Are the General Factors From Different Child And Adolescent Intelligence Tests the Same? Results From a Five-Sample, Six-Test Analysis

Article *in* School psychology review · December 2013

4 authors:

Randy Granville Floyd
The University of Memphis

Matthew Reynolds
University of Kansas

John Kranzler
University of Florida

Ryan L Farmer
Western Kentucky University

School Psychology Review,
2013, Volume 42, No. 4, pp. 383–401

Are the General Factors From Different Child And Adolescent Intelligence Tests the Same? Results From a Five-Sample, Six-Test Analysis

Randy G. Floyd
University of Memphis

Matthew R. Reynolds
University of Kansas

Ryan L. Farmer
University of Memphis

John H. Kranzler
University of Florida

Abstract. Psychometric *g* is the largest, most general, and most predictive factor underlying individual differences across cognitive tasks included in intelligence tests. Given that the overall score from intelligence tests is interpreted as an index of psychometric *g*, we examined the correlations between general factors extracted from individually administered intelligence tests using data from five samples of children and adolescents ($n = 83$ to $n = 200$) who completed at least two of six intelligence tests. We found strong correlations between the general factors indicating that these intelligence tests measure the same construct, psychometric *g*. A total of three general-factor correlations exceeded .95, but two other correlations were somewhat lower (.89 and .92). In addition, specific ability factors correlated highly across tests in most (but not all) cases. School psychologists and other professionals should know that psychometric *g* and several specific abilities are measured in remarkably similar ways across a wide array of intelligence tests.

The term *positive manifold* refers to the pattern of universally positive correlations between cognitive task scores (i.e., scores from tasks on which performance results from dif-

We thank NCS Pearson, Inc., Lawrence Weiss, William Schryver, and Ying Meng for providing data from the DAS-II and KABC-II validity studies (Sample 1) and the Woodcock–Munoz Foundation, Richard Woodcock, Fredrick Schrank, and Kevin McGrew for providing data from the WJ III validity studies (Samples 2 and 3). We also appreciate manuscript proofing provided by Triche Roberson.

Please address correspondence regarding this article to Randy G. Floyd, University of Memphis, Department of Psychology, 202 Psychology Building, 400 Innovation Drive, Memphis, TN 38120; phone: 901-678-4846, fax: 901-678-2579, e-mail: rgfloyd@memphis.edu.

ferences in mental processing and not sensory acuity or motor skill; Jensen, 1998). This ubiquitous pattern seems to indicate that a general mental ability underlies individual differences on every kind of cognitive task. Spearman (1927) referred to this general mental ability as simply *g*, but it is often referred to as psychometric *g*. On the most widely used intelligence tests, the largest independent source of variance across persons is psychometric *g*. In addition, the predictive validity of intelligence tests is largely a function of psychometric *g* (Jensen, 1998; Neisser et al., 1996). The more closely a cognitive task measures psychometric *g*, the better it is at predicting a wide range of important social outcomes, such as success in school and work (Gottfredson, 2008).

At the current time, Carroll's (1993) three-stratum theory is arguably the most widely accepted taxonomy of human cognitive abilities. From patterns evident across the results of hierarchical factor analysis of more than 460 data sets from a variety of cognitive tasks, Carroll (1993) identified three levels of cognitive abilities representing varying degrees of generality: (a) Stratum I abilities that are narrow and highly specialized, (b) Stratum II abilities that are broader in nature, and (c) one very broad Stratum III ability, or psychometric *g*. Three-stratum theory posits that cognitive abilities can be arranged from specific to general and that psychometric *g* is best represented as a higher order variable correlated with all factors and test scores beneath it. For reference, the term *order* refers to the level of the latent variable in a factor analysis. First-order factors, for example, represent the latent variables underlying a matrix of correlations between scores, and second-order factors represent the latent variables underlying first-order factor correlations. In contrast, the term *stratum* refers to the degree of generality indicated by a factor, regardless of the order at which it appears in the analysis (Carroll, 1993).

Psychometric g in the Practice of Psychology

In schools, federal special education law (Individuals with Disabilities Education Improvement Act, 2004) mandates the assessment of intelligence for the identification of intellectual disability. The use of intelligence tests is allowed, but no longer required, for the identification of specific learning disability. Finally, intelligence tests are used in most states to identify students for intellectually gifted and talented programs (McLain & Pfeiffer, 2012). In other applied settings, diagnostic criteria (e.g., American Psychiatric Association, 2000) require that intelligence tests be administered. In particular, the American Association on Intellectual and Developmental Disabilities (2010) explicitly links its diagnostic criteria for intellectual disability to psychometric *g*. In all these instances, the overall score on intelligence tests (i.e., IQs), which is interpreted as an estimate of psychometric *g*, is most often used. Because intelligence tests (and IQs) are an integral aspect of applied psychology, it is important to determine the degree to which they converge in measuring the construct they intend to measure, psychometric *g*. It would be problematic if the psychometric *g* extracted from intelligence tests available to practitioners were vastly different.

Comparing General Factors Across Intelligence Tests

A criticism of psychometric *g* is that it is based on the specific battery of cognitive tasks from which it is derived. For example, Jensen (1998) stated, "There is no method of factor analysis that can yield exactly the same *g* when different tests are included in the battery. . . . The *g* is always influenced, more or less, by both the nature and the variety of tests from which it is extracted" (p. 85). Jensen (1998) described this effect as stemming from psychometric sampling error; he used this term to refer to imperfection in the measurement of psychometric *g* owing to inadequate sampling across a range of cognitive tasks. Psychometric sampling error is most evident when extracting a first-order general factor (as in nonhierarchical exploratory factor analysis) and cognitive tasks measuring the same Stratum

I or Stratum II ability are overrepresented in a factor analysis and others are underrepresented.

Breadth of psychometric sampling is particularly important for intelligence testing in practical settings, where the number of cognitive tasks (e.g., intelligence test subtests) that can be administered is quite limited, especially with children. Consequently, test developers must select a rather small number of cognitive tasks—typically, no more than 10—to measure the targeted Stratum II abilities in addition to psychometric *g*. Further, these cognitive tasks tend to differ across intelligence tests in the processes required (e.g., reasoning and memory) and in content (e.g., verbal and spatial). Because psychometric sampling produces varying batteries of cognitive tasks across intelligence tests, the general factor extracted from one battery may differ from that extracted from another.

Is the general factor the same when extracted from subtest scores from varying intelligence tests? This question addresses invariance under differential selection of variables (Thurstone, 1947; Mulaik, 2010). Two types of studies have been used to examine the invariance of general factors across intelligence tests. The first examines the consistency of the correlations between general factors and subtest scores, called *g* loadings, when the subtests used to identify the general factors vary. For example, these studies could determine if a subtest from the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV; Wechsler, 2003) has the same *g* loading when it is factor analyzed with other WISC-IV subtests as it does when it is factor analyzed with the Kaufman Assessment Battery for Children, Second Edition (KABC-II; A. S. Kaufman & N. L. Kaufman, 2004a) subtests. Because a subtest's *g* loading is expected to be relatively constant, the comparability of general factors would be seriously questioned if a *g* loading fluctuated widely across the general factors derived from different intelligence tests.

