Journal of Psychoeducational Assessment
2000, *18*, 27-38

THE RELATIONSHIP BETWEEN THE DIFFERENTIAL ABILITY SCALES (DAS) AND THE WOODCOCK-JOHNSON TESTS OF COGNITIVE ABILITY-REVISED (WJ-R COG) FOR STUDENTS REFERRED FOR SPECIAL EDUCATION EVALUATIONS

Ron Dumont
Fairleigh-Dickinson University

John O. Willis
Rivier College

Laurie P. Farr
North Country Education Foundation, Gorham, New Hampshire

Tim McCarthy
Laconia School District, Laconia, New Hampshire

Linda Price
Hudson-Litchfield School District, Hudson, New Hampshire

The Differential Ability Scales (DAS) and the Woodcock-Johnson Tests of Cognitive Ability-Revised (WJ-R COG standard battery) were administered to 81 children referred for special education services evaluation. Both the General Conceptual Ability (GCA) and the Nonverbal Reasoning scores of the DAS were significantly lower than the WJ-R COG Broad Cognitive Ability-Standard score (BCA-STD). The DAS Verbal and Spatial Composite scores were not significantly different from the BCA-STD. Low to moderate correlations were found between related constructs on the two tests.

Children referred for special education evaluations are often assessed with multiple diagnostic tools, each measuring several specific factors based on different theoretical constructs. Many questions need to be considered when selecting instruments. Will different instruments that purport to measure similar constructs (e.g., intelligence, spatial ability, math reasoning) produce comparable results? Should an instrument be used again for re-evaluations? If not, will differences in scores reflect differences between the instruments, changes in the student's abilities, or chance variation? Will changes in the apparent pattern of cognitive abilities reflect genuine changes in the student?

Sections in test manuals that address the issue of construct validity often provide helpful information, which allows for the prediction of scores from one test to another. However, validity data in manuals are often based on small samples, which may have group characteristics that are different from the characteristics of the individual being evaluated.

Studies that provide data about similarities and differences between tests are important for several reasons. First, they provide information about the amount of overlap and similarity of the constructs measured across instruments. Second, they predict expected differences between scores related to the constructs assessed in common across instruments. This information is important because identification of certain disabilities often depends on the student's test scores. Differences between scores on two tests that purport to measure the same abilities could result in different classifications and different interpretations.

The extent to which normative data generalize to students with disabilities is often unknown, especially when evaluators use different, though highly correlated instruments. For example, Prewett and Matavich (1994) found that when the Wechsler Intelligence Scale for Children, 3rd ed. (WISC-III; Wechsler, 1991) was compared to the Stanford-Binet: Fourth Edition (SB: IV; Thorndike, Hagen, & Sattler, 1986), the results, although highly correlated, "...did not give the same diagnostic impressions..."(p. 46).

Recently an integrated Carroll/Horn-Cattell *Gf-Gc* cross-battery approach to test interpretation (e.g., Flanagan & McGrew, 1997; McGrew & Flanagan, 1998) has been suggested for interpreting batteries of intellectual measures. Because no current test adequately incorporates all of the constructs from the Horn-Cattell Fluid-Crystallized (*Gf-Gc*) theory (Carroll, 1993; Cattell, 1941, 1957; Horn, 1985; Horn & Cattell, 1966), clinicians are encouraged to "fill the holes" by choosing appropriate subtests from other batteries. "Through supplementing their preferred test battery with tests from other intelligence batteries, clinicians can conduct and interpret assessments that are more comprehensive and theoretically sound..." (Flanagan & McGrew, 1997, p. 315). There is, however, continued debate in the field regarding the value of the *Gf-Gc* theory as a theoretical foundation, the abilities that underpin the *Gf-Gc* theory, and whether these abilities are distinct (Carroll, 1993). Interpretation of cognitive tests according to test factor structures can be a powerful tool for understanding a student's intellectual abilities (e.g., Carroll, 1993; Flanagan & McGrew, 1997; Kaufman, 1994; McGrew, 1997; McGrew & Flanagan, 1998), but not when factors with the same name (e.g., "crystallized ability" or "*Gc*") do not measure the same abilities. Throughout this paper, the *Gf-Gc* broad and narrow factor designations are derived from McGrew and Flanagan (1998), but see also Carroll (1993), Flanagan and McGrew (1997), and Woodcock (1990) for further explanation of the *Gf-Gc* theory.

