Available online at www.sciencedirect.com

SCIENCE DIRECT®

Intelligence 33 (2005) 393–416

INTELLIGENCE

The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized

Wendy Johnson*, Thomas J. Bouchard Jr.

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

Received 2 September 2004; received in revised form 20 November 2004; accepted 7 December 2004
Available online 17 February 2005

Abstract

In a heterogeneous sample of 436 adult individuals who completed 42 mental ability tests, we evaluated the relative statistical performance of three major psychometric models of human intelligence—the Cattell–Horn fluid-crystallized model, Vernon's verbal–perceptual model, and Carroll's three-strata model. The verbal–perceptual model fit significantly better than the other two. We improved it by adding memory and higher-order image rotation factors. The results provide evidence for a four-stratum model with a *g* factor and three third-stratum factors. The model is consistent with the idea of coordination of function across brain regions and with the known importance of brain laterality in intellectual performance. We argue that this model is theoretically superior to the fluid-crystallized model and highlight the importance of image rotation in human intellectual function.

Keywords: *g* factor; Fluid and crystallized intelligence; Verbal and perceptual abilities; Mental rotation; Spatial visualization; VPR theory

1. Introduction

Psychometric models of the structure of human intelligence contribute to scientific understanding in two important ways. First, they provide an organized and objective framework for evaluating the construct and predictive validity of the measurement tools that have been developed to assess the

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

0160-2896/$ - see front matter
doi:10.1016/j.intell.2004.12.002

abilities of individuals. Such tools are widely used to forecast outcomes in school as well as the worlds of work and everyday life (Frey & Detterman, 2004; Gottfredson, 2003; Kuncel, Hezlett, & Ones, 2004; Sackett, Schmitt, Ellington, & Kabin, 2001; Schmitt & Hunter, 2004). Second, by providing specific measurement models, they provide a falsifiable framework (c.f. Hunt, 2001; Lubinski & Benbow, 1995; Platt, 1964) for theoretical and empirical work in the neurosciences (Gray & Thompson, 2004), behavior genetics (Bouchard, 1998; Posthuma et al., 2003), epidemiology (Gottfredson, 2004; Gottfredson & Deary, 2004; Hart et al., 2003), cognitive psychology (Deary, 2001; Kane et al., 2004; Mackintosh, 1998), and aging (Finkel, Reynolds, McArdle, Gatz, & Pedersen, 2003; Salthouse, 2004). Much has been written about psychometric models of the structure of human intelligence, and they are routinely used as underlying assumptions in designing psychological research studies and for developing assessment tools. Surprisingly, however, the most well-established models have been subject to almost no empirical scrutiny in the form of assessment of comparative performance using modern confirmatory factor analytic techniques. In particular, Carroll's (1993) thorough and methodical exploratory analysis of more than 460 data sets of mental ability tests did not address this issue, a point he acknowledged in his final (2003, p. 12) publication, noting that his methodology "suffered from a lack of adequate procedures for establishing the statistical significance of findings". This is an important omission, as the objective evaluation of these models and the theories that generated them should result in more powerful theories, thereby making better use of monetary and intellectual resources and avoiding conceptual dead ends. The purpose of this study was to correct this omission, in the process using confirmatory factor analysis as a form of "strong inference" (Platt, 1964).

The dominant theoretical model of the structure of human intellect in the psychometric tradition is the theory of fluid and crystallized (Gf–Gc) intelligence. The Gf–Gc theory was developed initially by Cattell (1963, 1943), but since the early 1970s has been elaborated in greater detail by Horn (1976, 1985, 1998). Relative to other theories, the most important distinction made by this theory is between fluid (Gf) and crystallized (Gc) abilities, and the distinction has influenced research in virtually all domains of psychology. Gf reflects the capacity to solve problems for which prior experience and learned knowledge and skills are of little use. It is considered to be measured best by tests having little scholastic or cultural content such as perceptual and figural tasks, or verbal tasks that rely on relationships among common and familiar words. Gc reflects consolidated knowledge gained by education, access to cultural information, and experience. According to this theory, Gc reflects an individual's Gf as well as access to and selection of learning experiences. Consequently, among people of similar educational and cultural background, individual differences in Gf should strongly influence individual differences in Gc. Yet persons from different cultural backgrounds with the same level of Gf should differ in Gc. This is the theoretical basis for arguing that there are culture-fair (free) and culture-loaded intelligence tests.

This hypothetical causal model of the Gf–Gc distinction has led to the prediction that Gf should be under greater genetic influence than Gc. As acknowledged by Horn (1998), however, this prediction has been repeatedly disconfirmed. While there are other ways to posit relative genetic and environmental influences on Gf and Gc, this finding is a rather strong refutation of one of the major and novel predictions based on the theory as conceived by its conceptualizers, calling into question at least their developmental framework.

Gf–Gc theory has also been used to argue against the existence of general intelligence (Cattell, 1971; Horn, 1989), primarily out of the belief that the higher-order general intelligence factors arising from different batteries of tests would vary. This, however, was not the case for three widely known test batteries (Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004). That is, the three different test batteries yielded general intelligence factors that were completely correlated. The Johnson et al.'s study and the work of others (Gustafsson, 1988, 1999) suggest that, to a large extent, *g*, or general intelligence, and Gf are effectively equivalent. If one accepts the empirical findings noted above that contradict the theory, the question that remains is whether the structural distinction between fluid and crystallized intelligence is valid and continues to make a useful theoretical contribution.

Though the fluid-crystallized model has tended to dominate discussion, other models have been suggested. One of the most prominent of these was proposed by Vernon (1964, 1965). He stressed the importance of general intelligence in contributing to all mental abilities, but observed that, once a general intelligence factor is extracted from any collection of ability tests, the correlations among the residuals fall into two main groups. He labeled one of these *v:ed* to refer to verbal and educational abilities, and the other *k:m* to refer to spatial, practical, and mechanical abilities. The *v:ed* group, he noted, generally consists of verbal fluency and divergent thinking, as well as verbal scholastic knowledge and numerical abilities. The *k:m* group generally consists of perceptual speed, and psychomotor and physical abilities such as proprioception in addition to spatial and mechanical abilities. These broad groupings of residuals have practical importance and empirical support: working with large military samples throughout the 1950s, Humphreys (1962) found that *v:ed* and *k:m* added significantly to the prediction of work performance beyond *g*, but more narrow dimensions did not.

Noticeably absent from Vernon's theoretical discussion is the mention of memory as a separable ability, as Vernon suggested that memory demands tend to be distributed relatively evenly over the other abilities (Vernon, 1964). In addition, he believed there were various and important cross-links between the verbal and perceptual groups. For example, clerical tests often combine verbal ability and perceptual speed, and mathematics and science often depend both on number and spatial abilities. He also argued that the two main groupings reflect different cultural and educational pressures, with the *v:ed* factor arising from common schooling experiences, and the *k:m* factor arising from skills developed during non-educational experiences. According to this theory, there should not be any reason to expect that these factors would be under differing degrees of genetic influence. Vernon explicitly did not incorporate a factor comparable to Gf. He argued that, "Sometimes an inductive reasoning ability (also very relevant to science) can be distinguished, though most of the common variance of reasoning tests is apt to be absorbed into *g*." (Vernon, 1965, p. 725).

In more recent years, a consensus, to the extent it can be said to exist, has developed around Carroll's (1993) three-strata theory. Based on his monumental and systematic exploratory factor analysis of more than 460 data sets previously analyzed separately, Carroll proposed that mental abilities can be identified at three levels, or *strata*. Beginning with a description of the middle level following Carroll (p. 633), the second-stratum factors can be characterized as very general abilities lying in broad domains of behavior, yet they can be distinguished from the single, general, third-stratum factor because they involve moderate specialization of ability into spheres such as ratiocinative processes (Gf), prior acquisition of knowledge (Gc), learning and memory, visual perception, auditory perception, facile production of ideas, and speed. In contrast, the first-stratum factors are much narrower. They reflect greater specialization of abilities in specific ways that reflect particular experiences, exposure to learning opportunities, or the adoption of particular performance strategies. Carroll's theory can be considered to be a synthesis of the ideas of the many researchers who have discussed the structure of mental ability over the last hundred years. It is worth noting that he expressed strong reservations about the distinctiveness of Gf from *g*. Consider his recent analysis of two related data sets explicitly designed to articulate Gf–Gc theory (the Woodcock–Johnson Psycho-Educational Battery-Revised). With regard to the first data set he said, "In view of the undoubtedly careful and persistent efforts that were made in constructing these tests at the time the battery was being developed, the low Gf factor loadings most likely indicate that the factor Gf is inherently difficult to measure reliably independently of its dependence on *g* (as indicated by the high *g* loadings for these tests)." (Carroll, 2003, p. 14). This argument is virtually identical to Vernon's given above. With regard to the second data set he argued, "Some doubt is cast on the view that emphasizes the importance of a Gf factor in view of the relatively low factor loadings of some tests (numbered 07 and 21) on this factor." (p. 17). Indeed, in his discussion he concluded, "that more and better tests of factor Gf are needed to establish this factor as linearly independent of factor *g*, if indeed this is possible..." (p. 19).