Researchers have used data from sizable samples of adults and exploratory factor analytic techniques to demonstrate that subtest *g* loadings do not vary substantially across general factors based on the characteristics and number of subtests used to identify them (Thorndike, 1987; Floyd, Shands, Rafael, Bergeron, & McGrew, 2009; Major, Johnson, & Bouchard, 2011). In particular, Floyd et al. (2009) and Major et al. (2011) showed strong consistency in the magnitude of subtest *g* loadings across analyses even when evaluating the effects of various targeted and error-related sources of variance. Across these studies, some evidence of psychometric sampling error affecting the measurement of psychometric *g* has been apparent, but variations in *g* loadings have neither been too great nor too consistent to weaken the construct validity of the general factor representing psychometric *g*.

The second type of study used to examine the invariance of psychometric *g* examines the correlations between general factors derived from factor analysis across intelligence tests. These studies provide a more direct test of the relations between factors representing psychometric *g* than those examining the consistency of subtest *g* loadings. In studies of this second type, participants complete two or more intelligence tests. A general factor is modeled within each intelligence test using confirmatory factor analysis. The correlation between the general factors modeled for each test is then estimated. Large correlations between the general factors indicate that the rank ordering of individuals on the general factor is consistent across the intelligence tests. One advantage of modeling these relations within a confirmatory factor-analytic framework is that the general factor for each intelligence test may be modeled as a higher order factor. A higher order factor should control for overrepresentation and underrepresentation of measures of the same Stratum I or Stratum II ability that may contribute to psychometric sampling error (Jensen, 1998).

Johnson, Bouchard, Krueger, McGue, and Gottesman (2004) used a confirmatory factor-analytic framework to examine the correlations of general factors using data from three intelligence tests yielded by more than 400 adult twins. Models for two of the tests included five first-order factors, and the model for the third test included three first-order factors; all test models included a second-order general factor. Correlations between second-order general factors were .99, .99, and 1.0, which were near-perfect and perfect correlations. Johnson, te Nijenhuis, and Bouchard (2008) replicated the Johnson et al. (2004) study and examined correlations between general factors across five intelligence tests completed by Dutch seamen. Models for four intelligence tests included three or four first-order factors and a second-order *g* factor. The model for one four-subtest intelligence test was the exception; it included one general factor because it was narrow in scope. Correlations between the general factors ranged from .77 and 1.0, but when the uniquely specified general factor from the test that was narrow in scope was omitted, general factor correlations across second-order factors were .95 or higher in every case.

Although all of the studies described thus far have used adult samples to examine the relations between general factors, two studies of note have examined these relations using data obtained from individually administered intelligence tests completed by children and adolescents. First, Keith, Kranzler, and Flanagan (2001) examined the correlations of general factors across two tests, the Woodcock–Johnson III Tests of Cognitive Abilities (WJ III; Woodcock, McGrew, & Mather, 2001) and the Cognitive Assessment System (CAS; Naglieri & Das, 1997). Keith and colleagues (2001) constructed models based on the Cattell–Horn–Carroll (CHC) theory (Schneider & McGrew, 2012), which is closely related to Carroll's three-stratum theory, using confirmatory factor analysis. Their test-specific models included either four or seven first-order factors and a second-order general factor. The correlation between the second-order general factors for the WJ III and the CAS was .98; this correlation was not statistically significantly different from one, indicating that the general factors were statistically indistinguishable.

Floyd, Bergeron, Hamilton, and Parra (2010) also examined the correlations of general factors across two tests, the WJ III (Woodcock et al., 2001) and the Delis–Kaplan Executive Function System (DKEFS; Delis, Kaplan, & Kramer, 2001). Floyd and colleagues constructed models based on CHC theory using confirmatory factor analysis. The correlation between a second-order general factor for the WJ III and a first-order general factor for the DKEFS was .99. Further, when the DKEFS model was altered to include first-order Verbal and Nonverbal factors and a second-order factor, the correlation between the two general factors across tests was 1.0. As evident in these studies, the near-perfect and perfect factor correlations for all appropriately identified general factors indicate that the same psychometric *g* is present across them.

Purpose of the Study

The purpose of this study is to extend research in this area by examining the relations between general factors derived from multidimensional intelligence tests administered to children and adolescents. Most of the prior research has focused on data from intelligence tests administered to adults, and only two peer-reviewed journal articles have reported results from analysis of data obtained from children and adolescents. Studies including children and adolescents are important to ensure that the findings generalize across age groups and tests designed for these populations. In addition, prior studies with children and adolescents have derived results from the WJ III and another test with a strong foundation in neuropsychological models (i.e., Delis et al., 2001; Naglieri & Das, 1997). However, no studies have examined the relations between general factors from pairs of the most frequently used intelligence tests for children and adolescents. From a practical perspective, school psychologists should know if the psychometric *g* measured by these intelligence tests is essentially the same or whether (and to what degree) it differs across tests. To address these issues, we identified five archival data sets from children and adolescents who completed at least two intelligence tests measuring a variety of ability domains. We hypothesized that all general factor correlations will be higher than .95.

Method

Data Sources and Participants

This study reports the results of analysis of data sets from five samples drawn during development of the Differential Ability Scales, Second Edition (DAS-II; Elliot, 2007), the KABC-II (A. S. Kaufman & N. L. Kaufman, 2004a), and the WJ III (Woodcock et al., 2001).

Sample 1. As described by Elliot (2007), the first sample consisted of 200 children and adolescents ages 6–17. Of those sampled, 100 were girls (50%) and 100 were boys (50%). A total of 70 were White (35%), 54 were Hispanic (27%), 50 were African American (25%), 13 were Asian (6.5%), and 13 were listed as Other (6.5%). Children completed the DAS-II and the WISC-IV, with the DAS-II completed first in all instances (Elliot, 2007).

Sample 2. As described by McGrew and Woodcock (2001), Phelps, McGrew, Knopik, and Ford (2005), Floyd, Bergeron, McCormack, Anderson, and Hargrove-Owens (2005), and Floyd, Clark, and Shadish (2008), the second sample consisted of 150 children ages 8–12 who were randomly selected from three public elementary schools. Of those sampled, 66 were girls (44%), and 84 were boys (56%). A total of 148 (98.7%) were White, and 2 (1.3%) were African American. They completed the Wechsler Intelligence Scale for Children, Third Edition (WISC-III; Wechsler, 1991) and the WJ III in counterbalanced order (McGrew & Woodcock, 2001; Phelps, McGrew, Knopik, & Ford, 2005).

Sample 3. As described by McGrew and Woodcock (2001), Floyd et al. (2005), Sanders, McIntosh, Dunham, Rothlisberg, and Finch (2007), and Floyd et al. (2008), the third sample consisted of 135 children and adolescents ages 8–13 randomly selected from public and private elementary schools. Of those sampled, 69 were girls (51.9%), and 64 were boys (48.1%). A total of 127 were White (95.5%), and 6 were African American (4.5%). They completed the DAS (Elliott, 1990) and the WJ III in counterbalanced order (McGrew & Woodcock, 2001).

Sample 4. Sample 4 and Sample 5 are treated as distinct in this study, but there is considerable overlap in participants across these samples. Participants were selected to represent the general child population and completed the KABC-II, the WISC-III, and the WJ III in counterbalanced order (A. S. Kaufman & N. L. Kaufman, 2004a). As described by A. S. Kaufman and N. L. Kaufman (2004a), Floyd et al. (2005), and Floyd et al. (2008), the fourth sample consisted of 116 children and adolescents ages 8–13. Of those sampled, 63 were girls (54.3%), and 53 were boys (45.7%). A total of 71 were White (61.2%), 18 were Hispanic (15.5%), 9 were African American (7.8%), 9 were Asian (7.8%), 1 was Native American (0.9%), 3 were listed as Other (2.6%), and 5 did not report race/ethnicity (4.3%).

