Psychological Assessment
2015, Vol. 27, No. 2, 689–697

1040-3590/15/$12.00 http://dx.doi.org/10.1037/pas0000052

Structure of the Woodcock–Johnson III Cognitive Tests in a Referral Sample of Elementary School Students

Tracy Strickland
Arizona State University

Marley W. Watkins
Baylor University

Linda C. Caterino
Arizona State University

The structure of the Woodcock–Johnson Cognitive Battery-Third Edition (WJ Cog) has been extensively explored via confirmatory factor analysis (CFA) with its normative sample, but there has been little research to verify that the same structure holds for students referred for special education services. Likewise, research on the structure of the WJ Cog with exploratory factor analysis (EFA) methods has been rare. Consequently, this study applied both EFA and CFA methods to the scores of 529 elementary school students referred for special education services (95.5% eligible) on the 14 tests of the WJ Cog extended battery. EFA results suggested only 2 or 3 factors, whereas CFA results favored the theoretical 7 factors posited by McGrew and Woodcock. In this theoretical model, a strong general factor accounted for 27% of the total variance and 57% of the common variance, whereas the 7 group factors combined accounted for 21% of the total variance and 43% of the common variance. Reliability, as quantified by ω_H, was good for the general factor, marginal for the *G*s factor, and poor for the other group factors. Nine of the 14 WJ Cog tests displayed uniqueness values that exceeded their communality. On the basis of this evidence from a referral sample, interpretation of the WJ Cog should be restricted to the *G*s and *g* factors: the *G*s factor because it exhibited considerable independence and precision of measurement and the *g* factor because it has emerged in all investigations of the WJ Cog.

Keywords: intelligence, factor analysis, validity, Woodcock–Johnson Cog

According to the Individuals with Disabilities Education Act of 2004, all students being considered for special education must receive a comprehensive individual evaluation to determine their eligibility for special education services. Assessment of cognitive ability with individual intelligence tests is included in many of those mandated special education eligibility evaluations (Braden, 2013). One individual intelligence test battery that is popular with clinicians (Braden, 2013) and often taught in assessment courses (Braden & Alfonso, 2003) is the Woodcock–Johnson Cognitive Battery-Third Edition (WJ Cog; Woodcock, McGrew, & Mather, 2001).

Because of its use in high-stakes special education eligibility decisions, it is vital that strong evidence support the reliability and validity of the WJ Cog. To this end, McGrew and Woodcock (2001) presented considerable evidence of reliability and validity. Content validity evidence, convergent and discriminant validity evidence, and evidence based on relations to other variables have also been published (Braden & Alfonso, 2003; Floyd, Shaver, & McGrew, 2003; McGrew & Wendling, 2010; Sattler, 2001; Schrank, 2011).

McGrew and Woodcock (2001) also provided evidence based on internal structure, or structural validity, which indicates the degree to which a test's structure aligns with the constructs on which it was based (Messick, 1995; Wasserman & Bracken, 2013). The theoretical foundation of the WJ Cog is the Cattell–Horn–Carroll (CHC) theory of cognitive abilities (McGrew, 2009), which posits a hierarchical structure of intelligence with three levels: (a) an overarching general intelligence (*g*) at strata III, (b) 10 broad abilities at strata II, and (c) more than 70 narrow abilities at strata I. The WJ Cog is designed to measure seven of the 10 broad CHC abilities (Schrank, 2011). Extensive confirmatory factor analyses (CFAs) have been conducted with the WJ Cog normative sample and have generally aligned with the theoretical model (Floyd, McGrew, Barry, Rafael, & Rogers, 2009; Keith & Reynolds, 2012; Keith, Reynolds, Patel, & Ridley, 2008; Locke, McGrew, & Ford, 2011; Phelps, McGrew, Knopik, & Ford, 2005; Taub & McGrew, 2004).

Although the structure of the WJ Cog has been extensively explored with its normative sample, there has been little research to verify that the same structure holds for clinical samples. Al-

This article was published Online First February 2, 2015.

Tracy Strickland, Department of Educational Psychology, Arizona State University; Marley W. Watkins, Department of Educational Psychology, Baylor University; Linda C. Caterino, Division of Educational Leadership and Innovation, Arizona State University.

Tracy Strickland is now with the Joliet Public School District, Joliet, Illinois.

This study is based on Tracy Strickland's dissertation. The contributions of Linda C. Caterino and Samuel Green are gratefully acknowledged. A preliminary version was reported at the 2013 annual meeting of the International Society for Intelligence Research, Melbourne, Australia.

Correspondence concerning this article should be addressed to Marley W. Watkins, Department of Educational Psychology, Baylor University, Waco, TX 76798-7301. E-mail: Marley_Watkins@baylor.edu

though the WJ Cog Technical Manual (McGrew & Woodcock, 2001) provided extensive information about factor analysis for special study samples (preschool, attention deficit hyperactivity disorder, learning disabled, and across age groups) that supported the theoretical structure, those studies included additional cognitive tests and may not reflect test administration as actually performed by practitioners (Taub & McGrew, 2004). However, a recent CFA with a large sample of children with learning disorders, as defined by the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; American Psychiatric Association, 2000), found general support for the theoretical model (Benson & Taub, 2013). Unfortunately, data for this analysis were extracted from archival records provided to the Woodcock–Muñoz Foundation by an unknown number of clinicians from unknown regions of the country and were reviewed by unknown judges using unknown criteria of fitness for inclusion. Thus, "the sample examined in this study may differ from the population of students identified in a school setting" (Benson & Taub, 2013, p. 259).