The purpose of the current study was to develop and compare models reflecting each of the three major theories of the structure of intellect discussed above, making use of a set of 42 mental ability tests spanning a broad range of specific abilities administered to a single adult sample heterogeneous for age and educational background. To do this, it was necessary to specify the elements distinguishing each theory from the others because all are stated in relatively general terms. In addition, the theories have quite a bit in common with each other and we believed that it would be important to assess the influences of each of those elements consistently from model to model.

The fluid-crystallized model can be distinguished from the Vernon model in two major ways. First, Vernon very clearly emphasized the central role of general intelligence in the performance of all mental ability tasks. He described more specific ability factors in terms of relationships among residuals remaining after a general intelligence factor had been extracted. In contrast, both Cattell and Horn recognized an explicit causal link from fluid to crystallized intelligence and acknowledged substantial correlations between the two factors in the data sets they analyzed, though their conceptions of the manner in which they were linked differed to some degree. Both also denied that correlation between the two factors provided evidence for a general intelligence factor in any consistent way, relying on the claim that differences in the composition of test batteries would result in inconsistent measures of general intelligence from test battery to battery. Though this claim would appear to be unfounded (Johnson et al., 2004), their conception meant that they described their specific ability factors in terms of relationships among full scores rather than in terms of relationships among residuals after extraction of a general factor. The techniques of analysis for the two approaches are commonly different, but the results can be mathematically transformed one into the other (Gustafsson & Undheim, 1996), showing that they are merely alternative ways of specifying the same sets of relationships. For our purposes, we thus stated both the fluid-crystallized and Vernon models in terms of relationships among full scores to ensure consistency of interpretation. To the extent that the conception of general intelligence in the fluid-crystallized model is correct, we would expect a lower correlation between the fluid and crystallized factors than between the *v:ed* and *k:m* factors. To the extent that the conception in the verbal–perceptual model is more accurate, we would expect the opposite.

The second major way in which the fluid-crystallized model can be distinguished from the verbal–perceptual model is in the definition of the processes underlying the identification of the higher-order factors of fluid and crystallized intelligence and verbal and perceptual abilities. Clarity about these definitions is complicated by the fact that many researchers who have worked with the terms have tended to conflate 'fluid' intelligence with perceptual abilities and 'crystallized" intelligence with verbal abilities. This has occurred, no doubt, because the two sets of terms overlap to a substantial degree. Still, we believe that they reflect processes that can be distinguished in a straightforward way. As described by Cattell (1971), fluid and crystallized intelligence can be distinguished from each other by the relative absence of contributions of learned knowledge and skill to manifestations of fluid intelligence and the heavy emphasis of contributions of learning to manifestations of crystallized intelligence. Both Cattell (1971) and Horn (1989) were clear that this distinction applies across content boundaries, as they referred many times in their writings to high loadings on crystallized intelligence factors from mechanical knowledge and numerical reasoning tests. In contrast, Vernon's (1965) distinction between verbal and perceptual abilities followed the content areas and he was clear that, where the abilities involved in individual tests spanned the boundary between the two areas, the tests should receive cross-loadings on both in factor solutions. Thus, we would expect substantial loadings from tests involving the explicit use of pre-existing perceptual knowledge on a crystallized intelligence factor, but we would not expect such loadings on a *v:ed* factor. In addition, we would expect that tests that involve abstract reasoning with factual knowledge would receive substantial loadings on both fluid and crystallized intelligence factors, but that such tests would not receive substantial loadings on a *k:m* factor.

Because it is in many ways a synthesis of the ideas of many researchers in the field, it is more difficult to distinguish Carroll's (1993) three-strata model from either of the other two. In particular, as we noted above, Carroll questioned the distinction between fluid and crystallized intelligence. Nevertheless, he not only did not reject it, he explicitly incorporated it into his model at the second stratum level, below *g*, and treated it as accepted in his final published analyses (Carroll, 2003) providing additional empirical support for *g*. With regard to Vernon's theory he argued, "There is good evidence, for example, for clustering of variables around higher-order verbal–educational and spatial–mechanical factors, and for domination of these factors by some sort of general factor." (Carroll, 1993, p. 60). Thus, as he specified it, the aspect of the three-strata model that distinguishes it from both of the other two is the specification of an important factor of general intelligence and two strata of lower-order factors. Cattell's version of the Gf–Gc model includes three strata but the highest-order stratum consists generally of original fluid and crystallized intelligence, with the other two levels being broad groupings of abilities including memory, perceptual speed, visualization, and fluency; and narrow abilities such as would often be measured by individual tests; Horn's version of the Gf–Gc theory is a two-strata model; that is, Gf and Gc occur only at the second stratum along with various other factors. The Vernon model includes four strata: (1) general intelligence; (2) *v:ed* and *k:m*; (3) lower-order but still broad abilities such as fluency, language abilities, mathematical abilities, scientific and technical abilities, and spatial abilities; and (4) narrow abilities such as would often be measured by individual tests. Strictly speaking, it is not appropriate to distinguish the three-strata model from the other two in this way, as Carroll noted that there may be intermediate strata between the three distinct levels (1993, p. 635). Still, he wrote of these potential intermediate levels as tending to result from idiosyncratic compositions of test batteries rather than from more general properties of the structure of intellect, and we wished to test explicitly the idea that the additional complexity of both the fluid-crystallized and verbal–perceptual models is important in describing the structure of intellect.

The Vernon model can also be distinguished from the other two models because it lacks a memory factor.

2. Method

2.1. Research participants

The 436 (188 males, 248 females) research participants for this analysis came from the Minnesota Study of Twins Reared Apart (MISTRA). In addition to adult twins who were reared apart, the sample also includes adoptive and biological family members, friends, partners, and spouses of the twins. In most cases, the twins were separated early in life, reared in adoptive families, and not reunited until adulthood. They came from a broad range of occupations and socio-economic backgrounds and were primarily from North America, Great Britain, and Australia, and ranged in age from 18 to 79 (mean=42.7). Education levels spanned a broad range, varying from less than high school to post-graduate experience. The sample included 128 twin pairs, 2 sets of triplets, 117 spouses of twins, and 57 other biological and adoptive family members of the twins. MISTRA was initiated in 1979 and continued until 2000, with some participants returning for a second assessment 7 to 12 years after the initial one. The assessment consisted of a week-long battery of tests evaluating psychological (cognitive abilities, personality, interests, attitudes, etc.) and medical and physical traits. Details of recruitment and assessment are reported in Bouchard, Lykken, McGue, Segal, and Tellegen (1990) and Segal (2000). The mental ability tests were largely administered in blocks lasting 60 to 90 min across the 6 days of assessments. We describe each of the three cognitive ability batteries included in the assessment and relevant to this study in turn.

2.2. Measures

2.2.1. Comprehensive Ability Battery (CAB)

The CAB was developed by Hakstian and Cattell (1975), specifically to operationalize the fluid-crystallized model (Hakstian & Cattell, 1978). Because of this, comparison of the performance of the fluid-crystallized model with other theoretically driven models in this data set is particularly relevant. The CAB consists of 20 brief (5–6 min each) primary ability tests developed with the goal of measuring a broad range of well-replicated primary abilities. To avoid duplication of tasks in the extensive MISTRA assessment and make maximal use of available time, 6 of the tests in the CAB were not administered to the participants (Auditory Ability, Originality, Representational Drawing, Aiming, Spontaneous Flexibility, and Ideational Fluency). In addition, for this analysis we eliminated the test of Esthetic Judgment as we judged it not directly relevant to cognitive ability. The tests included in our version are described briefly in Table 1. Hakstian and Cattell (1978) reported split-half and retest reliabilities from the tests ranging from 0.64 for Perceptual Speed and Accuracy to 0.96 for Memory Span. As the Verbal Ability test consists of two completely different tasks, we tabulated the scores on the two parts separately, which meant that we had a total of 14 test scores.