Sample 5. As described by A. S. Kaufman and N. L. Kaufman (2004a), Floyd et al. (2005), and Floyd et al. (2008), the fifth sample consisted of 83 children and adolescents ages 8–13. Of those sampled, 37 were girls (44.6%), and 46 were boys (55.4%). A total of 50 were White (60.2%), 16 were Hispanic (19.3%), 8 were Asian (9.6%), 5 were African American (6.0%), 1 was Native American (1.2%), and 3 were listed as Other (3.6%).

Measures

Differential Ability Scales. The study included data from eight DAS subtests: Matrices, Pattern Construction, Recall of Designs, Recall of Objects-Immediate, Sequential and Qualitative Reasoning, Similarities, Speed of Information Processing, and Word Definitions. Subtest scores had median reliability coefficients above .75 across ages 8–12 in the norming sample. Please note that we refer to reliability coefficients in general in this article, but recognize that there was variation in how these coefficients were obtained (e.g., split-half reliability and test–retest reliability methods).

Differential Ability Scales, Second Edition. The study included data from 10 DAS-II subtests: Matrices, Pattern Construction, Recall of Designs, Rapid Naming, Recall of Digits Backward, Recall of Sequential Order, Sequential and Quantitative Reasoning, Speed of Information Processing, Verbal Similarities, and Word Definitions. Subtest scores had median reliability coefficients above .75 across ages 6–17 in the norming sample.

Kaufman Assessment Battery for Children, Second Edition. The study included data from 14 KABC-II subtests: Atlantis, Atlantis Delayed, Block Counting, Expressive Vocabulary, Number Recall, Pattern Reasoning, Rebus, Rebus Delayed, Riddles, Rover, Story Completion, Triangles, Verbal Knowledge, and Word Order. Subtests had median reliability coefficients above .75 across ages 7–12 years in the norming sample.

Wechsler Intelligence Scale for Children, Fourth Edition. The study included data from 14 WISC-IV subtests: Arithmetic, Block Design, Cancellation, Coding, Comprehension, Digit Span, Information, Letter-Number Sequencing, Matrix Reasoning, Picture Completion, Picture Concepts, Similarities, Symbol Search, and Vocabulary. Subtests had median reliability coefficients above .75 across ages 6–16 in the norming sample.

Wechsler Intelligence Scale for Children, Third Edition. The study included data from 12 WISC-III subtests: Arithmetic, Block Design, Coding, Comprehension, Digit Span, Information, Object Assembly, Picture Arrangement, Picture Completion, Similarities, Symbol Search, and Vocabulary. All subtests, excluding Object Assembly (.68) and Picture Arrangement (.72), had median reliability coefficients above .75 across ages 8–13 in the norming sample.

Woodcock–Johnson III Test of Cognitive Abilities. The study included data from 16 WJ III subtests: Analysis-Synthesis, Auditory Attention, Auditory Working Memory, Concept Formation, Decision Speed, Incomplete Words, Memory for Words, Numbers Reversed, Picture Recognition, Rapid Picture Naming, Retrieval Fluency, Sound Blending, Spatial Relations, Verbal Comprehension, Visual–Auditory Learning, and Visual Matching. All subtests, excluding Picture Recognition (.70), had median reliability coefficients above .75 across ages 8–13 in the norming sample.

Analysis

Models and model comparisons. Analyses were conducted using Mplus Version 6.11 (L. K. Muthén & B. O. Muthén, 1998–2011). Age-based standardized subtest scores were used as input in each analysis. Multiple stand-alone indicators of model fit were used, including χ^2, the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). CFI and TLI values greater than .95 indicate adequate fit, whereas values approaching .97 indicate excellent fit. For the RMSEA and SRMR, values of .05 or below indicate excellent fit, whereas the RMSEA may be as high as .08 and the SRMR may be as high as .10 for adequate fit (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Nested models were compared using the likelihood ratio ($\Delta\chi^2$) test, and non-nested models were compared using the Akaike Information Criterion (AIC). Lower AIC values indicate better fitting models.

Several steps in modeling, informed primarily by prior research findings and CHC theory (Schneider & McGrew, 2012), were followed across samples. The first step was to estimate baseline models for each data set. In these models, the factor structure of each intelligence test was modeled independently (i.e., all cross-test correlations were fixed to zero) for each sample. These models were expected to have poor fit because it was unlikely that the general factor and corresponding first-order factors were independent across tests; they were expected to be substantially correlated. In the steps that followed, the general factor correlation was freed first, and correlations between like first-order unique variances were freed second. Because measurement residuals (associated with subtests) contain both measurement error and reliable specific variance, it was likely that some measurement residuals were not independent across tests; correlated specific variances likely represented Stratum I factors or common method variance, and in our models, they were almost always highly predictable. Therefore, some subtest residual correlations were freed across tests primarily based on prior research and theory (i.e., subtests apparently measuring the same specific ability) and on rare occasions, based on model-based modification indexes above 10 (i.e., associated with the $p < .001$ level of statistical significance). We freed these additional correlations because we did not want restrictions on variance shared between tests at lower orders to interfere with estimates of higher order factor correlations. After each cross-test factor model was deemed acceptable and model parameters were within reasonable limits, the focus shifted to interpretation of the factor correlations between tests.

Intelligence research in general and research with intelligence tests used in this study (e.g., Reynolds & Keith, 2007) commonly found that the first-order Fluid Reasoning factor was often statistically indistinguishable from a second-order general factor (see Gustafsson, 1984). That is, the Fluid Reasoning unique variance is frequently not statistically significantly different from zero. We expected that this pattern would be evident in our models.

Missing data. There were few missing data across samples. Missing data never exceeded 10% for any variable—with one exception. In Sample 4, 38 participants were missing WISC-III Symbol Search subtest scores. All of the available data were analyzed, and maximum likelihood estimation was used under the assumption of data missing at random (Baraldi & Enders, 2010).

Results

Descriptive statistics for all subtests are presented in Tables 1 through 5 in the supplemental materials. Most subtest score means were slightly higher than the population mean (based on normative data from each test). Restriction of range in norm-referenced subtest scores was frequently evident (as indicated by standard deviation values that are smaller than population values); expansion of range was occasionally evident. All univariate skewness values were less than an absolute value of 1.0, and all but a small minority of kurtosis values (11.5%) were also less than an absolute value of 1.0. For those exceptions, kurtosis values ranged from 1.02 to 4.28. Thus, skewness and kurtosis values were well below values that have been found to give rise to problems with multivariate non-normality (skewness = 2.0 and kurtosis = 7.0; Curran, West, & Finch, 1996). Across all analyses, statistical significance was defined as $p < .05$.

Sample 1: DAS-II and WISC-IV

An ideal cross-test factor model is shown in Figure 1. In it, only corresponding factors are correlated freely across intelligence tests. Not shown in Figure 1 are correlations among subtest residuals across batteries that were previously described in the models and model comparisons section. The DAS-II model included five first-order factors with two subtest indicators per factor (see the supplemental materials for factor-to-indicator specification), and those factors regressed on a second-order general factor. Its first-order factors were Comprehension–Knowledge, Short-Term Memory, Processing Speed, Visual Processing, and Fluid Reasoning. The WISC-IV model also included five first-order factors, with two to five subtest indicators per first-order factor. Those first-order factors regressed on a second-order general factor. The five WISC-IV first-order factors were the same as those in the DAS-II model. For both models, the general factor was scaled by fixing the variance to one.

Model fit statistics for Sample 1 are presented in the top section of Table 1. The baseline model included two test battery-specific, higher order models that were independent of each other. Its fit was poor (Model 1).