Historically, structural validity has been studied with exploratory factor analysis (EFA). For example, EFA was used by Carroll (1993) to create the structural model that was subsequently merged with Horn's (1968) *Gf–Gc* theory (1968) to create the CHC theoretical model for the WJ Cog. Carroll (1995) argued that "results should be shown on the basis of orthogonal factors, rather than oblique, correlated factors. I insist, however, that the orthogonal factors should be those produced by the Schmid–Leiman, 1957, orthogonalization procedure" (p. 437) for its ease of interpretation and parsimony.

More recently, unrestricted EFA methods have been eclipsed by restricted CFA methods in the study of structural validity (Keith & Reynolds, 2012). For instance, all extant analyses of the 14 WJ Cog tests have used CFA methods (Benson & Taub, 2013; Keith et al., 2008; McGrew & Woodcock, 2001; Phelps et al., 2005; Sanders, McIntosh, Dunham, Rothlisberg, & Finch, 2007; Taub & McGrew, 2004). Frazier and Youngstrom (2007) contended that the use of CFA has resulted in overfactoring of cognitive ability tests, including the WJ Cog. That hypothesis was supported by Dombrowski (2013), who found only three or four factors, instead of the theoretically posited seven, when he applied EFA methods to the correlation matrix from the 20 WJ Cog test scores of school-aged member of the WJ Cog normative sample. Expressing concern about differences between the results from CFA and EFA, Carroll (1995) recommended that "a confirmatory analysis of a dataset should not be published without an accompanying statement or report on one or more appropriate exploratory analyses" (p. 437).

Establishing the validity of the structure of the WJ Cog in a school referral sample can provide further support for its use in samples that may show different patterns of performance (Strauss, Sherman, & Spreen, 2006) that may not be reflected in the WJ Cog normative sample. In addition, replication of a model in multiple samples by independent researchers using both EFA and CFA methods may enhance validity (Raykov & Marcoulides, 2000). If the structure cannot be generalized, this indicates a lack of structural validity evidence for the test's use in a clinical sample and may lead to inappropriate interpretation of an individual's test scores and inappropriate actions being taken on the basis of those scores (Messick, 1995). Consequently, this study examined the extent to which the WJ-III Cog structure established in the normative sample was replicated in a sample of students referred for special education services.

Method

Participants

The 529 participants in this study were 6–13 years of age (M = 9.47, SD = 1.81) and predominately male (62%). Ethnic/racial background of the participants was reported to be 48.6% White, 32.3% Hispanic, 8.3% Black, 4.7% Native American, 0.6% Asian/Pacific Islander, 1.1% other, 2.8% multiracial, and 1.5% not reported. Special education eligibility of the participants included 68.8% with a learning disability, 11.3% with a speech–language impairment, 9.6% with an other health impairment, 3.2% with an emotional disability, 0.9% with an intellectual disability, 0.6% with multiple disabilities, 0.8% with autism, 0.2% with a hearing impairment, 0.2% with an orthopedic impairment, and 4.5% who were not disabled and therefore not eligible for special education services. English was the home language of 85% of the participants, and Spanish was the home language of 12% of the participants.

Internal review board and school district confidentiality and procedural policies did not allow additional identifying information to be collected on student participants. However, the ethnic/racial background of the students in this suburban southwestern U.S. school district was 38% White, 48% Hispanic, 7% Black, 4% Native American, and 3% Asian/Pacific Islander. Approximately 64% of district students were eligible for free or reduced-price lunch, and around 24% were English language learners.

Instruments

The WJ Cog battery is a test of intellectual ability that is designed to be individually administered to individuals between 24 months and 90-plus years of age (Woodcock et al., 2001). The norming sample was selected using stratified random sampling and was representative of the U.S. population ages 24 months to 90 years and older. In total, the WJ Cog battery contains 14 tests that make up the General Intellectual Ability (GIA)-Standard Battery and the GIA-Extended Battery, plus six supplemental tests that can provide additional information on an individual's cognitive strengths and weaknesses. The GIA-Standard scale consists of seven tests, one for each of the broad CHC abilities. The GIA-Extended scale comprises 14 tests, the seven tests from the GIA-Standard scale plus seven additional tests, resulting in two tests for each of the seven broad CHC abilities. An additional six tests can be used to supplement the standard or extended scales. A list of the 14 tests used in this study that measured the seven CHC factors in the standard and extended batteries is provided in Table 1.

Procedures

The sample for this study was extracted from a database of 1,954 students who received psychoeducational evaluations from 2001 through 2007 in a southwestern U.S. suburban school district that served students in grades K–8. Inclusion in the study was contingent on the special education record including a score for

Table 1
Woodcock–Johnson Cognitive Battery-Third Edition Cognitive General Intellectual Ability (GIA)-Extended Factor Structure

Test	Stratum II Broad Factor	Stratum III
Concept Formation Analysis–Synthesis	Fluid Reasoning (*Gf*)	General Intelligence (*g*)
Verbal Comprehension General Information	Comprehension–Knowledge (*Gc*)	
Spatial Relations Picture Recognition	Visual–Spatial Thinking (*Gv*)	
Visual Matching Decision Speed	Processing Speed (*Gs*)	
Visual Auditory Learning Retrieval Fluency	Long-Term Retrieval (*Glr*)	
Sound Blending Auditory Attention	Auditory Processing (*Ga*)	
Numbers Reversed Memory for Words	Short-Term Memory (*Gsm*)	

each of the 14 WJ Cog tests included in the GIA-Extended battery, a reported age of 6–13 years, and testing conducted in English. A total of 529 cases met these criteria. All assessments were conducted by certified school psychologists employed in the district.