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

The HB was developed to assess familial resemblance in cognitive ability in the Hawaii Family Study of Cognition (DeFries et al., 1974; Kuse, 1977). The HB consists of 15 tests of primary abilities; each test is short, requiring 3 to 10 min for administration. To avoid duplication of tasks and make maximal use of available time, two tests in this battery were not administered (Number Comparison and Social Perception). In order to articulate more clearly certain factors thought likely to be important, the battery

Table 1
Tests included in the three batteries

Test	Assessment activity
Comprehensive Ability Battery	
1. Numerical Ability	Computations including fractions, decimal divisions, square roots, etc.
2. Spatial Ability	Interpretation of two-dimensional figural rotation or reversal.
3. Memory Span	Recall of digits presented aurally.
4. Flexibility of Closure	Identification of embedded figures.
5. Mechanical Ability	Identification of mechanical principles and tools.
6. Speed of Closure	Completion of gestalt.
7. Perceptual Speed	Evaluation of symbol pairs.
8. Word Fluency	Production of anagrams.
9. Inductive Reasoning	Identification of pattern in sequences of letter sets.
10. Associative Memory	Rote memorization of meaningless pairings.
11. Meaningful Memory	Rote memorization of meaningful pairings.
12. Verbal—Vocabulary	Multiple choice among possible synonyms.
13. Verbal—Proverbs	Interpretation of proverbs.
14. Spelling	Multiple-choice identification of misspellings.
Hawaii Battery with Raven	
15. Card Rotations	Matching of rotated alternatives to probe.
16. Mental Rotation	Identification of rotated versions of two-dimensional prepresentation of three-dimensional objects.
17. Paper Form Board	Outline of cutting instructions to form the target figure.
18. Hidden Patterns	Identification of probe figures in more complex patterns.
19. Cubes	Identification of matched figures after rotation.
20. Paper Folding	Identification of unfolded version of a folded probe.
21. Raven	Identification of analogous figure to follow a sequence of figures.
22. Vocabulary	Multiple choice among possible meanings.
23. Subtraction/Multiplication	Completion of two-digit subtractions and two-digit by one-digit multiplications.
24. Word Beginnings/Endings	Generation of words beginning and ending with specified letters.
25. Pedigrees	Identification of familial relationshjps 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 two-dimensional designs using three-dimensional blocks.
41. Picture Arrangement	Chronological sequencing of pictures.
42. Object Assembly	Reassembly of cut-up figures.

was supplemented with four tests from the Educational Testing Services (Cubes and Paper Folding for spatial relations, Identical Pictures for perceptual speed and accuracy, and Different Uses for fluency), so there were 17 tests in the battery in total. The Hawaii study included a printed and shortened version of the Raven Progressive Matrices Test (1941). MISTRA utilized an untimed version of the Raven presented via slides (Lykken, 1982). Internal consistency and retest reliabilities for the tests ranged from 0.58 for Immediate Visual Memory to 0.96 for Vocabulary (DeFries et al., 1974).

2.2.3. The Wechsler Adult Intelligence Scale (WAIS)

The WAIS (Wechsler, 1955) is probably the best known and most widely used individually administered test of general intellectual ability. Wechsler believed that intelligence involved both abstract reasoning and the ability to handle practical situations involving performance and manipulative skills; thus, the WAIS includes Verbal and Performance subcomponents and many of the subtests require overt verbal articulation of reasoning based on factual knowledge. The subtests were also chosen to be suitable over a wide range of ages and for both sexes and to be appealing to examinees in the sense that they were not tedious or irrelevant. There are 11 subtests of the WAIS. Internal consistency reliabilities range from 0.79 for Comprehension to 0.94 for Vocabulary (Wechsler, 1955). For this sample, average WAIS full-scale IQ was 109.6 (range 79–140), normed at the 1955 level. The standard deviation was 11.8. Adjusted for secular changes in IQ using the average rates of change in WAIS scores summarized by Jensen (1998, p. 319),[1] the average WAIS full-scale IQ for the sample was 101.2 (range 61.1–139.9), with a standard deviation of 14.8. The standard deviation increased with this adjustment because IQ was *positively* correlated with age in this sample.

Table 1 summarizes the tests administered to MISTRA from the three batteries.

2.3. Statistical analyses

We used maximum likelihood confirmatory factor analysis to implement the models we evaluated. In doing so, we made no explicit adjustment in the results we present for the correlated nature of the observations for the twin pairs within our sample. This should have little effect on parameter estimates, but it will tend to make models appear to fit better than they actually do (McGue, Wette, & Rao, 1984). To avoid inappropriate conclusions resulting from this, we fit the same models using samples including only one randomly selected twin from each pair, with the identical pattern of results. Because the standard and most readily available chi-square measure of model fit generally indicates significant lack of fit in large samples, we used a chi-square statistic less than 2*degrees of freedom and Root Mean Square Error of Approximation (RMSEA) less than 0.05 (Browne & Cudeck, 1992) as indications of good model fit. RMSEA measures the extent of discrepancy between the model and data per degree of freedom; thus both alternative fit statistics provide some benefit to more parsimonious models. Model comparison among non-nested models by necessity relies upon information-theoretic fit statistics that emphasize minimization of the amount of information required to express the modeled data, which results in the selection of models that are the most parsimonious or efficient representations of the observed data. To compare our models, we made use of one of these fit statistics, the Bayesian

[1] The rates of change for verbal and performance IQ were weighted 0.6 and 0.4, respectively, to estimate change in full-scale IQ. Scores were individually adjusted downward from date of assessment to 1955 and upward by age at assessment in excess of 25. Other adjustment terms yielded similar results.

Information Criterion (BIC; Raftery, 1995). BIC estimates the natural log of the ratio of the posterior to the prior odds of the model in comparison to its saturated version, favoring the saturated model when positive and the presented model when negative. A difference in BIC of 10 between non-nested models is considered clear evidence in favor of the model with the more negative BIC. BIC heavily favors more parsimonious models.

In developing our models for comparison, we allowed only six second-stratum factors in each in order to ensure that the models were as directly comparable as possible. We chose the number six because the resulting factors could be justified within each of the theoretical frameworks and because the data correlation matrix had six eigenvalues greater than 1.0, giving the use of six factors an objective and generally accepted rationale. To maximize consistency of results, we assigned first-stratum tests to load on second-stratum factors uniformly unless there were substantive theoretical reasons or reasons of factor content to assign them differently. We allowed secondary or cross-factor loadings for individual tests where they were significant and theoretically appropriate. In addition, for the CAB, we followed the primary loading pattern for that battery given in Hakstian and Cattell (1978). This meant, in particular, that the fluid-crystallized and three-strata models had identical loading patterns from the first to second strata. In all three models, we allowed four sets of residual correlations resulting from analysis of modification indexes and consideration of test content. These residual correlations were between Immediate and Delayed Visual Memory from the HB (which have the same picture content as well as the same test format), between WAIS Digit Span and CAB Memory span (which have the same basic format), between WAIS Picture Completion and Picture Arrangement and between WAIS Object Assembly and HB Identical Pictures (all of which rely on picture identification).

The only constraints placed on the models we implemented were the structural forms given to them, the restrictions that tests load only on theoretically appropriate factors, and the restriction that most tests not have residual correlations. We did not develop expectations for or place constraints on the magnitudes of the factor loadings that would emerge. Instead, we estimated them freely and used the estimates that resulted to help evaluate the relative appropriateness of the models. This means, among other things, that when tests loaded on more than one factor, the definition of primary and secondary factors was determined by the magnitudes of the loadings that resulted rather than by assignment. There is always some disagreement about where particular tests should load—what is important here is the consistency of the approach from model to model. In particular, where we have allowed loadings that others might dispute, this should act to improve rather than impair model fit, and the resulting factor loadings should be insubstantial.

3. Results

For the fluid-crystallized model (Fig. 1), we obtained a chi-square of 1984.68 on 792 *df*, $p<0.00001$, with RMSEA=0.059 and BIC=−2828.81. The correlation between the two third-stratum factors was 0.85, strongly indicating the presence of a general intelligence factor given that such factors do not vary from test battery to test battery (Johnson et al., 2004). In addition, the loading of the second-stratum fluid factor on the third-stratum original fluid factor was 1.00, as was the loading of the second-stratum crystallized factor on the third-stratum school factor. This indicated that both the second-stratum fluid and crystallized factors were indistinguishable from their third-stratum counterparts.

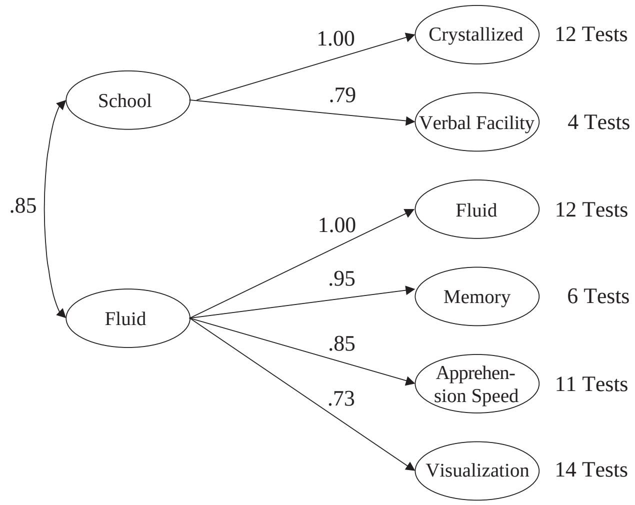

Fig. 1. Structural portion of fluid-crystallized model. Chi-square=1984.68, *df*=792, RMSEA=0.059, BIC=−2828.81. Some tests have primary and secondary factor loadings.