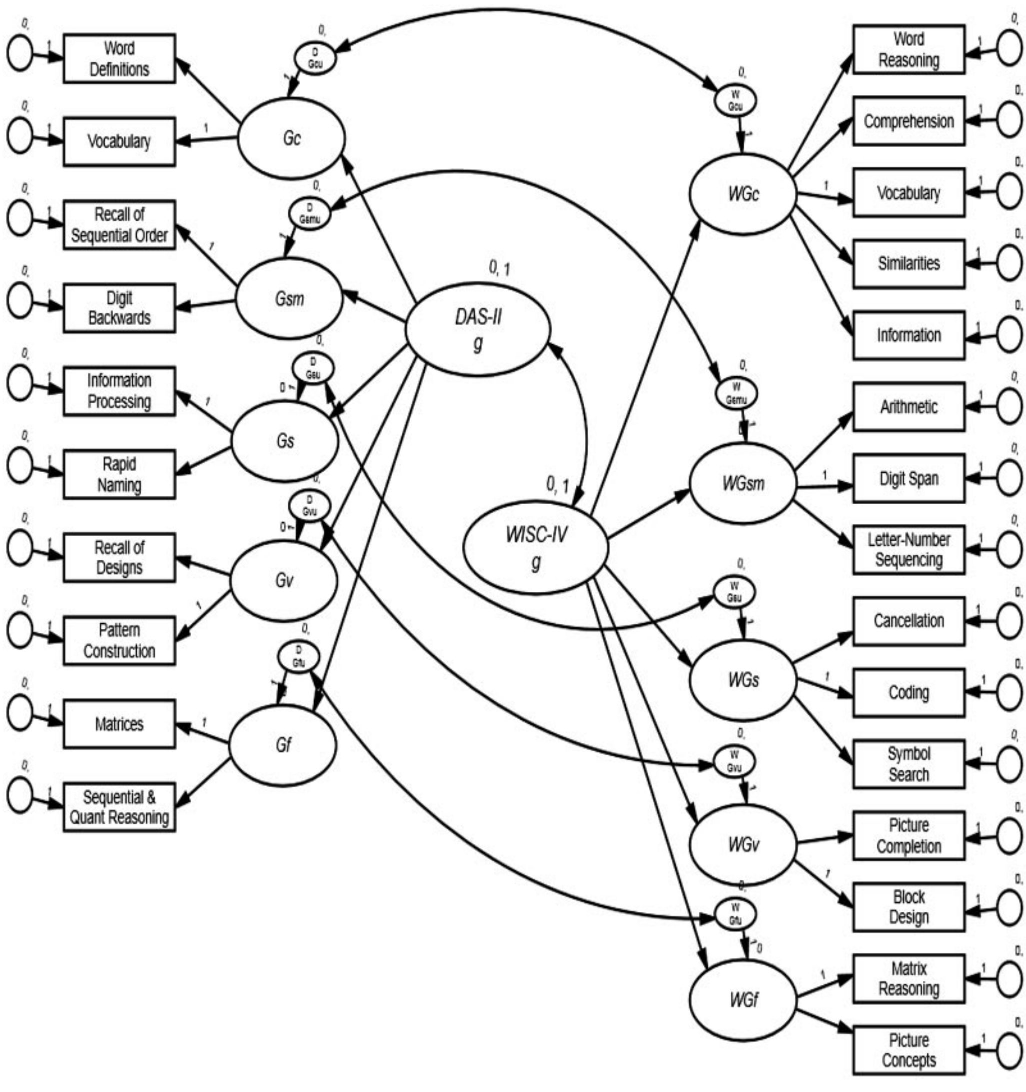

Figure 1. A cross-test factor model of the relations between second-order general factors and first-order unique variances from the DAS-II and the WISC-IV. Rectangles represent subtest scores, and large ovals represent the general factor and first-order factors. Small ovals above the large ovals represent first-order factor unique variances (measuring specific abilities not from the general factor or variance specific to subtests). Small circles represent measurement residuals that contain specific variance and measurement error associated with subtest scores. Curved arrows represent correlations between like abilities across batteries. DAS-II = Differential Ability Scales, Second Edition; WISC-IV = Wechsler Intelligence Scale for Children, Fourth Edition; g = the general factor, Gc = Comprehension–Knowledge, Gsm = Short-Term Memory, Gs = Processing Speed, Gv = Visual Processing, and Gf = Fluid Reasoning.

The Fluid Reasoning unique variances were zero in both the DAS-II and WISC-IV models; these unique variances represent Fluid Reasoning variance that is distinct from variance because of the general factor and from measurement error and subtest-specific reliable variance. Here, the general factor variance explained all of the Fluid Reasoning unique variance (see also Keith, Fine, Reynolds, Taub, & Kranzler, 2006). We fixed the Fluid Reasoning unique variances to zero, rendering a perfect correlation (i.e., loading) between Fluid Reasoning and the general factors.

There was not a statistically significant degradation in model fit; the general and Fluid Reasoning factors were statistically indistinguishable (Model 2). The correlation between the second-order general factors was freed next, improving model fit substantially (Model 3). Model fit, however, was inadequate. Thus, correlations between the pairs of Comprehension–Knowledge, Short-Term Memory, Processing Speed, and Visual Processing unique variances across batteries were freed. There was a statistically significant improvement in model fit (Model 4), but the correlation between the Short-Term Memory unique variances was greater than one, and the correlation between the Visual Processing unique variances was slightly greater than one. It is not unlikely for these factors to be correlated perfectly because they are such pure representations of the Stratum II abilities. Like the Fluid Reasoning unique variance, the Short-Term Memory loading on the general factor in the WISC-IV model approached one, and the WISC-IV Short-Term Memory unique variance was not statistically significant from zero. Therefore, the Short-Term Memory unique variance correlation across tests was deleted, the Short-Term Memory unique variance from the WISC-IV was fixed to zero, and the specific correlations between the WISC-IV Digit Span residual and both the residuals from the DAS-II Short-Term Memory subtests were freed (Model 5). Because the Visual Processing unique variance correlation was slightly greater than one (and not necessarily significantly different from one), the correlation between the WISC-IV Block Design subtest and DAS-II Pattern Construction subtest residuals was freed because of similarities in the design and response requirements of these subtests. Although model fit did not improve significantly (Model 6), it reduced the Visual Processing correlation to less than one. All parameter estimates were within reasonable limits.

The correlations between the first-order factor unique variances were Comprehension–Knowledge = .81, Visual Processing = .94, and Processing Speed = .90. The correlation between the general factors was .97. An additional model was estimated in which the correlation between the general factors was fixed to one; the fit of this model was compared with the previous model. The likelihood ratio test was then used to test whether the correlation between the general factors was statistically significantly different from one. Fixing the correlation between the general factors to one resulted in a degradation in model fit with a p value of .049 (Model 7); thus, the correlation between the general factors was statistically significantly different from 1.0 using a p value of $<.05$. The WISC-IV Arithmetic subtest may cross-load on the Fluid Reasoning factor in addition to the Short-Term Memory factor. We estimated models with this cross-loading. Although the model fit did not improve, the Arithmetic's factor loading on the Fluid Reasoning factor was indeed stronger than its loading on the Short-Term Memory factor. The Short-Term Memory factor did not collapse as a result, and the Short-Term Memory factors were strongly correlated. The magnitude of the general factor correlation was unaltered in this model, but the p value for the test of a perfect correlation was .024. Thus, the general factor correlation was not statistically significant from one.

