16. Memsp																		
17. Spdclos																		
18. Wordfl																		
19. Wordbe																		
20. DS																		
21. Asscmem																		
22. Mnmem																		
23. Ivmem																		
24. Dvmem																		
25. Percspd	1.000																	
26. Ident	.547	1.000																
27. Crdrot	.335	.375	1.000															
28. Cubes	.374	.361	.509	1.000														
29. Number	.589	.416	.258	.436	1.000													
30. SM	.526	.406	.203	.236	.712	1.000												
31. Linedot	.306	.325	.217	.272	.312	.278	1.000											
32. Dsym	.579	.526	.219	.349	.562	.547	.265	1.000										
33. Arith	.426	.322	.199	.404	.684	.539	.241	.465	1.000									
34. Space	.475	.460	.594	.513	.375	.295	.255	.384	.299	1.000								
35. Flxclos	.430	.435	.286	.470	.521	.354	.261	.445	.443	.322	1.000							
36. Mchancl	.165	.322	.313	.392	.301	.138	.256	.195	.353	.294	.387	1.000						
37. Induct	.468	.421	.313	.462	.561	.390	.239	.391	.520	.402	.463	.312	1.000					
38. Mentrot	.331	.360	.469	.415	.255	.137	.220	.194	.246	.450	.308	.393	.342	1.000				
39. Pprform	.413	.512	.480	.501	.439	.272	.336	.374	.399	.442	.448	.460	.435	.450	1.000			
40. Hidpat	.482	.550	.493	.469	.470	.403	.311	.459	.421	.473	.485	.386	.467	.488	.571	1.000		
41. Pprfld	.344	.384	.426	.536	.429	.232	.215	.302	.451	.424	.395	.550	.470	.454	.590	.455	1.000	
42. BD	.495	.479	.470	.545	.453	.272	.309	.409	.464	.466	.500	.458	.498	.479	.643	.602	.312	1.000

61.1–139.9), with standard deviation of 14.8. Because IQ was positively correlated with age in this sample, the adjustment increased the standard deviation.

1.2.4. Statistical analysis

We adjusted all test scores for the effects of age, age^2, sex, age x sex, and age$^2 \times$sex by regressing them on these terms and saving the standardized residuals. Because there were small amounts of ignorable missing data, we used maximum likelihood estimation to produce the variance–covariance and correlation matrices of test scores. We recognized the biological relationships between members of twin pairs using the sandwich estimator as implemented in Mplus 6.0 (Asparouhouv 2005) (Muthen & Muthen 1998–2010). This reflects a technological advance since preparation of the analyses in Johnson and Bouchard (2005a); thus, the variance-covariance matrix presented here is very similar but not identical to the one we used then. Failure to reflect these relationships has little effect on parameter estimates in models using these data, though it tends to inflate indices of model fit (McGue, Wette, & Rao 1984). In our prior studies using these data without adjustment for these relationships, we verified the results presented by repeating the analyses in data files excluding one member of each twin pair.

2. Results

Table 2 shows the correlation matrix. Electronic versions of both the correlation and covariance matrices are available as supplementary material on the Elsevier *Intelligence* website or at www.ISIRonline.org. The matrices there are formatted in plain text using scientific notation and include the principal diagonals and the upper triangular portions. They may be read in a word processor (e.g., Word Pad) or imported

Notes to Table 2:

1. CAB vocabulary
2. CAB proverbs
3. CAB spelling
4. HB vocabulary
5. WAIS information
6. WAIS comprehension
7. WAIS vocabulary
8. WAIS similarities
9. HB pedigrees
10. HB raven
11. HB things categories
12. HB different uses
13. WAIS picture completion
14. WAIS picture arrangement
15. WAIS object assembly
16. CAB memory span
17. CAB speed of closure
18. CAB word fluency
19. HB word beginnings/endings
20. WAIS digit span
21. CAB associative memory
22. CAB meaningful memory
23. HB immediate visual memory
24. HB delayed visual memory
25. CAB perceptual speed
26. HB identical pictures
27. HB card rotation
28. HB cubes
29. CAB numerical ability
30. HB subtraction/multiplication
31. HB lines and dots
32. WAIS digit symbol
33. WAIS arithmetic
34. CAB spatial ability
35. Flexibility of closure
36. CAB mechanical ability
37. CAB inductive reasoning
38. HB mental rotation
39. HB paper formboard
40. HB hidden patterns
41. HB paper folding
42. WAIS block design

into a spreadsheet (e.g., Excel). If you have difficulty opening them, save the file, right click on the file name, select "Properties,"and change the program used to open them. Scientific notation can be removed in Excel by formatting as numbers with 8 digits beyond the decimal point.