For the Vernon model (Fig. 2), we obtained a chi-square of 1646.07 on 783 *df*, $p<0.00001$, with RMSEA=0.050 and BIC=−3112.73. The correlation between the *v:ed* and *k:m* factors was 0.76, which clearly defined a general intelligence factor, though the loadings of the *v:ed* and *k:m* factors on the general factor were constrained equal in order to identify the model. Though the loading of the second-stratum verbal factor on the third-stratum *v:ed* factor was very high (0.97), each of the second-stratum factors was distinguishable from the third-stratum factors.

For the Carroll model (Fig. 3), we obtained a chi-square of 1979.92 on 791 *df*, $p<0.00001$, with RMSEA=0.059 and BIC=−2827.50. The loading of the second-stratum fluid factor on the third-stratum general intelligence factor was 1.00, indicating that it was indistinguishable from the general intelligence factor.

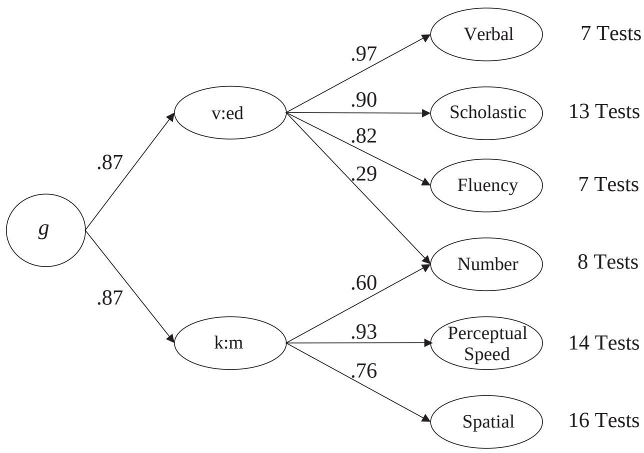

Fig. 2. Structural portion of verbal-perceptual model. Chi-square=1646.07, *df*=783, RMSEA=0.050, BIC=−3112.73. Some tests have primary and secondary factor loadings.

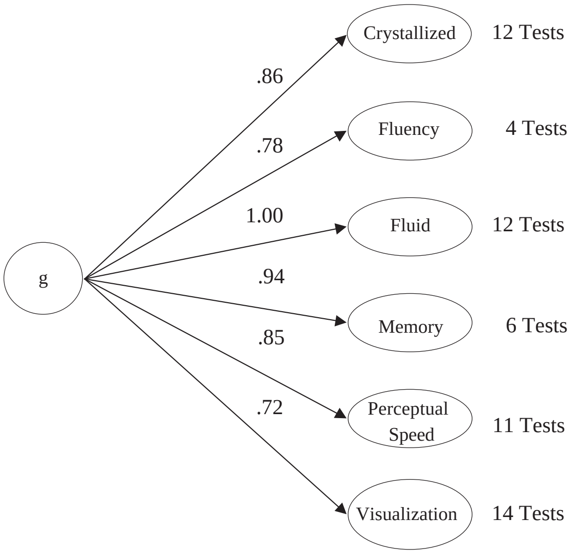

Fig. 3. Structural portion of three-strata model. Chi-square=1979.92, *df*=791, RMSEA=0.059, BIC=−2827.50. Some tests have primary and secondary factor loadings.

These model-fitting results suggested several conclusions. First, none of the three models fit really well, though all performed reasonably. Second, the fluid-crystallized model was effectively statistically equivalent to the three-strata model. The specification of the third stratum with distinct but correlated original fluid and school factors did not improve model fit, nor did the more parsimonious specification with a single general intelligence factor at that level provide any substantive improvement even in BIC, which heavily favors more parsimonious models. Third, the Vernon model fit substantially better than did the other two, as indicated by both RMSEA and BIC. The difference in BIC of about 285 suggests that the odds of the Vernon model being correct over those of either of the other two are in excess of 10^{100} to 1, rendering their actual quantification unimportant.

In addition to the model-fitting results, evaluation of the loadings from the first to the second strata is also relevant to assessment of the appropriateness of any of the three models. These loadings for the fluid-crystallized and Vernon models are shown in Table 2. We do not show the loadings for the Carroll model because they were effectively identical to those from the fluid-crystallized model. As the table indicates, there were several tests that we expected to show substantive loadings on both the second-stratum fluid and crystallized factors in that model, as they draw on both factual knowledge and reasoning ability. Yet three of these tests (CAB Numerical Ability, WAIS Comprehension, and WAIS Similarities) showed positive loadings on one factor and *negative* loadings on the other. Of these tests, only the WAIS Arithmetic test performed as we would have expected, generating a loading of 0.51 on the second-stratum fluid factor and a loading of 0.25 on the second-stratum crystallized factor. The Vernon model did not generate anomalous results of this nature, though the absence of a memory factor appeared to be important, especially for the Immediate and Delayed Visual Memory tests from the HB, which did not generate important loadings on any second-stratum factor. (The unique variable and factor variances from these models are shown in Table 3).

The results indicated clearly that the Vernon model had better verisimilitude than did either of the other two models. Yet the Vernon model still did not fit well by the criteria we had established, and it had

Table 2
Loadings from first to second strata for the Cattell–Horn fluid-crystallized, Vernon verbal–perceptual, and VPR models

Test	Primary loadings						Secondary loadings					
	Fluid-crystallized		Verbal–perceptual		VPR		Fluid-crystallized		Verbal–perceptual		VPR	
Comprehensive Ability Battery												
1. Numerical Ability	fluid	0.87	number	0.90	number	0.92	crystallized	−0.12	–		–	
2. Spatial Ability	visualization	0.62	spatial	0.63	rotation	0.78	–		–		–	
3. Memory Span	memory	0.59	number	0.45	number	0.40	–		fluency	0.21	fluency	0.26
4. Flexibility of Closure	visualization	0.35	perceptual speed	0.39	spatial	0.40	apprehension speed	0.35	spatial	0.29	number	0.32
5. Mechanical Ability	visualization	0.50	spatial	0.58	spatial	0.61	crystallized	0.15	–		–	
6. Speed of Closure	apprehension speed	0.58	perceptual speed	0.59	perceptual speed	0.35	–		–		fluency	0.32
7. Perceptual Speed	apprehension speed	0.75	perceptual speed	0.64	perceptual speed	0.52	–		number	0.13	number	0.31
8. Word Fluency	verbal facility	0.56	fluency	0.62	fluency	0.74	apprehension speed	0.39	perceptual speed	0.31	perceptual speed	0.19
9. Inductive Reasoning	fluid	0.70	number	0.41	spatial	0.40	–		spatial	0.36	number	0.38
10. Associative Memory	memory	0.54	number	0.48	cont. memory	0.48	–		–		number	0.21
11. Meaningful Memory	memory	0.64	scholastic	0.58	cont. memory	0.50	–		–		scholastic	0.32
12. Verbal-Vocabulary	crystallized	0.87	verbal	0.89	verbal	0.89	–		–		–	
13. Verbal-Proverbs	crystallized	0.66	verbal	0.80	verbal	0.80	fluid	0.14	–		–	
14. Spelling	crystallized	0.51	verbal	0.64	verbal	0.54	apprehension speed	0.34	perceptual speed	0.20	fluency	0.31
Hawaii Battery with Raven												
15. Card Rotations	visualization	0.58	spatial	0.58	rotation	0.69	–		–		–	
16. Mental Rotation	visualization	0.62	spatial	0.62	spatial	0.36	–		–		rotation	0.32
17. Paper Form Board	visualization	0.73	spatial	0.68	spatial	0.63	apprehension speed	0.05	perceptual speed	0.09	perceptual speed	0.17
18. Hidden Patterns	visualization	0.54	spatial	0.48	spatial	0.45	apprehension speed	0.28	perceptual speed	0.33	perceptual speed	0.37
19. Cubes	visualization	0.69	spatial	0.69	spatial	0.44	–		–		rotation	0.32
20. Paper Folding	visualization	0.74	spatial	0.74	spatial	0.75	–		–		–	