Analyses

EFA. The EFA followed best practice guidelines (Fabrigar, Wegener, MacCullum, & Strahan, 1999; Hoelzle & Meyer, 2013; Kline, 2013; Osborne, Costello, & Kellow, 2007) and was implemented with the *psych* package in the R programming language (R Development Core Team, 2013). Bartlett's (1950) test of sphericity was used to ensure that the correlation matrix was not random, and the Kaiser–Meyer–Olkin statistic was required to be above a minimum standard of .6 (Kaiser, 1974). The principal axis extraction method was used because of its relative tolerance of nonnormality and demonstrated ability to recover weak factors (Briggs & MacCallum, 2003). Squared multiple correlations were used to estimate the initial communalities (Gorsuch, 2003). As suggested by Velicer, Eaton, and Fava (2000), the procedures used to determine the appropriate number of factors for retention and rotation included parallel analysis (Horn, 1965) and minimum average partials (MAPs; Velicer, 1976), supplemented by a visual scree test (Cattell, 1966). Parsimony and theoretical convergence were also considered. Because of the nature of the construct, it was assumed that the factors would be correlated. Therefore, a Promax rotation (Hendrickson & White, 1964) with a *k* value of 4 was used (Tataryn, Wood, & Gorsuch, 1999). Criteria for determining factor adequacy were established a priori. Pattern coefficients greater than or equal to .30 were considered salient. Complex loadings that were salient on more than one factor were rejected in the interest of parsimony and to honor simple structure (Thurstone, 1947). The oblique first-order factors were orthogonalized with the Schmid–Leiman procedure (Schmid & Leiman, 1957), as recommended by Carroll (1995).

CFA. Mplus 7 for the Macintosh (Muthén & Muthén, 2013) was used to conduct CFAs using maximum likelihood estimation, with Satorra and Bentler (1994) scaling to correct for multivariate nonnormality. The optimal EFA solution was compared with the seven-factor theoretical model (see Table 1). Multiple fit indices that represented a variety of criteria were examined (Kline, 2013): (a) the chi-square, (b) the comparative fit index (CFI), (c) the root mean square error of approximation (RMSEA), (d) the standardized root mean square residual (SRMR), and (e) Akaike's information criterion (AIC). Although there are no universally accepted cutoff values for model fit indices, the guidelines for adequate model fit were as follows: (a) CFI $\geq$ .90 and (b) SRMR and RMSEA $\leq$ .08. The guideline for good model fit was (a) CFI $\geq$ .95 and (b) SRMR and RMSEA $\leq$ .06 (Hu & Bentler, 1999). There are no specific criteria for information-based fit indices like the AIC, but smaller values may indicate better approximations of the true model (Vrieze, 2012).

Results

Descriptive statistics for the WJ Cog tests are reported in Table 2. Score distributions were relatively normal, with −1.37 the largest univariate skew and 5.61 the largest univariate kurtosis

Table 2
Descriptive Statistics for the Woodcock–Johnson Tests of Cognitive Ability-Third Edition Test Scores of 529 Elementary School Students Tested for Special Education Eligibility

Test	*M*	*SD*	Skewness	Kurtosis
Concept Formation	94.54	12.98	−0.68	1.45
Analysis-Synthesis	97.17	12.32	−0.17	0.65
Verbal Comprehension	89.73	12.51	−0.04	0.40
General Information	88.59	13.01	−0.41	0.84
Spatial Relations	96.75	9.97	−0.73	4.00
Picture Recognition	100.90	10.50	−1.37	5.61
Visual Matching	86.29	14.15	−0.51	0.94
Decision Speed	96.43	14.48	−0.23	0.29
Visual Auditory Learning	84.57	14.98	−0.24	-0.34
Retrieval Fluency	86.80	13.83	−0.58	0.98
Sound Blending	101.10	11.86	−0.08	0.82
Auditory Attention	97.65	12.31	−0.69	1.47
Numbers Reversed	89.27	12.74	−0.16	0.37
Memory for Words	90.17	13.10	0.07	0.63
General Intellectual Ability	88.54	10.86	−0.53	1.18

(Fabrigar et al., 1999). However, multivariate kurtosis was large (23.38) and indicative of nonnormality (Bentler, 2005).

EFA

The results of Bartlett's test of sphericity indicated that the correlation matrix was not random, $\chi^2(91, N = 529) = 2{,}273.8$, $p < .001$, and the Kaiser–Meyer–Olkin statistic was .83, well above the minimum standard for conducting factor analysis (Kaiser, 1974). Therefore, the correlation matrix was appropriate for factor analysis. Parallel analysis suggested that two factors should be retained, MAP suggested one factor, and visual inspection of the scree plot indicated that two factors were sufficient. However, parallel analysis of data with a strong general factor, as anticipated with these data, has been found to underfactor (Crawford et al., 2010). Therefore, both the two-factor and three-factor solutions and their orthogonalizations were examined.

In the oblique two-factor solution, the Fluid Reasoning (*Gf*), Comprehension–Knowledge (*Gc*), Visual–Spatial Thinking (*Gv*), Long-Term Retrieval (*Glr*), Auditory Processing (*Ga*), and Short-Term Memory (*Gsm*) tests combined to create one factor, whereas the two Processing Speed (*Gs*) tests loaded on the second factor. In the oblique three-factor solution, the *Gc* and *Gs* factors were distinct, but the other tests combined to form one factor. Orthogonalization of the three-factor oblique solution (see Table 3) revealed that all 14 tests were *g*-loaded, with the Concept Formation test exhibiting the strongest loading (.73) and the Decision Speed test the weakest loading (.20). The *Gs*, *Gc*, and *Gsm* factors emerged after taking *g* into account. In contrast, there was little evidence of the *Gf*, *Gv*, *Glr*, or *Ga* factors beyond *g*.