Sample 2: WISC-III and WJ III

Results for Sample 2 are presented in the second section of Table 1. The WISC-III model included four first-order factors with two or four indicators per factor and with the first-order factors regressed on a second-order general factor. Its first-order factors were

Table 1
Fit Statistics and Indexes for All Models Across Samples

Sample 1: WISC-IV and DAS-II

	Models	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	CFI	TLI	RMSEA (90% CI)	AIC	SRMR
1	Baseline	820.89	265				.739	.705	.102 (.095, .110)	27320.0	.232
2	Gf unique variance = 0	822.09	267	1.20	2	.550	.740	.708	.102 (.094, .110)	27317.2	.232
3	Correlation between general factors	562.21	266	259.88	1	<.001	.864	.846	.074 (.066, .083)	27341.5	.066
4	Correlations between corresponding first-order factor unique variances	406.45	262	155.77	4	<.001	.932	.922	.053 (.042, .062)	26911.6	.056
5	Gsm unique variance = 0; correlations between Digit Span residual and both Recall of Sequential Order and Recall of Digits-Backward residuals	387.06	262				.941	.933	.049 (.038, .059)	26892.2	.052
6	Correlation between Block Design residual and Pattern Construction residual	385.01	261	2.06	1	.151	.942	.933	.049 (.038, .059)	26892.1	.052
7	Perfect *g*-to-*g* correlation	388.89	262	3.88	1	.049	.941	.931	.049 (.039, .059)	26894.0	.052

Sample 2: WJ III and WISC-III

	Model	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	CFI	TLI	RMSEA (90% CI)	AIC	SRMR
1	Baseline, WJ III Glr unique variance = 0	694.22	341				.722	.692	.083 (.074, .092)	27192.1	.174
2	Gf unique variance = 0	694.34	342	0.10	1	.742	.722	.693	.083 (.074, .092)	27190.2	.174
3	Correlation between general factors	545.47	341	148.86	1	<.001	.839	.821	.063 (.053, .073)	27043.3	.077
4	Correlations between corresponding first-order factor unique variances and correlation between Gc unique variance and Verbal Comprehension residual	450.32	337	95.15	4	<.001	.911	.900	.047 (.035, .058)	26956.2	.069
5	Correlations between Digit Span residual and Numbers Reversed residual and between Retrieval Fluency and Rapid Picture Naming residual	427.25	335	23.07	2	<.001	.927	.918	.043 (.029, .055)	26937.1	.066
6	Perfect *g*-to-*g* correlation	430.66	336	3.42	1	.064	.925	.914	.043 (.030, .055)	26938.5	.066

(Table 1 continues)

Table 1 Continued

Sample 3: WJ III and DAS										
Model	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	CFI	TLI	RMSEA (90% CI)	AIC	SRMR
1 Baseline	486.59	182				.642	.587	.108 (.097, .119)	21408.2	.204
2 Correlation between general factors	323.13	181	163.46	1	<.001	.833	.806	.076 (.063, .090)	21246.8	.078
3 Correlations between corresponding first-order factor unique variances and similar residual variance	268.79	175	54.34	6	<.001	.890	.868	.063 (.048, .078)	21204.5	.072
4 Correlations between Memory for Words residual and Recall of Designs residual	235.12	174	33.67	1	<.001	.928	.913	.051 (.033, .067)	21172.8	.070
5 Perfect *g*-to-*g* correlation	235.14	175	0.02	1	.888	.929	.915	.050 (.032, .066)	21170.8	.069

Sample 4: KABC-2 and WISC-III										
Model	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	CFI	TLI	RMSEA (90% CI)	AIC	SRMR
1 Baseline	600.04	288				.792	.710	.097 (.086, .108)	13809.2	.232
2 Correlation between general factors	472.00	287	128.04	1	<.001	.877	.860	.075 (.062, .086)	13683.2	.075
3 Correlations between corresponding first-order factor unique variances	388.98	284	83.02	3	<.001	.930	.920	.074 (.066, .083)	13606.1	.066
4 Gc subtest residuals correlated	378.20	278	10.78	6	.095	.931	.920	.056 (.041, .070)	13607.4	.074
5 Additional subtest residuals correlated	370.44	275	7.76	3	.051	.936	.925	.055 (.039, .069)	13605.6	.072
6 Perfect *g*-to-*g* correlation	388.94	276	18.51	1	<.001	.925	.911	.059 (.045, .073)	13622.1	.073

Sample 5: KABC-2 and WJ III										
Model	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	CFI	TLI	RMSEA (90% CI)	AIC	SRMR
1 Baseline	547.20	311				.792	.766	.096 (.082, .109)	13870.1	.244
2 Gf unique variance = 0	547.42	312	0.23	1	.635	.793	.767	.095 (.082, .108)	13868.3	.244
3 Correlation between general factors	444.91	311	102.51	1	<.001	.882	.867	.072 (.056, .087)	13767.8	.068
4 Correlations between corresponding first-order factor unique variances and subtest residuals	394.89	307	50.03	4	<.001	.923	.912	.059 (.040, .075)	13725.8	.064
5 Additional subtest residuals correlated	374.21	303	20.67	4	<.001	.937	.927	.053 (.032, .070)	13713.1	.063
6 Perfect *g*-to-*g* correlation	409.35	304	35.14	1	<.001	.907	.893	.065 (.047, .080)	13746.3	.114

Note. WISC-IV = Wechsler Intelligence Scale for Children, Fourth Edition; DAS-II = Differential Ability Scales, Second Edition; WJ III = Woodcock–Johnson III Tests of Cognitive Abilities; KABC-2 = Kaufman Assessment Battery for Children, Second Edition; CFI = comparative fit index; TLI = Tucker Lewis index; RMSEA = root mean square error of approximation; AIC = Akaike information criterion; SRMR = standardized root mean square residual; Gf = Fluid Reasoning; *g* = the general factor; Glr = Long-Term Storage and Retrieval; Gsm = Short-Term Memory; Gc = Comprehension–Knowledge. All nested comparisons of model fit (as evidenced in the $\Delta\chi2$ and Δdf columns) are made with the model immediately prior for each sample.

Comprehension–Knowledge, Visual Processing, Processing Speed, and Short-Term Memory. The WJ III model included six first-order factors, with two or three indicators per first-order factor. Its first-order factors were Visual Processing, Processing Speed, Fluid Reasoning, Short-Term Memory, Long-Term Storage and Retrieval, and Auditory Processing. First-order factors and the Verbal Comprehension subtest were regressed on a second-order general factor. (Technically, a first-order factor was created for the Verbal Comprehension subtest, but its variance was fixed to 0. No other subtest targeting the Stratum II ability Comprehension–Knowledge was available in the data set. Thus, a Comprehension–Knowledge factor could not be modeled.) The model would not converge because of the WJ III Long-Term Storage and Retrieval factor. The Long-Term Storage and Retrieval factor unique variance was fixed to zero, and the model converged. The fit of the baseline model, with independent factor models for each test, was poor (Model 1). Consistent with the previous models in Sample 1, the first-order WJ III Fluid Reasoning factor unique variance was fixed to zero; it correlated perfectly with the general factor (Model 2). Model fit did not degrade.