21. Raven	fluid	0.59	spatial	0.39	spatial	0.42	–		scholastic	0.32	scholastic	0.27
22. Vocabulary	crystallized	0.85	verbal	0.88	verbal	0.88	–		–		–	
23. Subtraction/ Multiplication[a]	apprehension speed	0.62	number	0.68	number	0.66	fluid	0.11	perceptual speed	0.09	perceptual speed	0.13
24. Word Beginnings/ Endings	verbal facility	0.88	fluency	0.88	fluency	0.85	–		–		–	
25. Pedigrees[b]	fluid	0.85	verbal	0.39	verbal	0.44	–		perceptual speed	0.33	perceptual speed	0.33
26. Things Categories	verbal facility	0.41	fluency	0.41	fluency	0.43	fluid	0.26	scholastic	0.26	scholastic	0.24
27. Different Uses	fluid	0.42	scholastic	0.43	scholastic	0.41	verbal facility	0.27	fluency	0.25	fluency	0.26
28. Immediate Visual Memory	memory	0.31	verbal	0.21	cont. memory	0.45	–		perceptual speed	0.09	–	
29. Delayed Visual Memory	memory	0.35	fluency	0.17	cont. memory	0.42	–		perceptual speed	0.14	–	
30. Lines and Dots	apprehension speed	0.38	perceptual speed	0.41	perceptual speed	0.42	–		–		–	
31. Identical Pictures	apprehension speed	0.48	perceptual speed	0.73	perceptual speed	0.80	visualization	0.29	–		–	
Weschler Adult Intelligence Scale												
32. Information	crystallized	0.82	scholastic	0.85	scholastic	0.86	–		–		–	
33. Comprehension	crystallized	0.94	scholastic	0.76	scholastic	0.77	fluid	−0.25	–		–	
34. Vocabulary	crystallized	0.85	scholastic	0.65	scholastic	0.64	–		verbal	0.28	verbal	0.30
35. Coding	apprehension speed	0.72	perceptual speed	0.50	perceptual speed	0.40	–		number	0.22	number	0.37
36. Arithmetic	fluid	0.51	number	0.44	number	0.43	crystallized	0.25	scholastic	0.39	scholastic	0.40
37. Similarities	crystallized	0.72	scholastic	0.75	scholastic	0.75	fluid	−0.02	–		–	
38. Digit Span	memory	0.56	number	0.32	fluency	0.36	–		fluency	0.31	number	0.27
39. Picture Completion	visualization	0.35	scholastic	0.37	scholastic	0.34	crystallized	0.29	spatial	0.30	–	
40. Block Design	visualization	0.76	spatial	0.77	spatial	0.78	–		–		–	
41. Picture Arrangement	visualization	0.33	scholastic	0.31	spatial	0.30	fluid	0.18	spatial	0.28	scholastic	0.28
42. Object Assembly	visualization	0.57	spatial	0.55	spatial	0.57	–		–		–	

[a] Subtraction/multiplication was constrained to 0 on crystallized intelligence due to model computational difficulties. For the same reason, Immediate and Delayed Visual Memory loaded on different factors in the verbal–perceptual model, which had no memory factor.

[b] Pedigrees had a tertiary loading of 0.22 on spatial in the verbal–perceptual model and 0.19 in the VPR model as well.

Table 3
Unique variable and latent factor variances from the Cattell–Horn fluid-crystallized, Vernon verbal–perceptual, and VPR models

Variable	Fluid-crystallized	Verbal–perceptual	VPR
Comprehensive Ability Battery			
1. Numerical Ability	0.35	0.19	0.16
2. Spatial Ability	0.62	0.61	0.38
3. Memory Span	0.65	0.64	0.63
4. Flexibility of Closure	0.60	0.60	0.59
5. Mechanical Ability	0.64	0.66	0.63
6. Speed of Closure	0.66	0.65	0.65
7. Perceptual Speed	0.44	0.44	0.43
8. Word Fluency	0.29	0.29	0.26
9. Inductive Reasoning	0.51	0.52	0.52
10. Associative Memory	0.71	0.77	0.62
11. Meaningful Memory	0.60	0.66	0.49
12. Verbal-Vocabulary	0.25	0.21	0.20
13. Verbal-Proverbs	0.38	0.36	0.35
14. Spelling	0.37	0.37	0.36
Hawaii Battery with Raven			
15. Card Rotations	0.67	0.67	0.53
16. Mental Rotation	0.61	0.62	0.60
17. Paper Form Board	0.42	0.44	0.43
18. Hidden Patterns	0.45	0.45	0.44
19. Cubes	0.53	0.52	0.51
20. Paper Folding	0.46	0.45	0.44
21. Raven	0.66	0.63	0.59
22. Vocabulary	0.27	0.22	0.22
23. Subtraction/Multiplication	0.49	0.44	0.43
24. Word Beginnings/Endings	0.23	0.23	0.28
25. Pedigrees	0.29	0.31	0.30
26. Things Categories	0.62	0.61	0.60
27. Different Uses	0.60	0.60	0.60
28. Immediate Visual Memory	0.90	0.92	0.69
29. Delayed Visual Memory	0.88	0.90	0.81
30. Lines and Dots	0.85	0.83	0.81
31. Identical Pictures	0.51	0.47	0.35
Weschler Adult Intelligence Scale			
32. Information	0.33	0.27	0.26
33. Comprehension	0.45	0.42	0.41
34. Vocabulary	0.20	0.18	0.18
35. Coding	0.49	0.52	0.53
36. Arithmetic	0.46	0.43	0.43
37. Similarities	0.50	0.44	0.43
38. Digit Span	0.68	0.67	0.66
39. Picture Completion	0.67	0.65	0.62
40. Block Design	0.42	0.41	0.38
41. Picture Arrangement	0.77	0.74	0.72
42. Object Assembly	0.68	0.70	0.65

Table 3 (*continued*)

Variable	Fluid-crystallized	Verbal–perceptual	VPR
Latent variables from fluid-crystallized model			
Crystallized intelligence	0.00	n/a	n/a
Fluid intelligence	0.00	n/a	n/a
Memory	0.10	n/a	n/a
Apprehension speed	0.28	n/a	n/a
Visualization	0.46	n/a	n/a
Verbal facility	0.38	n/a	n/a
Latent variables from verbal–perceptual models			
Verbal	n/a	0.06	0.08
Scholastic	n/a	0.21	0.23
Fluency	n/a	0.34	0.31
Number	n/a	0.27	0.40
Content memory	n/a	n/a	0.54
Perceptual speed	n/a	0.12	0.38
Spatial	n/a	0.45	0.31
Image Rotation	n/a	n/a	0.00

For the fluid-crystallized model, there were unique variable variance correlations of 0.17 between Picture Completion and Picture Arrangement, 0.22 between Things Categories and Different Uses, 0.23 between Digit Span and Memory Span, 0.10 between Identical Pictures and Object Assembly, and 0.42 between Immediate and Delayed Visual Memory. For the verbal–perceptual model, the analogous correlations were 0.15, 0.22, 0.23, 0.13, and 0.44. For the VPR model, the analogous correlations were 0.14, 0.22, 0.23, 0.12, and 0.33.

what seemed one obvious lack in the absence of a memory factor. We thus attempted in an exploratory manner to develop a Vernon-based model that did fit well according to the criteria we had established. We did this only after establishing the relative superiority of the Vernon model over the others in a tightly restricted comparison, in order to evaluate how the necessary alterations might refine Vernon's overall theory. We began by adding a memory factor (labeled content memory to distinguish it from the memory factor in the fluid-crystallized model). We also made three additional changes that improved model fit without compromising the integrity of the model as an implementation of Vernon's theory. These changes were as follows: (1) we moved the secondary loading for CAB Spelling from the perceptual speed factor to the fluency factor, (2) we moved the loading for CAB Flexibility of Closure from the Perceptual Speed factor to the number factor, making the loading on the spatial factor primary for that test, and (3) we added a loading on the fluency factor for CAB Speed of Closure. The loadings for these tests in the original Vernon model had been assigned to maximize consistency with the loadings for the other two models, and this constraint seemed no longer directly relevant. The resulting model, which we term the verbal–perceptual–memory model, had a chi-square of 1514.28 on 782 *df*, $p<0.00001$, with RMSEA=0.046 and BIC=−3238.44. It thus offered a substantial additional improvement in fit. RMSEA also indicated a close-fitting model, and the chi-square statistic was now less than 2**df*.

In the course of developing the verbal–perceptual–memory model we noted persistent and significant negative factor cross-loadings involving explicitly verbal and spatial ability tests, particularly those spatial ability tests involving three-dimensional mental rotation. Isolation of the tests involved resulted in identification of an additional third-stratum factor for image rotation, which also eliminated the contradictory cross-loadings. The resulting model, which we termed the verbal–perceptual–rotation (VPR) model, is shown in Fig. 4, and its factor loadings and unique variable and factor variances are

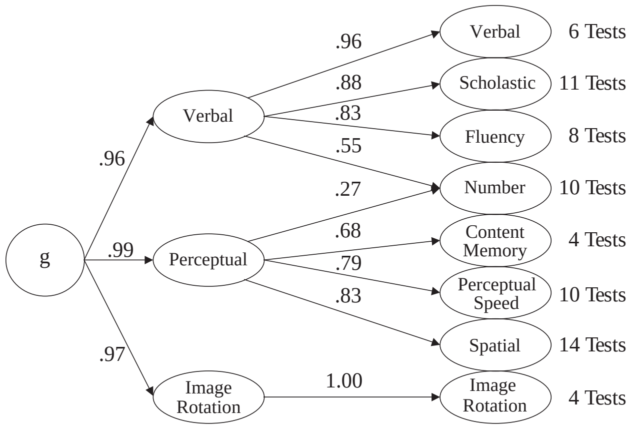

Fig. 4. Structural portion of verbal-perceptual-visualization model. Chi-square=1374.31, *df*=778, RMSEA=0.042, BIC=−3354.10. Some tests have primary and secondary factor loadings.