When the seven-factor theoretical model was forced, (a) all 14 tests loaded on the general factor, (b) one group factor contained the four *Gf* and *Gv* tests, (c) a second group factor was formed by the two *Gc* tests, (d) a third group factor consisted of the two *Gs* tests, (e) a fourth group factor contained the two *Gsm* tests, and (f) the *Glr* and *Ga* tests did not cohere. Thus, the *Gc*, *Gs*, and *Gsm* factors emerged, but EFA was unable to identify a model that fully aligned with the WJ Cog theoretical structure.

CFA

The oblique two- and three-factor EFA models were compared with an oblique seven-factor model and a higher order model (Keith & Reynolds, 2012) with seven first-order factors. The two- and three-factor models suggested by EFA were poor fits (see Table 4), whereas the oblique seven-factor model was a good fit, and its equivalent higher order version, illustrated in Figure 1, demonstrated adequate fit to the data.

An orthogonal version of the theoretical seven-factor model was also examined (see Table 5). Given the underidentification of the WJ Cog factors, test loadings within each factor were restrained to equality. This bifactor model (Reise, 2012), which has also been called a *direct hierarchical model* (Gignac, 2008) and a *nested factor model* (Gustafsson & Balke, 1993), has been recommended for hierarchically structured constructs such as intelligence (Brunner, Nagy, & Wilhelm, 2012), has been applied to other cognitive scales (Gignac & Watkins, 2013; Golay & Lecerf, 2011; Watkins, 2010; Watkins, Canivez, James, James, & Good, 2013), and was explicitly suggested by Carroll (1998). In this orthogonal model, the general factor accounted for more total and common variance (27% and 57%, respectively) than all group factors combined (21% and 43%, respectively).

As displayed in Table 5, nine of the 14 WJ Cog tests displayed uniqueness values (i.e., error variance or variance unique to that test alone) that exceeded their communality (i.e., variance contributed by the common factors). The *Gc*, *Gs*, and *Gf* tests retained little specific variance (i.e., specificity or variance unique to that test alone). Thus, these six tests were measuring relatively large amounts of reliable variance from general and group factors. Comparatively, the *Gs* construct exerted greater influence than the *g*

Table 3
Exploratory Factor Analysis With Oblique and Orthogonalized Pattern Coefficients of the Woodcock–Johnson Tests of Cognitive Ability-Third Edition Test Scores of 529 Elementary School Students Tested for Special Education Eligibility

	Oblique solution			Orthogonalized solution					
Test	I	II	III	*g*	I	II	III	h^2	u^2
Concept Formation	**.715**	.041	−.013	**.731**	.120	.013	−.118	.563	.437
Analysis-Synthesis	**.717**	−.026	.015	**.693**	.152	−.061	−.140	.527	.473
Verbal Comprehension	−.034	**.896**	.015	**.613**	−.028	**.591**	.093	.734	.266
General Information	−.011	**.825**	.007	**.589**	−.030	**.617**	−.073	.733	.267
Spatial Relations	**.710**	−.188	.023	**.557**	.179	−.166	.054	.372	.628
Picture Recognition	**.670**	−.119	.024	**.572**	.163	−.110	−.100	.377	.623
Visual Matching	.145	−.040	**.625**	**.249**	**.712**	−.025	.032	.571	.429
Decision Speed	−.054	.029	**.908**	**.203**	**.776**	−.008	−.021	.644	.356
Visual Auditory Learning	**.468**	.220	−.061	**.611**	.015	.134	−.170	.421	.579
Retrieval Fluency	.130	.220	.164	**.325**	.173	.137	.047	.157	.843
Sound Blending	**.400**	.127	−.042	**.475**	.026	.045	.189	.264	.736
Auditory Attention	**.509**	−.001	−.024	**.494**	.073	−.026	−.088	.258	.742
Numbers Reversed	.292	.097	−.022	**.353**	.044	.027	**.375**	.269	.731
Memory for Words	**.450**	.101	−.081	**.515**	−.007	.002	**.397**	.423	.577
Total variance (%)				27.3	8.9	5.8	3.1	45.1	54.9
Common variance (%)				60.5	19.7	12.9	6.9		

Note. Salient loadings ≥.30 for the oblique solution and ≥.20 for the orthogonalized solution (Dombrowski, 2013) are represented in bold. h^2 = communality; u^2 = uniqueness.

Table 4
Fit Statistics for Competing Structural Models of the Woodcock–Johnson Tests of Cognitive Ability-Third Edition Test Scores of 529 Elementary School Students Tested for Special Education Eligibility

Model	χ^2	*df*	CFI	SRMR	RMSEA	90% RMSEA	AIC
Oblique two factors	468.2	76	.794	.062	.099	.090–.107	56,947
Oblique three factors	292.9	74	.890	.056	.073	.064–.082	56,746
Oblique seven factors	156.4	56	.947	.035	.058	.048–.069	56,639
Higher order seven factors	208.8	70	.927	.046	.061	.052–.071	56,667

Note. CFI = comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; AIC = Akaike's information criterion.

construct on the two *Gs* tests (.50 vs. .10, respectively), and the *g* construct exerted greater influence than the *Gf* construct on the two *Gf* tests (.48 vs. .16, respectively), whereas the *g* and *Gc* constructs had roughly equal influence on the two *Gc* tests (.34 vs. .38, respectively). In contrast, the Retrieval Fluency, Sound Blending, Auditory Attention, and Numbers Reversed tests exhibited specificity values greater than their communality values. These four tests were measuring relatively small amounts of reliable common variance (.18–.29), and their constituent factor scores (*Glr*, *Ga*, and *Gsm*) were correspondingly contaminated with variance specific to each test.