There has as yet been no published study comparing the Differential Ability Scales (DAS; Elliott, 1990) and the Woodcock-Johnson Tests of Cognitive Ability (WJ-R COG; Woodcock, 1997) scores of a sample of referred students. Although both tests are based on the *Gf-Gc* theory, the designs for measuring those constructs differ (Elliott, 1997; Woodcock, 1997).

Six core subtests on the DAS School-Age Battery combine to assess three (i.e., Crystallized Intelligence [*Gc*], Fluid Intelligence [*Gf*], and Visual Intelligence [*Gv*]) of the eight factors in the McGrew and Flanagan integrated Carroll/Horn-Cattell *Gf-Gc* cross-battery model (Elliott, 1997). From these factors, the General Conceptual Ability (GCA) score is determined. Additional diagnostic subtests (not taken into account in the GCA) include measures of

Processing Speed (*Gs*), Short-Term Memory (*Gsm*), and Long-Term Storage and Retrieval (*Glr*). In contrast to the DAS, the WJ-R COG Broad Cognitive Ability (BCA) is derived from seven of the eight Horn and Cattell factors (Woodcock, 1990, 1997). Besides *Gc, Gf,* and *Gv,* the WJ-R COG BCA also includes Processing Speed (*Gs*), Short-Term Memory (*Gsm*), Long-Term Storage and Retrieval (*Glr*), and Auditory Processing (*Ga*). The WJ-R COG provides a BCA-Standard (BCA-STD) or a BCA-Extended (BCA-EXT) score based upon 7 or 14 subtests, respectively. With 14 subtests, seven factor scores can be obtained, each based on 2 subtests.

The purpose of this study was to examine the relationship between the DAS and the WJ-R COG subtests of similar theoretical orientation and the tests' overall scores. Table 1 presents both the broad and narrow factors of the *Gf-Gc* theory and identifies the subtests of the DAS and the WJ-R COG with their respective factors, according to McGrew and Flanagan (1998).

Table 1 suggests that moderate to high correlations would exist among the following broad *Gf-Gc* groupings of DAS and WJ-R COG subtests:

1. *Gc*—DAS Word Definitions and Similarities and WJ-R COG Picture Vocabulary and Memory for Sentences;
2. *Gf*—DAS Matrices and Sequential & Quantitative Reasoning and WJ-R COG Analysis-Synthesis;
3. *Gv*-DAS Pattern Construction, Recall of Designs, and Recall of Objects (Immediate and Delayed) and WJ-R COG Visual Closure;
4. *Gsm*—DAS Recall of Digits and WJ-R COG Memory for Sentences; and
5. *Glr*—DAS Recall of Objects (Immediate and Delayed) and WJ-R COG Memory for Names.

Three additional, narrow-domain comparisons were:

1. Language Development and Lexical Knowledge—DAS Word Definitions and Similarities and WJ-R COG Memory for Sentences and Picture Vocabulary;
2. Quantitative Reasoning—DAS Sequential & Quantitative Reasoning and WJ-R COG Analysis-Synthesis; and
3. Memory Span—DAS Recall of Digits and WJ-R COG Memory for Sentences.

METHOD

Participants

Participants were 81 students (47 males, 34 females; 78 White, 3 Black) attending two rural/suburban school districts in southern New Hampshire. The mean age of the sample was 12 years 4 months (*SD* = 3 months; range = 6 years 6 months to 17 years 8 months). Each student was referred for a special education evaluation after extensive "pre-referral" interventions. A multidisciplinary team evaluated all students. Of this referred sample, some were subsequently found to meet criteria for special education services (New Hampshire *Standards for the Education of Handicapped Students,* 1988), with a deficit in one or more of the basic neurological "processes" and a significant difference between their actual and predicted achievement scores.