shown in Tables 2 and 3. It had a chi-square of 1374.31 on 778 *df*, $p<0.00001$, with RMSEA of 0.042 and BIC=−3354.10, indicating both a degree of fit which met our criteria for good fit and great improvement in fit over the verbal–perceptual–memory model in spite of the additional complexity. Interestingly, though the correlations between the verbal and perceptual and perceptual and image rotation factors were high (0.80 and 0.85), the correlation between the verbal and image rotation factors was much lower, 0.41. Each third-stratum factor loaded highly on the general intelligence factor. In fact, the loading of the perceptual factor on the general intelligence factor was .99, indicating their effective equivalence. We present the figure in the full hierarchical form because of its familiarity and generality. It is effectively equivalent to a model specifying no fourth-stratum general intelligence factor but with a third-stratum perceptual factor that contributes directly to both the second-stratum verbal and image rotation factors. The third-stratum ability groupings were important: when the image rotation factor was restricted to the second stratum, loading on the third-stratum perceptual factor, the chi-square was 1415.38 on 779 *df*, $p<0.00001$, with RMSEA of 0.043. This model clearly fit well by our criteria, but BIC was larger (−3319.10 vs. −3354.10), in spite of the fact that this model was more parsimonious than the full VPR model.

4. Discussion

Our primary goal in this study was to make use of a set of 42 mental ability tests administered to a single sample heterogeneous for age and educational background in order to develop and compare models reflecting Cattell and Horn's fluid-crystallized, Vernon's verbal–perceptual, and Carroll's three-strata theories about the structure of human intellect. The results were very clear: all three models provided reasonable structural descriptions, yet none met our criteria for a well-fitting model. At the same time, the Vernon model performed much better than either of the other two. In addition, the predictions derived from it regarding the relationship between third-stratum factors were accurate: the original fluid and school factors were more highly correlated than were the *v:ed* and *k:m* factors. This led us to use the Vernon model as the basis for the development of a model that did meet our fit criteria by adding second-stratum memory and image rotation factors. In addition, it appeared that the image rotation factor deserved representation at the level of the third stratum. In combination with our earlier findings regarding the consistency of general intelligence factors across test batteries (Johnson et al., 2004), our results point unequivocally to the existence of a general intelligence factor contributing substantively to all aspects of intelligence. They also call into question the appropriateness of the pervasive distinction between fluid and crystallized intelligence in psychological thinking about the nature of the structure of intellect. Instead, they suggest that a distinction between verbal and perceptual abilities may be more informative for many purposes and that the visualization processes involved in mental image rotation tasks have not been given the attention they deserve as important and relatively independent contributors to the manifestation of human intelligence.

4.1. Why the VPR theory is better than the fluid-crystallized theory

We began this paper by noting that psychometric models of the structure of human intellect offer (a) objective and rigorous (falsifiable) frameworks for studying genetic and neurobiological processes, and (b) insight into the relative accuracy of the measurement tools we use to assess the ability of individuals and predict their success in educational and occupational domains as well as everyday life. Factor analysis of paper-and-pencil tests of ability is a blunt tool to apply to these problems for two basic reasons. First, performance on any task reflects learned behavior to at least some degree. In addition, people likely differ in their prior exposure to any task presented as much as in their innate ability to approach any truly novel task. Consequently, it is never possible to measure innate ability per se and innate ability is always reflected to varying degrees in individual test scores. Second, most problems can be solved using multiple strategies, making it impossible to be sure that any specific task measures any specific ability. Nevertheless, it should be possible to reconcile the results of factor analytic studies of this type with findings in cognitive neuroscience, and the process of reconciliation should shed light on ways in which to improve both cognitive neuroscience studies and psychometric measurements.

What we need is a neurological conceptualization of intellectual performance that helps us choose between competing structural theories. Garlick (2002) has outlined one process that might explain the structures that we observe. He suggested that we build knowledge and skills through the growth of specific neuronal connections triggered by exposure to trait-relevant environmental stimuli. Individuals would thus be expected to differ in performance on mental ability tasks due to differences in their at least partly genetically influenced capacity to form neuronal connections in various parts of the brain as well as in their experiences, opportunities, and motivation to have previously developed such connections related to the task in question (Bouchard, Lykken, Tellegen, & McGue, 1996). This is borne out by studies comparing, for example, relevant brain structures and function in trained and untrained musicians and others with particular skills (e.g., Gaab & Schlaug, 2003). It is easy to draw an analogy between Gf (or *g*) and Gc and capacity to form neuronal connections and existence of task-relevant connections, raising the possibility that the process Garlick (2002) articulated may offer a neurological conceptualization of Gf–Gc theory.

The analogy may be apt but inappropriate to thinking about the structure of human abilities. Neuronal growth in response to exposure to stimuli is a process that takes place within an individual, however universal a process it may be. We seek to understand what individual differences in the manifestation of this presumably universal process reveal about the structure of human intellect, which means that we seek factors, or patterns of test performance *across* individuals. Jensen (1998, p. 95) nicely distinguished between processes and factors in their implications for understanding intellectual performance. We can understand *how* intellectual performance emerges in the individual through the use of Gf–Gc theory, but understanding *why* intellectual performance takes the particular pattern it does requires comparison across individuals. The distinction between fluid and crystallized intelligence can only emerge to the extent that it is possible to develop a battery of tests that accurately distinguishes between material to which test takers *should* have been exposed and tasks that they *should* find novel. This can be accomplished to some degree, but the high correlation (0.85) between the two third-stratum factors in our fluid-crystallized model suggests that the link between fluid ability and realized crystallized skills is very strong. People also differ, however, in their intrinsic motivation to select exposure to varying materials that may be related to tests (cf. Raine, Reynolds, Venables, & Mednick, 2002) as well as in the materials that have been presented to them due to differences in level and quality of education and in cultural background. A test that taps one aspect of ability in a person who has considerable experience with problems of the type presented taps another aspect of ability in a person to whom the problems are novel. Complicating this is the probability that general (fluid) intelligence can be used to address any problem, but it may compete relatively poorly with well-developed specific skills where they exist and are applicable. At the same time, as Mackintosh (1998, p. 171) pointed out, there is no reason to suppose that the ability to solve arbitrary abstract problems, such as those found in Raven's tests, is any less a learned skill than the ability to do mental arithmetic or to answer questions about the meaning of words.

Gf–Gc theory predicts that fluid abilities decline with age more rapidly than crystallized abilities (Cattell, 1971), and empirical evidence offers some mixed support for this prediction, particularly for abilities associated with processing speed or reaction time (e.g., Li et al., 2004). Understanding the Gf–Gc distinction from the perspective of neuronal development associated with specific skills and knowledge helps to explain how this might be the case: speed and the capacity to form new connections may decline with age much more rapidly than ability to access and use already established pathways. This would imply that, to whatever extent ability tests can consistently contrast between demanding use of deteriorating speed and capacities to form new connections and tapping existing pathways in the individuals tested, we would expect to see differences in performance with age in the one area but not to the same degree in the other. The relative inconsistency with which ability tests generally are able to make this distinction across individuals may be one of the major reasons for the muddiness of results in this area. The same neuronal development perspective can also help to explain the empirical failure of Gf–Gc theory's prediction that fluid abilities would be under greater genetic influence than would crystallized abilities, as we would expect that the complexity of the neuronal patterns developed would directly reflect the capacity for development.

Gf–Gc theory has been used to refine observations surrounding the secular increase in scores on intelligence test scores otherwise known as the Flynn effect (Dickens & Flynn, 2002). Abilities such as performance on vocabulary tests that are typically identified as crystallized show much smaller cohort effects than do those such as performance on inductive reasoning tasks that are typically identified as fluid (Jensen, 1998). This too can be understood from the perspective of neuronal development associated with the acquisition of specific skills and knowledge. Though most people's opportunities for exposure to classical vocabulary have undoubtedly increased in the past hundred years or so due to improved access to schooling, their opportunities for exposure to reasoning tasks in the form of computer technology, standardized tests, and games and household tasks requiring the use of inductive processes have exploded (Williams, 1998). It seems at least possible that, whatever secular trends may be occurring involving innate human mental capacity, the greater increase in performance on so-called fluid tasks may merely reflect greater development of related neuronal patterns.