The correlation between the general factors was freed, which resulted in a substantial and statistically significant improvement in model fit (Model 3). Model fit, however, was not acceptable. The correlation between the WISC-III Comprehension–Knowledge unique variance and the Verbal Comprehension residual was freed. In addition, the correlations between the Visual Processing, Short-Term Memory, and Processing Speed unique variances between tests were freed. Fit improved substantially (Model 4). The Short-Term Memory unique variance correlation across tests was greater than one. The correlation between the WISC-III Digit Span subtest and WJ III Numbers Reversed residuals were freed because they are very similar tasks. The modification indexes indicated that, if the residuals of the WJ III Rapid Picture Naming and the WJ III Retrieval Fluency subtests were free to correlate, model fit would improve. Both of these tests require rapid verbal responses, and other research has found that these two subtests either load on the same factor or have correlated residuals (e.g., S. B. Kaufman, Reynolds, Liu, A. S. Kaufman, & McGrew, 2012). These additional parameters were freed, and the resulting model fit was acceptable. All parameter estimates were with reasonable limits.

The correlations between the first-order factor unique variances varied substantially: Visual Processing = .34, Short-Term Memory = .84, and Processing Speed = .97. The correlation between the general factors was .95. The likelihood ratio test indicated that the correlation between the general factors was not statistically significantly different from one (Model 6).

Sample 3: WJ III and DAS

Results for Sample 3 are presented in the third section of Table 1. The DAS model included three first-order factors with two indicators per factor, and each regressed on a second-order general factor. Its first-order factors were Comprehension–Knowledge, Visual Processing, and Fluid Reasoning. Two subtests, Recall of Objects and Speed of Information Processing, were also regressed directly on the general factor because no other subtests targeting the same Stratum II ability were available in the data set. The WJ III model included four first-order factors with two to three indicators per factor, and each regressed on a second-order general factor. Its first-order factors were Visual Processing, Fluid Reasoning, Short-Term Memory, and Auditory Processing. Three subtests, Verbal Comprehension, Visual–Auditory Learning, and Decision Speed were regressed directly on the general factor because no other subtests targeting the same Stratum II ability were available in the data set. The initial model fit poorly (Model 1).

The correlation between the general factors was freed. Model fit improved substantially; however, overall model fit was not acceptable (Model 2). Several expected correlations were freed in the next model, including the correlation between the Fluid Reasoning unique variances, the Visual Processing unique variances, the DAS Word Definition residual and WJ III Verbal Comprehension residual, the DAS Speed of Information Processing residual and the WJ III Decision Speed residual, and the DAS Object Recall residual and (a) the WJ III Picture Recognition residual and (b) the WJ III Visual–Auditory Learning residual. There was a statistically significant improvement in model fit (Model 3). Overall model fit, however, was not optimal. Examination of the modification indexes suggested that the WJ III Memory for Words residual and the DAS Recall of Designs residual should be correlated. Both subtests target memory abilities. The result was a statistically significant improvement in model fit (Model 4); model fit was considered adequate.

The correlation between Fluid Reasoning unique variances was .94. The correlation between Visual Processing unique variances was not statistically significantly different from zero. The correlation between the general factors was 1.00. The likelihood ratio test indicated that the correlation between the general factors was not statistically significantly different from one (Model 5).

Sample 4: KABC-2 and WISC-III

Results for Sample 4 are presented in the fourth section of Table 1. The KABC-II model included five first-order factors with two to four indicators per factor, and each first-order factor regressed on a second-order general factor. Its first-order factors included Comprehension–Knowledge, Visual Processing, Short-Term Memory, Long-Term Storage and Retrieval, and Fluid Reasoning. The residuals for the Atlantis subtest and the Rebus Learning subtest were correlated with their corresponding Delayed subtest residuals. The WISC-III model included four first-order factors with two to four indicators per factor, and each first-order factor regressed on a second-order general factor. The first-order factors were Comprehension–Knowledge, Visual Processing, Short-Term Memory, and Processing Speed. The baseline model fit poorly (Model 1). Model fit improved when the general-factor correlation was freed (Model 2).

Correlations between the Comprehension–Knowledge, Visual Processing, and Short-Term Memory unique variances were freed, and there was a statistically significant improvement in fit (Model 3). Nevertheless, the correlation between the Comprehension–Knowledge unique variances was greater than one. An additional model was estimated with the residuals of subtests used as indicators of the Comprehension–Knowledge factor correlated across tests (e.g., residuals for WISC-III Vocabulary and KABC-II Expressive Vocabulary subtests, which target word knowledge). Model fit improved (Model 4), and the correlation between Comprehension–Knowledge unique variances was reduced to .93. All of the correlated residuals were statistically significant except for two. Because the Visual Processing unique variance correlation was also greater than one, an additional model was estimated. In this model, correlations between the residuals for (a) the KABC-II Triangles subtest and both the WISC-III Block Design subtest and WISC-III Object Assembly subtest, and (b) the KABC-II Story Completion subtest and the WISC-III Picture Arrangement subtest, were freed based on prior research and theory. All parameter estimates were within reasonable limits, and model fit was adequate (Model 5).

The correlations between the first-order factor unique variances were Comprehension–Knowledge = .94, Visual Processing = .94, and Short-Term Memory = .85. The correlation between the general factors was .89. The likelihood ratio test indicated that the correlation between the general factors was statistically significantly different from 1.0 (Model 6).

Sample 5: KABC-II and WJ III

Results for Sample 5 are presented in the last section of Table 1. The KABC-II model included five first-order factors, with two to four indicators per factor, and each first-order factor regressed on a second-order general factor. Its first-order factors were the same as those for Sample 4, and the residuals for the

Atlantis and Rebus Learning subtests were correlated with their corresponding Delayed subtest residuals as in Sample 4. The WJ III model included six first-order factors, with two or three indicators per first-order factor. Its first-order factors were Comprehension–Knowledge, Visual Processing, Processing Speed, Fluid Reasoning, Short-Term Memory, and Auditory Processing. Those first-order factors and the Visual–Auditory Learning subtest were regressed on a second-order general factor. No other subtests targeting the same Stratum II ability as Visual–Auditory Learning subtest were available in the data set.

The fit of the baseline model indicated poor fit (Model 1). The KABC-II Fluid Reasoning unique variance was zero, and it was fixed to zero (Model 2). Model fit improved by allowing for the correlation between the two general factors to be estimated freely (Model 3). Correlations between the Comprehension–Knowledge, Visual Processing, and Short-Term Memory unique variances were freed. In addition, the KABC-II Long-Term Storage and Retrieval unique variance and the residual from the WJ III Visual–Auditory Learning subtest were correlated because they target associative memory ability. Allowing for these four correlations to be estimated freely improved model fit (Model 4). The Comprehension–Knowledge unique variance correlation was larger than one, so one additional model was estimated with these correlations freed: (a) the WJ III General Information residual and both the KABC-II Verbal Knowledge subtest and KABC-II Expressive Vocabulary residuals, (b) the WJ III Verbal Comprehension residual and the KABC-II Riddles residual, and (c) the WJ III Concept Formation residual and the KABC-II Pattern Reasoning residual. These correlations improved model fit (Model 5); parameter estimates were within reasonable limits.

The correlations between the first-order factor unique variances varied substantially: Comprehension–Knowledge = .98, Visual Processing = .43, and Short-Term Memory = .88. The correlation between the general factors was .92. The likelihood ratio test indicated that the correlation between the general factors was statistically significantly different from 1.0 (Model 6).

Discussion

School psychologists and others involved in assessment should know if the latent general ability measured by intelligence tests, psychometric *g*, is essentially the same or whether it differs across them. Thus, this study investigated the relations between general factors from varying intelligence tests frequently used to assess school-age children and adolescents.