Traditional reliability estimates such as coefficient alpha can be misleading when test scores are influenced by more than one source of variance (Yang & Green, 2011). In these situations, model-based reliability indices such as omega (ω) and omega hierarchical (ω_H) have been recommended (Zinbarg, Revelle, Yovel, & Li, 2005). Omega estimates the reliable variance accounted for by all constructs (i.e., general *and* group factors) influencing indicators, whereas ω_H estimates the variance accounted for by a single target construct (i.e., either general *or* group factor). Although subjective, values for ω and ω_H should exceed .50, with values of .75 or above preferred (Reise, Bonifay, & Haviland, 2013). When multiple sources of variance were considered via ω, the *g*, *Gc*, *Gf*, and *Gs* factors exceeded the .75 preferred guideline. When a single source of variance was considered via ω_H, only the *g* factor was above .75, and only the *Gs* factor exceeded the minimum guideline of .50. Omega hierarchical estimates for the other group factors ranged from .06 (*Ga*) to .45 (*Gc*). These values indicate that these WJ Cog tests, with the exception of the *Gs* tests, contain little reliable variance that is unique from the general factor. Similar results have been found for other tests of intelligence (Brunner et al., 2012; Gignac & Watkins, 2013; Reynolds, Floyd, & Niileksela, 2013; Watkins et al., 2013).

Discussion

The structure of the WJ Cog has been extensively explored via CFA with its normative sample, but there has been little research to verify that the same structure holds for students referred for special education services. Likewise, research on the structure of the WJ Cog with EFA methods has been rare. Consequently, this study applied both EFA and CFA methods to the scores of 529 elementary school students referred for special education services on the 14 tests of the WJ Cog extended battery. EFA results suggested only two or three factors, whereas CFA results favored the theoretical seven factors posited by McGrew and Woodcock (2001). In this theoretical model, a strong general factor accounted for 27% of the total variance and 57% of the common variance, whereas all seven group factors combined accounted for 21% of the total variance and 43% of the common variance. This result

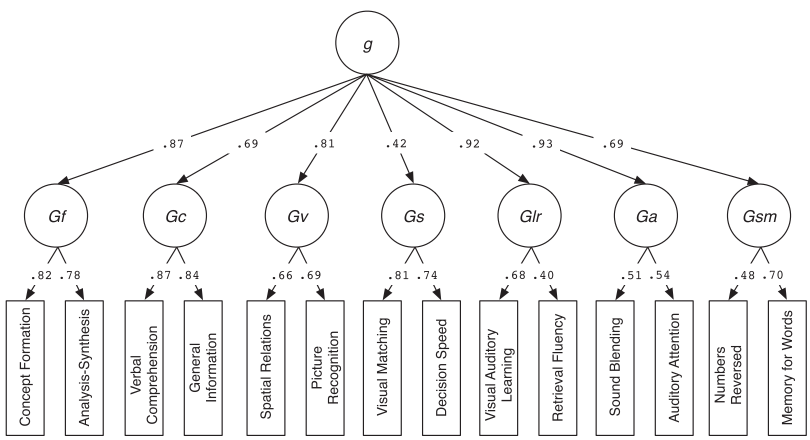

Figure 1. Theoretical higher order structure of the Woodcock–Johnson Cognitive Extended battery among 529 students referred for special education services. *Gf* = Fluid Reasoning; *Gc* = Comprehension–Knowledge; *Gv* = Visual–Spatial Thinking; *Gs* = Processing Speed; *Glr* = Long-Term Retrieval; *Ga* = Auditory Processing; *Gsm* = Short-Term Memory; *g* = General Intelligence.

Table 5
Standardized Coefficients for Bifactor Version of the Theoretical Structure of the Woodcock–Johnson Tests of Cognitive Ability-Third Edition Test Scores of 529 Elementary School Students Tested for Special Education Eligibility

	Factor											
Test	*g*	*Gc*	*Glr*	*Gv*	*Ga*	*Gf*	*Gs*	*Gsm*	h^2	u^2	Error	s^2
Concept Formation	.713					.389			.660	.340	.055	.285
Analysis–Synthesis	.671					.410			.618	.382	.115	.267
Verbal Comprehension	.595	.632							.753	.247	.105	.142
General Information	.574	.608							.699	.301	.130	.171
Spatial Relations	.536			.402					.449	.551	.190	.361
Picture Recognition	.561			.382					.461	.539	.305	.234
Visual Matching	.343						.712		.625	.375	.125	.250
Decision Speed	.312						.696		.582	.418	.135	.283
Visual Auditory Learning	.626		.202						.433	.567	.150	.417
Retrieval Fluency	.369		.218						.184	.816	.190	.626
Sound Blending	.475				.191*				.262	.738	.165	.573
Auditory Attention	.508				.184*				.292	.708	.115	.593
Numbers Reversed	.334							.425	.292	.708	.140	.568
Memory for Words	.481							.413	.402	.598	.255	.343
Total Variance (%)	27.2	5.5	0.6	2.2	0.5	2.3	7.1	2.5	47.9	52.1	15.5	36.5
Common Variance (%)	56.8	11.5	1.3	4.6	1.0	4.8	14.8	5.2				
ω	.885	.841	.457	.625	.434	.780	.752	.511				
ω_H	.794	.445	.069	.211	.055	.195	.619	.263				