Another neurological conceptualization of the structure and development of intellectual performance is consistent with the verbal–perceptual theory as articulated by Vernon (e.g., 1965) and with our VPR extension. This involves our general understanding of the functions of the left and right hemispheres of the brain. The left hemisphere of the brain is generally understood to be more in control of verbal and logical thought processes, while the right hemisphere is more in control of non-verbal, spatially oriented thought processes (e.g., Gray, 1999; Hugdahl, 2000; Toga & Thompson, 2003), though it is clear that all tasks of any complexity involve contributions from both hemispheres (Gray & Thompson, 2004). The factor analytic results of this study can be interpreted clearly within this general framework and the verbal–perceptual theory generates related predictions that can be tested. Specifically, we would expect that general intelligence will prove to be influenced by several to many genes responding to environmental stimuli to control biochemical processes acting throughout the brain. At the same time, we would expect that there will be several to many genes that influence brain functions that affect primarily verbal abilities and others that influence brain functions that affect primarily spatial and perceptual abilities. Some of these genes may even enhance abilities in one area at the expense of abilities in another, contributing to the lower correlation between verbal and perceptual abilities than between fluid and crystallized abilities in our models. Such environmentally mediated genetic processes may also help to explain differences in strategies used to approach cognitive tasks, individual differences in stimuli that attract attention, and sex differences in performance on various kinds of tasks. Evidence in support of these kinds of predictions is beginning to emerge.

4.2. Using Hakstian and Cattell's (1978) data to evaluate our conclusions

We used one other approach to examine the appropriateness of the fluid-crystallized model as a description of intellectual structure. In their important paper describing the fluid-crystallized factor structure of the CAB (which was designed to articulate Gf–Gc theory), Hakstian and Cattell (1978) provided their exploratory factor solution and the test correlation matrix. Probably limited by the numerical techniques available at the time, however, they did not provide any fit statistics for the factor solution they presented. We applied modern confirmatory factor analytic techniques to their test correlation matrix in order to evaluate the fit of their solution as well as to develop and compare our own solution. We specified the second stratum of their model with the loadings not indicated in bold in their paper constrained to 0. We were not able to specify their model with three third-stratum factors, as the model with the indicated constraints would not converge. A model with two third-stratum factors did, however, fit well using the fit criteria described above: chi-square was 240.02 with 156 *df*, p=0.00002, RMSEA=0.044, BIC=−639.01. Our solution contained a substantial negative loading for Speed of Closure on the general memory capacity factor, as did their solution, and the loading of Perceptual Speed and Accuracy on the general perceptual speed factor was trivial (−0.04) while the loading of Verbal on the crystallized factor was anomalously high (1.01). The solution we developed also had six second-stratum factors, which we labeled words (with loadings from Speed of Closure, Spelling, Auditory Ability, and Word Fluency), reasoning (with loadings from Numerical Ability, Spatial Ability, Perceptual Speed and Accuracy, Inductive Reasoning, Flexibility of Closure, and Spontaneous Flexibility), ideas (with loadings from Memory Span, Esthetic Judgment, Spontaneous Flexibility, Ideational Fluency, and

Originality), memory (with loadings from Associative Memory, Meaningful Memory, and Memory Span), knowledge (with loadings from Verbal and Mechanical Ability), and aim/draw (with loadings from Aiming and Representational Drawing). The two third-stratum factors generally reflected the verbal–perceptual distinction, with loadings from words, ideas, and knowledge on the verbal factor, and loadings from reasoning, memory, and aim/draw on the perceptual factor. This model fit substantially better, with chi-square of 220.14, 161 *df*, p=0.00136, RMSEA=0.036, BIC=−687.16. It also contained no anomalous loadings (negative, trivial, or greater than 1). Recall that the CAB was developed specifically in order to operationalize the fluid-crystallized model, yet it would appear that a verbal–perceptual model can provide a better description of its structure.

4.3. Complications in distinguishing between models

The potential to distinguish between the relative appropriateness of the fluid-crystallized and Vernon models is complicated by the paper-and-pencil format commonly used and by the existence of an educational system that is at least *generally* a common experience in developed nations. Testing vocabulary, commonly considered to be the quintessential example of both verbal and crystallized ability, serves to highlight the nature of the complications. The development of vocabulary is a fluid intellectual process; most people do not acquire their vocabularies by studying lists of words and definitions, but by inference in spoken or written situations they have sought or to which they are paying attention. Yet we test not the ability (and motivation) to engage in this inferential process, but the results of having engaged in it in the past (but see Fagan & Holland, 2002, for an example of an attempt to test the inferential ability). In addition, all human cultures are saturated with verbal content and this facility to deal with this content is transmitted to children by both parents (genetically and environmentally) and peers as well as by a generally common formal or informal educational system. The words selected for vocabulary tests tend to reflect the assumption that test takers have participated in the culture of those selecting the words to the same degree. Similar fluid-to-crystallized processes are involved in the development of perceptual abilities used in fields such as chess playing, engineering, physical science, and architectural design, but we tend to test these kinds of abilities by presentations that most people find relatively novel, thus requiring the use of fluid abilities.

Probably as a result of these complications, some have suggested (e.g., Gustafsson, 1984; Undheim, 1981) that Cattell's crystallized intelligence is equivalent to Vernon's *v:ed* factor. This study, however, makes clear that this is not the case. The crystallized intelligence and school culture factors (based closely on Cattell's definitions of those factors) included loadings from tests that were perceptual in nature, with no verbal component. Neither the primary nor the secondary verbal factors in our Vernon-based models included such loadings. At the same time, there were definite similarities between the sets of factors in the fluid-crystallized and Vernon-based models.

4.4. The role of spatial ability in human intellectual performance

The additions we made to the Vernon model to develop a model that met our fit criteria suggest that the structure of human cognitive ability is richly hierarchical, and that, in test batteries of sufficient breadth and depth, recognition of the strata within the hierarchy adds descriptive power. Our findings add to the growing body of evidence that measured intellectual ability is more than simply aptitude for schooling (Kuncel et al., 2004). A general factor (g) highly related to all aspects of intellectual function is apparent. At the same time, the more modest correlation between the third-stratum verbal and image rotation factors also suggests differential roles in intellectual performance for the two hemispheres of the brain, with some individuals more verbally than spatially adept, and vice versa. This has important implications for the development of a neurological understanding of intellectual performance, as visualization ability such as mental image rotation shows both genetic and environmental influences, as well as associations with handedness, sex, hormonal levels, and the development of brain hemispheric specialization (Halpern, 2000; Kimura, 1999; McGee, 1979; Sowell, Thompson, Holmes, Jernigan, & Toga, 1999). It is clearly an important aspect of human intellectual performance, and is often overlooked or assessed only poorly in cognitive batteries (e.g., the Woodcock–Johnson, McGrew & Woodcock, 2001).

Factor analytic studies of spatial ability tasks have typically provided strong support for the existence of two distinct spatial abilities, usually termed visualization and orientation (Hegarty & Waller, 2004; McGee, 1979). Visualization refers to the ability to mentally rotate, manipulate, and twist two- and three-dimensional objects in tasks such as those grouped here in our latent image rotation factor, while orientation refers to comprehension of the arrangement of elements within a visual pattern and the ability to retain spatial orientation with respect to one's body, even in changing conditions. Tests tapping both types of ability were included in our batteries; those involving orientation were subsumed under the second-stratum spatial factor. The abilities that formed our image rotation factor were effectively equivalent to visualization. Thus our results for a broad spectrum of abilities indicated an organization of spatial abilities consistent with those of other studies of specifically spatial abilities. Interestingly, it is the image rotation abilities that have repeatedly shown the most robust sex differences among cognitive abilities (favoring males, Voyer, Voyer, & Bryden, 1995).

The identification of a third-stratum factor associated with mental image rotation in these data also has important implications for education and career selection. It has been known for some time that performance on spatial tasks, particularly those involving image rotation, predicts success in fields such as airplane piloting, engineering, physical sciences, and fine arts better than does general intelligence, and especially verbal ability (Gottfredson, 2002; Humphreys & Lubinski, 1996; Shea, Lubinski, & Benbow, 2001; Sheppard, 1978). There is also evidence (Humphreys, Lubinski, & Yao, 1993) that failure to include assessment of such abilities in the standard batteries used for college and graduate school admissions is resulting in loss to those fields of potentially highly talented individuals. Perhaps of even greater concern, however, is the possibility that effects of this type may not be limited to the gifted and talented. Elementary school curricula tend to be used to educate those of all ability levels and they are generally much more highly focused on verbal than on image rotation abilities. This may be resulting in alienation from school of individuals unlikely to attend college as well as reducing the achievement of those who may. The social costs associated with early school leaving are well documented (e.g., Henry, Caspi, Moffitt, Harrington, & Silva, 1999).

The VPR model we have presented here must of necessity be considered somewhat open-ended. All models of this type are dependent to some degree on the specific tests included, with the range of abilities sampled as well as the level of specificity with which they are sampled determining the ability to distinguish among them (Horn, 1989). In addition, the ability to distinguish among them depends on the range of general intelligence existing within the sample tested (Vernon, 1965), with restriction of this range to higher levels of ability making more detailed distinctions possible. The range of general intelligence in the current sample is typical of an ordinary population. There is no reason to believe that the distinctions involving spatial and related abilities that we made arose from restriction of range. On the other hand, though the sample of abilities tested is extremely broad, other ability tests exist, and an even more extensive battery of tests might reveal additional factors at either the second or third strata without changing the basic form of the model. Were a specific ability to be investigated in great detail, additional levels involving that ability might also be observed. Still, our data strongly corroborate the evidence available from other, more achievement-oriented, sources that spatial image rotation ability is highly relevant to the overall structure of human intellect and that it is only moderately related to the verbal abilities that are so highly stressed in our educational systems.