General Factor Relations

We hypothesized that all general-factor correlations would be higher than .95, but this hypothesis was not supported. The average general-factor correlation across the five samples was .95, three correlations were .95 or higher, and two correlations were not statistically significantly different than 1.0. Two other general-factor correlations, however, were .89 and .92. These last two correlations were slightly lower in magnitude than expected based on previous research. Johnson et al. (2004) reported general-factor correlations from .99 or 1.0 across analyses, and Johnson et al. (2008) reported them as being .95 or higher with one exception. Keith et al. (2001) reported a general-factor correlation of .98, and Floyd et al. (2010) reported general-factor correlations of .99 and 1.0. Although both of the lowest correlations found in this study involved the KABC-II, we have no substantive explanation for these findings based on our modeling results as well as prior research. Sampling error associated with the two smallest data sets (*N* values = 116 and 83), which happened to involve the KABC-II, is one explanation.

Collectively, these findings suggest that the general factors yielded by the targeted intelligence tests are nearly indistinguishable. One of the reasons we found instances of lower-than-expected general-factor correlations is that we employed a slightly different methodology than many other researchers. For example, neither Johnson et al. (2004) nor

Johnson et al. (2008) appear to have included correlated first-order unique variances to represent shared Stratum II factors across their test-specific models. Our modeling indicated that failure to correlate these variances caused the second-order general-factor model correlations to be inflated.

Additional referencing of prior research findings is necessary to interpret the general-factor correlations. Recently, S. B. Kaufman et al. (2012) employed similar methods as those used in this study to examine the relations between second-order general factors from the KABC-II and WJ III and second-order factors from their corresponding achievement tests that target reading, mathematics, and writing skills (A. S. Kaufman & N. L. Kaufman, 2004b; Woodcock et al. 2001). Correlations ranged from .77 to .94 across age levels and yielded a median correlation of .80. These comparisons indicate that our methods do not produce near-perfect correlations inherently, and they make apparent that correlations between general factors targeting psychometric *g* are notably higher than correlations between these same general factors and general factors of academic achievement.

Floyd et al. (2008), using four of the same data sets (Samples 2, 3, 4, and 5), reported correlations between IQs across tests that were notably lower than the general-factor correlations found in this study. The IQ correlations were .78 on average, which is in the typical range (Jensen, 1998). When we calculated the correlations between the most global IQs from each test yielded from each data set (including Sample 1), the average correlation was .76. A discrepancy between these average correlations across methods (.76 to .78 for IQs and .95 for general factors) is likely explained by general factors (as latent variables) being perfectly reliable. However, a discrepancy this large (considering many IQs have internal consistency reliability estimates of .95 or higher) is surprising; in fact, when we corrected our IQ correlations for attenuation, the average correlation rose from .76 to .82 (range = .75 to .86). Although the general factors in this study included some subtest indicators that are not subtests that contribute to the IQs in question, we surmise that overrepresentation of some Stratum II ability measures and underrepresentation of others (i.e., psychometric sampling error) in the composition of the IQs leads to their lower correlations with other IQs.

Specific Ability Relations and Independence from Psychometric *g*

Although we did not plan to investigate the relations between factors representing Stratum II abilities (as represented by first-order unique variances) across intelligence tests, we found strong correlations between them in most cases. For example, correlations between Comprehension–Knowledge factors ranged from .98 to .81, correlations between Short-Term Memory factors ranged from .88 to .84, and correlations between Processing Speed factors ranged from .97 to .90. One model permitted a correlation between Fluid Reasoning factors; this correlation was .94. There was, in contrast, much variation across the correlations between Visual Processing factors; two were strong with correlations of .94, but two others were weak to moderate (with correlations of .43 and .34). A final correlation between Visual Processing factors was not statistically significant from zero. The weaker correlations between these factors were found in the three samples with the WJ-III, suggesting Visual Processing may be measured differently by that test than by the other three other tests we included. In general, the variation in the size of the correlations between factors representing these Stratum II abilities mirrors that evident across correlations between composite scores representing these same abilities (see Floyd et al., 2005).

In previous research, similarities between psychometric *g* and Fluid Reasoning have been widely noted (see Gustafsson, 1988). In half of the eight test-specific models in this study, the Fluid Reasoning factor was statistically indistinguishable from the general factor. In addition, two other factors, the WISC-IV Short-Term Memory factor and the WJ III Long-Term Retrieval factor, also were statistically indistinguishable from the general factor, but these findings were evident in only one of nine and one of three test-specific models, respectively. Such results suggest that these abilities, independent of psychometric *g,* may be inadequately measured by some intelligence tests (see Carroll, 2003). However, these perfect relations with the general factor were evident in only a small minority of the test-specific models in which these factors were specified. Independent Stratum II abilities have appeared repeatedly across factor-analytic studies (see Keith & Reynolds, 2010); the perfect relations noted here are likely from sampling error and the inability to distinguish the general factor from first-order factors because of low statistical power (see Matzke, Dolan, & Molenaar, 2010). These findings of perfect relations between Stratum II abilities and the general factor are probably not reliable.

Limitations, Caveats, and Directions for Future Research

Two sets of limitations were present in this study: those associated with our samples and those associated with our modeling. First, archival data were used in lieu of new data collection. Thus, variations in data collection that were not noted in previous publications may have affected our results in unknown ways. Similarly, although we relied on multiple samples rather than a single sample and considered patterns of findings across replications, our modest sample sizes limit our confidence in interpretation and the power of our statistical significance tests. Although we had adequate power to find statistically significant results in most of our analyses, our ability to detect whether the general factor correlations were significantly different from one may have been affected by our modest sample sizes. Larger samples are clearly needed to understand the invariance of these latent abilities (Reeve & Blacksmith, 2009). Second, some decisions we made while modeling the relations between factors appear to have affected our results. For example, based on prior research and theory, we correlated some subtest residuals—affecting the magnitude of the correlations between first-order factors—and some first-order factor unique variances—affecting the general factor correlations. We believe that these steps were prudent, but they are at odds with other models evident in recent research (Johnson et al., 2004, 2008).

It is important to note that there are alternative explanations provided for what may give rise to positive manifold that does not involve a single general factor (Thomson, 1951; van der Maas et al., 2006). Future researchers should evaluate these alternate explanations and the nature of psychometric *g*. Finally, we targeted the measurement of psychometric *g* and more specific abilities underlying intelligence tests, but we did not partition the ability-related variance associated with any individual scores from these tests. To extend this line of research, future investigators should determine how well the overall scores from intelligence tests, the IQs, measure psychometric *g*.

Implications for Research and Practice

We see three implications for research and practice stemming from our findings. First, although most intelligence tests consist of fairly small batteries of cognitive tasks, the general factor derived from the intelligence tests used in this study were consistently highly correlated and sometimes perfectly correlated. Thus, intelligence tests are mature technologies that measure psychometric *g* with a high degree of fidelity (Kamphaus, 2009). Nevertheless, in some instances, these correlations were lower than expected, indicating that the psychometric *g* measured across some intelligence tests was not exactly the same. Second, correlations between the factors representing Stratum II abilities indicate that the most commonly used intelligence tests also measure more specific abilities in similar ways. In particular, Comprehension–Knowledge, Short-Term Memory, Processing Speed, and Fluid Reasoning are essentially the same across intelligence tests. There are, however, potential construct validity problems with Visual Processing factors that raise concerns about interpreting measures of this ability across batteries.