Note. *g* = General Intelligence; *Gc* = Comprehension–Knowledge; *Glr* = Long-Term Retrieval; *Gv* = Visual–Spatial Thinking; *Ga* = Auditory Processing; *Gf* = Fluid Reasoning; *Gs* = Processing Speed; *Gsm* = Short-Term Memory; h^2 = communality; u^2 = uniqueness; error = 1 − reliability for ages 9–10 years from McGrew, Schrank, and Woodcock (2007); $s^2 = u^2 -$ error.
* $p > .01$.

replicates the important role of the general factor found in previous analyses of the WJ Cog (e.g., Dombrowski, 2013; Floyd et al., 2009). Nine of the 14 WJ Cog tests displayed uniqueness values that exceeded their communality, indicating that much of the variability in these tests comprises test-specific and error variance. Likewise, omega coefficients indicated that only the general factor was measured with high precision. These results indicate that the WJ Cog tests contain little reliable variance that is unique from the general factor.

Contradictory EFA and CFA results have also been reported for the WJ Cog standardization sample (Dombrowski, 2013, 2014a, 2014b; Dombrowski & Watkins, 2013) and in joint analyses of the WJ Cog with other instruments (Chang, Paulson, Finch, McIntosh, & Rothlisberg, 2014; Floyd, Bergeron, Hamilton, & Parra, 2010). For example, EFA analyses by Dombrowski (2013, 2014b) found that tests that measured the *Gc*, *Gs*, and *Glr* factors aligned with the WJ Cog theoretical structure. The current EFA study with a referral sample found that tests that measured the *Gc*, *Gs*, and *Gsm* factors aligned with the WJ Cog theoretical structure. Thus, only the *Gc*, *Gs*, and a memory factor (either *Gsm* or *Glr*) have consistently emerged in EFA studies.

CFA results have sometimes failed to support the theoretical structure of the WJ Cog (Chang et al., 2014; Floyd et al., 2010). For example, a joint CFA with the Differential Ability Scales (Elliott, 1990) found problems with the *Gv* tests and redundant *g* and *Gf* factors (Sanders et al., 2007). In addition, joint CFAs with the Stanford-Binet scale (Roid, 2003) and the Kaufman Assessment Battery for Children-Second Edition (Kaufman & Kaufman, 2004) among preschool children failed to isolate the *Gf* factor (Chang et al., 2014; Hunt, 2008).

As noted by Gorsuch (2003), "the ultimate arbitrator in science is well established: replication" (p. 153). It was expected that CFA would support the structure identified by EFA. This expectation was not substantiated. Rather, CFA supported the theoretical seven-factor structure for the WJ Cog. It is possible that the WJ Cog has been overfactored, as suggested by Frazier and Youngstrom (2007). Alternatively, neurodevelopmental processes may make it possible to detect fewer dimensions among children (Mungas et al., 2013). It is also possible that cognitive theory needs more study (as recognized by Carroll, 1995). Finally, the power to detect converging EFA and CFA solutions may have been inadequate. Additional studies with diverse populations, multiple methods, and larger samples will be needed to resolve this dilemma.

On the basis of this evidence from a referral sample of elementary school students, interpretation of the WJ Cog should be restricted to the *Gs* and *g* factors: the *Gs* factor because it exhibited considerable independence and precision of measurement and the *g* factor because it has emerged in all investigations of the WJ Cog as well as in investigations of other intelligence batteries (e.g., Floyd, Reynolds, Farmer, & Kranzler, 2013; Reynolds et al., 2013). Several of the WJ Cog factors (e.g., *Ga*, *Glr*) accounted for so little common variance that they might be termed "factors of no importance" (Kelley, 1939, p. 140). In additionally, any interpretation of WJ Cog scores should take into consideration the fact that each score is influenced by general, group, and specific factors as well as error and is not a pure measure of any single trait (DeMars, 2013). For example, interpreting the Concept Formation score as a measure of *Gf* ignores the contribution of general intelligence (approximately 51% of the *Gf* score variance) and uniqueness (around 12% of the *Gf* score variance). Likewise, interpreting the General Information score as a measure of *Gc* ignores the contribution of general intelligence (approximately 34% of the *Gc* score variance) and uniqueness (around 8% of the *Gc* score variance).

Absent specific tools that allow conversion of latent construct scores (Schneider, 2013), identification of cognitive strengths and weaknesses based on tests or lower order factors may be inaccurate (Floyd et al., 2009) and should be eschewed.

References

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

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology, 3,* 77–85.

Benson, N., & Taub, G. E. (2013). Invariance of Woodcock–Johnson III scores for students with learning disorders and students without learning disorders. *School Psychology Quarterly, 28,* 256–272. http://dx.doi.org/10.1037/spq0000028

Bentler, P. M. (2005). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.

Braden, J. P. (2013). Psychological assessment in school settings. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Vol. 10. Assessment psychology* (2nd ed., pp. 291–314). Hoboken, NJ: Wiley.