Acknowledgements

We thank Matt McGue and Robert F. Krueger for technical advice regarding the analyses presented in this paper. We thank Nancy L. Segal, Margaret Keyes, and Kathy Corson for their many hours spent interviewing and administering tests to the participants of MISTRA. Finally, we thank the numerous co-investigators, MISTRA staff over the years, and the MISTRA participants themselves for making the project possible and for their enthusiastic participation.

References

Bouchard, T. J., Jr. (1998). Genetic and environmental influences on adult intelligence and special mental abilities. *Human Biology*, *70*, 257–279.

Bouchard, T. J., Jr., 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., Jr., Lykken, D. T., Tellegen, A., & McGue, M. (1996). Genes, drives, environment, and experience: EPD theory-revised. In C. P. Benbow, & D. Lubinski (Eds.), *Intellectual talent: Psychometrics and social issues* (pp. 5–43). Baltimore: Johns Hopkins University Press.

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

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

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports *g* and about 10 broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). Amsterdam: Pergamon.

Cattell, R. B. (1963). The theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1–22.

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

Cattell, R. R. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*, 153–193.

Deary, I. J. (2001). Human intelligence differences: Towards a combined experimental–differential approach. *Trends in Cognitive Sciences*, *5*, 164–170.

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

Dickens, W. T., & Flynn, R. J. (2002). The IQ paradox is still resolved: Reply to Loehlin (2002) and Rowe and Rodgers (2002). *Psychological Review*, *109*, 764–771.

Fagan, J. F., & Holland, C. R. (2002). Equal opportunity and racial differences in IQ. *Intelligence*, *30*, 361–387.

Finkel, D., Reynolds, C. A., McArdle, J. A., Gatz, M., & Pedersen, N. L. (2003). Latent growth curve analyses of accelerating decline in cognitive abilities in late adulthood. *Developmental Psychology*, *39*, 535–550.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or *g*? *Psychological Science*, *15*, 373–378.
Gaab, N., & Schlaug, G. (2003). Musicians differ from nonmusicians in brain activation despite performance matching. *Neuroscience and Music Annals of the New York Academy of Sciences*, *999*, 385–388.
Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review*, *109*, 116–136.
Gottfredson, L. S. (2002). The challenge and promise of cognitive career assessment. *Journal of Career Assessment*, *11*(2), 115–135.
Gottfredson, L. S. (2003). *g*, jobs, and life. In H. Nyborg (Ed.), *The scientific study of general intelligence: A tribute to Arthur R. Jensen* (pp. 293–342). Oxford: Pergamon.
Gottfredson, L. S. (2004). Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequities in health? *Journal of Personality and Social Psychology*, *86*, 174–199.
Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, *13*, 1–4.
Gray, J. R., & Thompson, P. M. (2004). Neurobiology of intelligence: Science and ethics. *Nature Reviews. Neuroscience*, *5*, 471–482.
Gray, P. (1999). *Psychology*. (3rd edition). New York: Worth.
Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*, 179–203.
Gustafsson, J. E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 35–71). Hillsdale, NJ: Erlbaum.
Gustafsson, J. E. (1999). Measuring and understanding *g*: Experimental and correlational approaches. In P. L. Ackerman, P. C. Kyllonen, & R. D. Edwards (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 275–289). Washington, DC: American Psychological Association.
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. (1975). *The comprehensive ability battery*. Champaign, IL: Institute for Personality and Ability Testing.
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. (2000). *Sex differences in cognitive abilities*. Mahwah, NJ: Erlbaum.
Hart, C. L., Taylor, M. D., Smith, G. D., Whalley, L. J., Starr, J. M., Hole, D. J., et al. (2003). Childhood IQ, social class, deprivation, and their relationship with mortality and morbidity risk in later life: Prospective observational study linking the Scottish Mental Survey of 1932 and the Midspan Studies. *Psychosomatic Medicine*, *65*, 877–883.
Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*, 175–191.
Henry, B., Caspi, A., Moffitt, T. E., Harrington, H., & Silva, P. A. (1999). Staying in school protects boys with poor self-regulation in childhood from later crime: A longitudinal study. *International Journal of Behavioral Development*, *23*, 1049–1073.
Horn, J. L. (1976). Human abilities: A review of research and theory in the early 1970's. *Annual Review of Psychology*, *27*, 437–485.
Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 267–300). New York: Wiley.
Horn, J. L. (1989). Models of intelligence. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29–73). Urbana, IL: University of Illinois Press.
Horn, J. L. (1998). A basis for research on age differences in cognitive capabilities. In J. J. McArdle, & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 57–91). Mahwah, NJ: Erlbaum.
Hugdahl, K. (2000). Lateralization of cognitive processes in the brain. *Acta Psychologia*, *105*, 211–235.
Humphreys, L. G. (1962). The organization of human abilities. *American Psychologist*, *17*, 475–483.
Humphreys, L. G., & Lubinski, D. (1996). Assessing spatial visualization: An underappreciated ability for many school and work settings. In C. P. Benbow, & D. Lubinski (Eds.), *Intellectual talent: Psychometric and social issues* (pp. 116–140). Baltimore, MD: John Hopkins University Press.
Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial ability in becoming an engineer, physical scientist, or artist. *Journal of Applied Paychology*, *78*, 250–261.
Hunt, E. (2001). Multiple views of multiple intelligence. *Contemporary Psychology*, *46*, 5–7.

Jensen, A. (1998). *The g factor*. Westport, CN: Praeger.
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.
Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology*, *133*, 189–217.
Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: MIT Press.
Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*, 148–161.
Kuse, A. R. (1977). *Familial resemblances for cognitive abilities estimated from two test batteries in Hawaii*. PhD dissertation, University of Colorado at Boulder.
Li, S., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science*, *15*, 155–163.
Lubinski, D., & Benbow, C. P. (1995). An opportunity for empiricism: Review of Howard Gardner's "Multiple intelligences: The theory in practice". *Contemporary Psychology*, *40*, 935–938.
Lykken, D. T. (1982). Research with twins: The concept of emergenesis. *Psychophysiology*, *19*, 361–373.
Mackintosh, N. J. (1998). *IQ and human intelligence*. Oxford: Oxford University Press.
McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influence. *Psychological Bulletin*, *86*, 889–918.
McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock–Johnson III technical manual*. Itasca, IL: Riverside Publishing Co.
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.
Platt, J. R. (1964). Strong inference. *Science*, *1466*, 347–353.
Posthuma, D., Baare, W. F. C., Pol, H. E. N., Kahn, R. S., Boomsma, D. I., & De Geus, E. J. C. (2003). Genetic correlations between brain volumes and the WAIS-III dimensions of working memory, perceptual organization, and processing speed. *Twin Research*, *6*, 131–139.
Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163.
Raine, A., Reynolds, C., Venables, P. H., & Mednick, S. A. (2002). Stimulation seeking and intelligence: A prospective longitudinal study. *Journal of Personality and Social Psychology*, *82*, 663–674.
Raven, J. C. (1941). Standardization of progressive matrices, 1938. *British Journal of Medical Psychology*, *19*, 137–150.
Sackett, P. R., Schmitt, N., Ellington, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education. *American Psychologist*, *56*, 302–318.
Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, *13*, 140–144.
Schmitt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*, 162–173.
Segal, N. L. (2000). *Entwined lives: Twins and what they tell us about human behavior*. New York: Plume.
Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*, 604–614.
Sheppard, R. N. (1978). Externalization of mental images and the act of creation. In B. S. Randhawa, & W. E. Coffman (Eds.), *Visual learning, thinking and communication* (pp. 133–190). New York: Academic Press.
Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L., & Toga, A. W. (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nature Neuroscience*, *2*, 859–861.
Toga, A., & Thompson, P. M. (2003). Mapping brain asymmetry. *Nature Reviews. Neuroscience*, *4*, 37–48.
Undheim, J. O. (1981). On intelligence II: A neo-Spearman model to replace Cattell's theory of fluid and crystallized intelligence. *Scandinavian Journal of Psychology*, *22*, 181–187.
Vernon, P. E. (1964). *The structure of human abilities*. London: Methuen.
Vernon, P. E. (1965). Ability factors and environmental influences. *American Psychologist*, *20*, 723–733.
Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270.
Wechsler, D. (1955). *Manual for the Wechsler adult intelligence scale*. New York: The Psychology Corporation.
Williams, W. M. (1998). Are we raising smarter children today? School- and home-related influences on IQ. In U. Neisser (Ed.), *The rising curve* (pp. 125–154). Washington, DC: American Psychological Association.