Third, school psychologists should consider how Stratum III and Stratum II abilities operate at the level of score interpretation. It is well known that no score yielded by intelligence tests (or any other measurement instrument) is a pure measure of the construct it targets. No matter the efforts taken to reduce measurement error, it is a ubiquitous and unremitting source of score differences. All the other reasons for score differences across individuals can typically be attributed to particular abilities (Carroll, 1993). The models tested in this study make distinct this separation of the effect of the psychometric *g* and the effects of specific abilities, but typical score interpretation does not currently allow for accurate distinctions to be made between them. Every intelligence test score represents the influence of these abilities as well as measurement error and should be interpreted as such. In light of the evidence of invariance across almost all of the latent variables we targeted across tests, we suggest that factor scores representing psychometric *g* and more specific abilities should be developed to accomplish the elusive goal of measuring cognitive abilities at varying strata in a relatively pure manner (Carroll, 1993; Gustafsson & Snow, 1997).

References

American Association on Intellectual and Developmental Disabilities. (2010). *Mental retardation: Definition, classification, and systems of supports* (11th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48,* 5–37.

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

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

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16–29.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis–Kaplan Executive Function System.* San Antonio, TX: Pearson.

Elliott, C. (1990). *Differential Ability Scales.* San Antonio, TX: Psychological Corporation.

Elliott, C. (2007). *Differential Ability Scales, Second Edition.* San Antonio, TX: Psychological Corporation.

Floyd, R. G., Bergeron, R., Hamilton, G., & Parra, G. (2010). How do executive functions fit with the Cattell–Horn–Carroll model? Some evidence from a joint factor analysis of the Delis–Kaplan Executive Functions System and The Woodcock–Johnson III Tests of Cognitive Abilities. *Psychology in the Schools, 47,* 721–738.

Floyd, R. G., Bergeron, R., McCormack, A. C., Anderson, J. L., & Hargrove-Owens, G. L. (2005). Are Cattell–Horn–Carroll (CHC) broad ability composite scores exchangeable across batteries? *School Psychology Review, 34,* 386–414.

Floyd, R. G., Clark, M. H., & Shadish, W. R. (2008). The exchangeability of IQs: Implications for professional psychology. *Professional Psychology: Research and Practice, 39,* 414–423.

Floyd, R. G., Shands, E. I., Rafael, F. A., Bergeron, R., & McGrew, K. S. (2009). The dependability of general-factor loadings: The effects of factor-extraction methods, test battery composition, test battery size, and their interactions. *Intelligence, 37,* 453–465.

Gottfredson, L. S. (2008). Of what value is intelligence? In A. Prifitera, D. Saklofske, & L. G. Weiss (Eds.), *WISC-IV applications for clinical assessment and intervention* (2nd ed., pp. 545–563). Amsterdam, The Netherlands: Elsevier.

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* (Vol. 4, pp. 35–71). Hillsdale, NJ: Erlbaum.

Gustafsson, J. E., & Snow, R. E. (1997). Ability profiles. In R. F. Dillon (Ed.), *Handbook on testing* (pp. 107–135). Westport, CT: Greenwood Press.

Individuals with Disabilities in Education Improvement Act, Pub. L. No. 108-446 (2004).

Jensen, A. (1998). *The* g *factor: The science of mental ability.* Westport, CT: Praeger.

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

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

Kamphaus, R. K. (2009). Assessment of intelligence and achievement. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (4th ed., pp. 230–246). New York, NY: Wiley.

Kaufman, A. S., & Kaufman, N. L. (2004a). *Kaufman Assessment Battery for Children, Second Edition.* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman Test of Educational Achievement, Second Edition.* Circle Pines, NM: American Guidance Service.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive *g* and academic achievement *g* one and the same *g*? An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence, 40,* 72–238.

Keith, T. Z., Fine, J. G., Reynolds, M. R., Taub, G. E., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure? *School Psychology Review, 35,* 108–127.

Keith, T. Z., Kranzler, J. H., & Flanagan, D. P. (2001). What does the Cognitive Assessment System (CAS) measure? Joint confirmatory factor analysis of the CAS and the Woodcock–Johnson Tests of Cognitive Ability (3rd Edition). *School Psychology Review, 30,* 89–119.

Keith, T. Z., & Reynolds, M. R. (2010). CHC theory and cognitive abilities: What we've learned from 20 years of research. *Psychology in the Schools, 47,* 635–650.

Major, J. T., Johnson, W., & Bouchard, T. J., Jr. (2011). The dependability of the general factor of intelligence: Why small single-factor models do not adequately represent *g*. *Intelligence, 39,* 418–433.

Matzke, D., Dolan, C. V., & Molenaar, D. (2010). The issue of power in the identification of "*g*" with lower-order factors. *Intelligence, 38,* 336–344.

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock–Johnson III technical manual.* Itasca, IL: Riverside Publishing.

McLain, M., & Pfeiffer, S. (2012). Identification of gifted students in the United States today: A look at state definitions, policies, and practices. *Journal of Applied School Psychology, 28,* 59–88.

Mulaik, S. A. (2010). *Foundations of factor analysis.* Boca Raton, FL: Chapman & Hall/CRC.

Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide, sixth edition.* Los Angeles, CA: Muthén & Muthén.

Naglieri, J. A., & Das, J. P. (1997). *Das Naglieri Cognitive Assessment System.* Itasca, IL: Riverside Publishing.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. A. (2005). The general (*g*), broad, and narrow CHC stratum characteristics of the WJ III and WISC-III tests: A confirmatory cross-battery investigation. *School Psychology Quarterly, 20,* 66–88.

Reeve, C. L., & Blacksmith, N. (2009). Equivalency and reliability of vectors of *g*-loadings across different methods of estimation and sample sizes. *Personality and Individual Differences, 47,* 968–972.

Reynolds, M. R., & Keith, T. Z. (2007). A test of Spearman's law of diminishing returns in hierarchical models of intelligence. *Intelligence, 35,* 267–281.

Sanders, S., McIntosh, D. A., Dunham, M., Rothlisberg, B. A., & Finch, H. (2007). Joint confirmatory factor analysis of the Differential Ability Scales and the Woodcock–Johnson Tests of Cognitive Abilities–Third Edition. *Psychology in the Schools, 44,* 119–138.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8,* 23–74.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment* (3rd ed., pp. 99–144). New York, NY: Guilford Press.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York, NY: MacMillan.

Thomson, G. H. (1951). *The factorial analysis of human ability* (5th ed.). London, UK: University of London Press.

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

Thurstone, L. L. (1947). *Multiple-factor analysis.* Chicago, IL: University of Chicago Press.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, 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.

Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children—Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *The Wechsler Intelligence Scale for Children—Fourth Edition.* San Antonio, TX: Psychological Corporation.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Cognitive Abilities.* Itasca, IL: Riverside Publishing.

Date Received: September 21, 2012
Date Accepted: September 29, 2013
Associate Editor: Robert Volpe ■

Randy G. Floyd is an Associate Professor of Psychology at The University of Memphis. He received his doctoral degree in school psychology from Indiana State University. His research interests include the structure, measurement, and correlates of cognitive abilities; the technical properties of early numeracy measures; and the process of professional publication.

Matthew R. Reynolds is an Assistant Professor in the Department of Psychology and Research in Education (School Psychology) at The University of Kansas. He received his PhD from the University of Texas at Austin. His research interests include the development, structure, and measurement of human cognitive abilities.

Ryan L. Farmer is a doctoral candidate in the School Psychology program at The University of Memphis. His research interests include psychometric properties of cognitive ability tests, violence risk assessment, and intervention efficacy.

John H. Kranzler is Professor and Director of the School Psychology Program at the University of Florida. He received his doctoral degree in school psychology from the University of California at Berkeley in 1990. His primary research interests include the nature, development, and assessment of intelligence.