3. Discussion

As in any analysis, best use of these data will be made if researchers have some background on the reasons for the selection of the tests. The CAB was chosen because it was developed specifically to operationalize the Cattell/Horn fluid and crystallized model of intelligence and the theory predicted that these two factors had different heritabilities. Additionally the theory predicted that there was no *g* factor. These three tenets were testable given the twin data being gathered in MISTRA. The Hawaii Battery was chosen because it represented the Spearman/Vernon tradition of sampling a very wide array of measures of mental ability and, within that tradition, was not driven by a particular factor model. The WAIS was chosen for four reasons. First, it represented what might be called the "American Clinical Tradition" and thus what works best in the individual testing situation rather than a theoretical or factorial model. Second, it is individually rather than group administered and we wished to have different test administrators test each of the twins in the study in order to avoid claims of tester bias. Third, it was then and still is widely considered to be the gold standard of IQ testing, having been in use for many years and having undergone refinement based on use by thousands of clinicians. Fourth, the array of tests differed considerably from those in the other batteries, in part because they were individually administered, and this increased the diversity of tests overall. The Hawaii Battery was supplemented with spatial ability tests because at the time MISTRA was initiated Bouchard and his students had a special interest in spatial ability. They considered it an important and under-explored domain (see Bouchard & McGee 1977; McGee 1979). That intuition has turned out to be correct as demonstrated by the work of Lubinski (2010), who took Bouchard's Individual Differences course at that time.

Supplementary materials related to this article can be found online at doi:10.1016/j.intell.2011.02.010.

Acknowledgments

Detailed discussion of the myriad sources of funding for MISTRA can be found in Segal (2012). Wendy Johnson has been supported since 2007 by the Research Councils of the United Kingdom [Grant number R40547]. The University of Edinburgh is a charitable body, registered in Scotland with registration number SC005336.

References

Asparouhouv, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling*, *12*, 411−434.

Bartholomew, D. J. (2004). *Measuring intelligence: facts and fallacies.* Cambridge: Cambridge University Press.

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.

Bouchard, T. J., & McGee, M. (1977). Sex differences in human spatial ability: not an x-linked recessive gene effect. *Social Biology*, *24*, 225−232.

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

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.

DeFries, J. C., Vandenberg, S. G., McClearn, G. E., Kuse, A. R., Wilson, J. R., Ashton, G. G., et al. (1974). Near identify of cognitive structure in two ethnic groups. *Science*, *183*, 338−339.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests.* Princeton, N. J.: Educational Testing Service.

Fanelli, D. (2010). "Positive" results increase down the heirarchy of the sciences. *PLoSOne*, *5*(4), e10068.

Flanagan, J. C., Dailey, J. T., Shaycoft, M. F., Orr, D. B., & Goldberg, I. (1962). *Studies of the American high School: Technical Report to the U.S. office of Education, Cooperative Research Project No. 226.* Pittsburgh: Project TALENT Office, University of Pittsburgh.

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.

Hakstian, A. R., & Cattell, R. B. (1975). *The comprehensive ability battery.* Champaign, IL: Institute for Personality and Ability Testing.

Ionnidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

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

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

Johnson, W., & Bouchard, T. J. (2005b). Constructive replication of the visual-perceptual-image rotation model in Thurstone's (1941) battery of 60 tests of mental ability. *Intelligence*, *33*, 417−430.

Johnson, W., Bouchard, T. J., 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. (2007). Replication of the hierarchical visual-perceptual-image rotation model in de Wolff and Buiten's (1963) battery of 46 tests of mental ability. *Intelligence*, *35*, 69−81.

Johnson, W., te Nijenhuis, J., & Bouchard, T. J. (2008). Still just 1 g: consistent results from five test batteries. *Intelligence*, *36*, 81−95.

Kuse, A. R. (1977). *Familial resemblances for cognitive abilities estimates from two test batteries in Hawaii.* PhD dissertation: University of Colorado at Boulder.

Lubinski, D. (2010). Spatial ability and STEM: a sleeping giant for talent identification and development. *Personality and Individual Differences*, *49*, 344−351.

Lykken, D. T. (1982). Research with twins: the concept of emergenesis. *Psychophysiology*, *19*, 361−373.

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

McGee, M. (1979). Human spatial abilities: psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, *86*, 889−918.

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

Muthen, L. K., & Muthen, B. O. (1998–2010). *Mplus: statistical analysis with latent variables user's guide* (6th Ed.). Los Angeles, CA: Muthen & Muthen.

Pietschnig, J., Voracek, M., & Forman, A. K. (2010). Mozart effect-Schmozart effect: a meta-analysis. *Intelligence*, *38*, 314−323.

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

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

Segal, N. (2012). *The real Minnesota twins: the landmark study of twins reared apart and reunited.* Cambridge, MA: Harvard University Press.

Spearman, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology*, *15*, 201−293.

Schmidt, F. (2010). Detecting and correcting the lies that data tell. *Perspectives on Psychological Science*, *5*, 233.

Thurstone, L. L., & Thurstone, T. G. (1941). *Factorial studies of intelligence.* Chicago: University of Chicago Press.

Van der Maas, H. L. M., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huiznega, 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.

Vinels, P. (2009). The skeptical epidemiologist. *International Journal of Epidemiology*, *38*, 675−677.

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