Braden, J. P., & Alfonso, V. C. (2003). Woodcock-Johnson III Tests of Cognitive Abilities in cognitive assessment courses. In F. A. Schrank & D. P. Flanagan (Eds.), *WJ III clinical use and interpretation: Scientist–practitioner perspectives* (pp. 377–401). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/B978-012628982-4/50013-1

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research, 38,* 25–56. http://dx.doi.org/10.1207/S15327906MBR3801_2

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80,* 796–846. http://dx.doi.org/10.1111/j.1467-6494.2011.00749.x

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytical studies*. New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511571312

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research, 30,* 429–452. http://dx.doi.org/10.1207/s15327906mbr3003_6

Carroll, J. B. (1998). Human cognitive abilities: A critique. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 5–23). Mahwah, NJ: Erlbaum.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245–276. http://dx.doi.org/10.1207/s15327906mbr0102_10

Chang, M., Paulson, S. E., Finch, W. H., McIntosh, D. E., & Rothlisberg, B. A. (2014). Joint confirmatory factor analysis of the Woodcock-Johnson Test of Cognitive Abilities, Third Edition, and the Stanford-Binet Intelligence Scales, Fifth Edition, with a preschool population. *Psychology in the Schools, 51,* 32–57. doi:10.1002/pits.21734

Crawford, A. V., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement, 70,* 885–901. http://dx.doi.org/10.1177/0013164410379332

DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing, 13,* 354–378. http://dx.doi.org/10.1080/15305058.2013.799067

Dombrowski, S. C. (2013). Investigating the structure of the WJ-III Cognitive at school age. *School Psychology Quarterly, 28,* 154–169. http://dx.doi.org/10.1037/spq0000010

Dombrowski, S. C. (2014a). Exploratory bifactor analysis of the WJ-III Cognitive in adulthood via the Schmid–Leiman procedure. *Journal of Psychoeducational Assessment, 32,* 330–341. http://dx.doi.org/10.1177/0734282913508243

Dombrowski, S. C. (2014b). Investigating the structure of the WJ-III Cognitive in early school age through two exploratory bifactor analysis procedures. *Journal of Psychoeducational Assessment, 32,* 483–494. http://dx.doi.org/10.1177/0734282914530838

Dombrowski, S. C., & Watkins, M. W. (2013). Exploratory and higher order factor analysis of the WJ–III full test battery: A school-aged analysis. *Psychological Assessment, 25,* 442–455. http://dx.doi.org/10.1037/a0031335

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

Fabrigar, L. R., Wegener, D. T., MacCullum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299. http://dx.doi.org/10.1037/1082-989X.4.3.272

Floyd, R. G., Bergeron, R., Hamilton, G., & Parra, G. R. (2010). How do executive functions fit with the Cattell–Horn–Carroll model? Some evidence from a joint factor analysis of the Delis–Kaplan executive function system and the Woodcock–Johnson III tests of cognitive abilities. *Psychology in the Schools, 47,* 721–738.

Floyd, R. G., McGrew, K. S., Barry, A., Rafael, F., & Rogers, J. (2009). General and specific effects on Cattell–Horn–Carroll broad ability composites: Analysis of the Woodcock–Johnson III normative update Cattell–Horn–Carroll factor clusters across development. *School Psychology Review, 38,* 249–265.

Floyd, R. G., Reynolds, M. R., Farmer, R. L., & Kranzler, J. H. (2013). Are the general factors from different child and adolescent tests the same? Results from a five-sample, six-test analysis. *School Psychology Review, 42,* 383–401.

Floyd, R. G., Shaver, R. B., & McGrew, K. S. (2003). Interpretation of the Woodcock–Johnson III Tests of Cognitive Abilities: Acting on evidence. In F. A. Schrank & D. P. Flanagan (Eds.), *WJ III clinical use and interpretation: Scientist–practitioner perspectives* (pp. 1–46). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/B978-012628982-4/50002-7

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence, 35,* 169–182. http://dx.doi.org/10.1016/j.intell.2006.07.002

Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: *g* as superordinate or breadth factor? *Psychology Science Quarterly, 50,* 21–43.

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research, 48,* 639–662. http://dx.doi.org/10.1080/00273171.2013.804398

Golay, P., & Lecerf, T. (2011). Orthogonal higher order structure and confirmatory factor analysis of the French Wechsler Adult Intelligence Scale (WAIS-III). *Psychological Assessment, 23,* 143–152. http://dx.doi.org/10.1037/a0021230

Gorsuch, R. L. (2003). Factor analysis. In J. A. Schinka & F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 143–164). Hoboken, NJ: Wiley.

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28,* 407–434. http://dx.doi.org/10.1207/s15327906mbr2804_2

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology, 17,* 65–70. http://dx.doi.org/10.1111/j.2044-8317.1964.tb00244.x

Hoelzle, J. B., & Meyer, G. J. (2013). Exploratory factor analysis: Basics and beyond. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (2nd ed., pp. 164–188). Hoboken, NJ: Wiley.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185. http://dx.doi.org/10.1007/BF02289447

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75,* 242–259. http://dx.doi.org/10.1037/h0025662

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Hunt, M. S. (2008). A joint confirmatory factor analysis of the Kaufman Assessment Battery for Children, Second Edition, and the Woodcock–Johnson Tests of Cognitive Ability, Third Edition, with preschool children. *Dissertation Abstracts International: Section A*. Humanities and Social Sciences, *68*(11), 4605.

Individuals with Disabilities Education Act of 2004, 20 U.S.C. § 1400 (2004).

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39,* 31–36. http://dx.doi.org/10.1007/BF02291575

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children* (2nd ed.). Circle Pines, MN: American Guidance Service.

Keith, T. Z., & Reynolds, M. R. (2012). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 758–799). New York: Guilford Press.

Keith, T. Z., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2008). Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock–Johnson III tests of cognitive abilities. *Intelligence, 36,* 502–525. http://dx.doi.org/10.1016/j.intell.2007.11.001

Kelley, T. L. (1939). Mental factors of no importance. *Journal of Educational Psychology, 30,* 139–142. http://dx.doi.org/10.1037/h0056336

Kline, R. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher, C. Schatschneider, & D. L. Compton (Eds.), *Applied quantitative analysis in education and the social sciences* (pp. 171–207). New York: Routledge.

Locke, S., McGrew, K. S., & Ford, L. (2011). A multiple group confirmatory factor analysis of the structural invariance of the Cattell–Horn–Carroll theory of cognitive abilities across matched Canadian and U.S. samples. *WMF Press Bulletin(1)* 1–18. Retrieved from http://woodcock-munoz-foundation.org/press/pressbulletins.html

McGrew, K. S. (2009). CHC theory and the Human Cognitive Abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37,* 1–10. http://dx.doi.org/10.1016/j.intell.2008.08.004

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Technical manual. Woodcock Johnson III Normative Update*. Rolling Meadows, IL: Riverside.

McGrew, K. S., & Wendling, B. J. (2010). Cattell–Horn–Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47,* 651–675.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical manual: Woodcock Johnson III*. Rolling Meadows, IL: Riverside.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749. http://dx.doi.org/10.1037/0003-066X.50.9.741

Mungas, D., Widaman, K., Zelazo, P. D., Tulsky, D., Heaton, R. K., Slotkin, J., . . . Gershon, R. C. (2013). VII. NIH Toolbox Cognition Battery (CB): Factor structure for 3 to 15 year olds. *Monographs of the Society for Research in Child Development, 78,* 103–118. http://dx.doi.org/10.1111/mono.12037

Muthén, L. K., & Muthén, B. O. (2013). *Mplus user's guide* (7th ed.). Los Angeles: Author.

Osborne, J. W., Costello, A. B., & Kellow, J. T. (2007). Best practices in exploratory factor analysis. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 86–99). Thousand Oaks, CA: Sage.

Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. (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. http://dx.doi.org/10.1521/scpq.20.1.66.64191

Raykov, T., & Marcoulides, G. A. (2000). *A first course in structural equation modeling*. Mahwah, NJ: Erlbaum.

R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667–696. http://dx.doi.org/10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95,* 129–140. http://dx.doi.org/10.1080/00223891.2012.725437

Reynolds, M. R., Floyd, R. G., & Niileksela, C. R. (2013). How well is psychometric *g* indexed by global composites? Evidence from three popular intelligence tests. *Psychological Assessment, 25,* 1314–1321. http://dx.doi.org/10.1037/a0034102

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales* (5th ed.). Itasca, IL: Riverside.

Sanders, S., McIntosh, D. E., 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. http://dx.doi.org/10.1002/pits.20211

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors on covariance structural analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 399–419). Thousand Oaks, CA: Sage.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). La Mesa, CA: Author.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53–61. http://dx.doi.org/10.1007/BF02289209

Schneider, W. J. (2013). What if we took our models seriously? Estimating latent scores in individuals. *Journal of Psychoeducational Assessment, 31,* 186–201. http://dx.doi.org/10.1177/0734282913478046

Schrank, F. A. (2011). Woodcock–Johnson III Tests of Cognitive Abilities. In A. S. Davis (Ed.), *Handbook of pediatric neuropsychology* (pp. 415–434). New York: Springer.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.

Tataryn, D. J., Wood, J. M., & Gorsuch, R. L. (1999). Setting the value of *k* in promax: A Monte Carlo study. *Educational and Psychological Measurement, 59,* 384–391. http://dx.doi.org/10.1177/00131649921969938

Taub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of Cattell–Horn–Carroll theory and cross-age invariance of the Woodcock–Johnson Tests of Cognitive Abilities III. *School Psychology Quarterly, 19,* 72–87. http://dx.doi.org/10.1521/scpq.19.1.72.29409

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

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41,* 321–327. http://dx.doi.org/10.1007/BF02293557

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41–71). Norwell, MA: Kluwer. http://dx.doi.org/10.1007/978-1-4615-4397-8_3

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods, 17,* 228–243. http://dx.doi.org/10.1037/a0027127

Wasserman, J. D., & Bracken, B. A. (2013). Fundamental psychometric considerations in assessment. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Vol. 10. Assessment psychology* (2nd ed., pp. 50–81). Hoboken, NJ: Wiley.

Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children—Fourth Edition among a national sample of referred students. *Psychological Assessment, 22,* 782–787. http://dx.doi.org/10.1037/a0020043

Watkins, M. W., Canivez, G. L., James, T., James, K., & Good, R. (2013). Construct validity of the WISC-IV^{UK} with a large referred Irish sample. *International Journal of School & Educational Psychology, 1,* 102–111. http://dx.doi.org/10.1080/21683603.2013.794439

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

Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment, 29,* 377–392. http://dx.doi.org/10.1177/0734282911406668

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70,* 123–133. http://dx.doi.org/10.1007/s11336-003-0974-7

Received March 12, 2014
Revision received September 24, 2014
Accepted October 1, 2014 ■

Correction to Strickland, Watkins, and Caterino (2015)

In the article "Structure of the Woodcock–Johnson III Cognitive Tests in a Referral Sample of Elementary School Students" by Tracy Strickland, Marley W. Watkins, and Linda C. Caterino (*Psychological Assessment*, Advanced online publication. February 2, 2015. http://dx.doi.org/10.1037/pas0000052), Linda C. Caterino's name and affiliation were not included in the byline or author note. All versions of this article have been corrected.

http://dx.doi.org/10.1037/pas0000160