Table 1

Broad and Narrow McGrew, Flanagan, and Ortiz Integrated Carroll-Cattell Gc-Gf Factors and the Subtests That Measure Them

Broad Factor of *G*	Subtests from	
Narrow Factor of *G*	DAS	WJ-R COG
Crystallized Intelligence (*Gc*)		
Language Development	Word Definitions	Memory for Sentences
	Similarities	Picture Vocabulary
Lexical Knowledge	Word Definitions	Picture Vocabulary
	Similiarities	
General (Verbal) Information		Picture Vocabulary
Fluid Intelligence (*Gf*)		
General Sequential Reasoning		Analysis/Synthesis
Induction	Matrices	
	Sequential & Quantitative Reasoning	
Quantitative Reasoning	Sequential & Quantitative Reasoning	Analysis/Synthesis
Visual Processing (*Gv*)		
Visualization	Pattern Construction	
Spatial Relations	Pattern Construction	
Closure Speed		Visual Closure
Visual Memory	Recall of Designs	
	Recall of Objects	
Short-term Memory (*Gsm*)		
Memory Span	Recall of Digits	Memory for Sentences
Long-term Storage and Retrieval (*Glr*)		
Associative Memory		Memory for Names
Free Recall Memroy	Recall of Objects	
Processing Speed (*Gs*)		
Perceptural Speed		Visual Matching
Rate-of-Test-Taking	Speed of Information Processing	Visual Matching
Auditory Processing (*Ga*)		
Phonetic Coding		Incomplete Words
Resistance to Auditory Stimulus Distortion		Incomplete Words
Decision/Reaction Time or Speed (*Gf*)		
Mental Comparison Speed	Speed of Information Processing	

Adapted from McGrew, K. S., & Flanagan, D. P. (1998), *The Intelligence Test Desk Reference (ITDR): Gf-Gc Cross-Battery Assessment* (Allyn & Bacon, 1998), ch. 1, 4, & 11 and Appendix A. See also Carroll (1993) and Flanagan & McGrew (1997).

Procedure

Each student was administered at least the seven-subtest WJ-R COG Battery and the six core subtests and two diagnostic subtests of the DAS: Recall of Digits and Recall of Objects. Most children were evaluated by at least two different examiners, usually one who administered the DAS and a second who administered the WJ-R. Time between testing ranged from 1 to 8 days. No attempt was made to counterbalance the administrations of the tests: 64% were administered the DAS first, and 36% were administered the WJ-R COG first. Because the cross-battery approach (McGrew & Flanagan, 1998) incorporates subtests from different measures standardized at different times and on different samples, the need to counterbalance should be limited.

Instruments

The DAS is an individually administered measure of cognitive ability, achievement, and information processing. The DAS was designed for ages 2 years 6 months through 17 years 11 months and was normed on 3,475 children. The battery provides separate norms for three different age levels: lower preschool, upper preschool, and school age. Subtests administered to school-age children combine to form three cluster scores (Verbal, Nonverbal Reasoning, and Spatial) and a General Conceptual Ability score (GCA). A Special Nonverbal Composite also is available. A number of diagnostic subtests provide additional information about a child's functioning but are not used in determining cluster scores or the GCA. In the *Gf-Gc* model, the DAS Verbal cluster is considered to assess crystallized intelligence (*Gc*), the Nonverbal Reasoning cluster assesses fluid intelligence (*Gf*), and the Spatial cluster assesses visual processing (*Gv*). Three of the DAS diagnostic subtests (Recall of Digits, Recall of Objects, and Speed of Information Processing) provide measures of *Gsm*, *Glr*, and *Gs*, respectively. Mean internal reliabilities for the DAS school-age subtests range from .76 to .91; clusters range from .88 to .92; and the school-age GCA mean reliability is .95. Test-retest reliability values indicate that the DAS scores are highly stable. Total time for administering the six core and three diagnostic subtests is approximately 45 minutes to 1 hour.

The Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R), a measure of cognitive ability and achievement, was designed to assess children and adults between the ages of 2 and 95. It was normed on 6,359 children and adults. The WJ-R is composed of two major parts: the Woodcock-Johnson Tests of Cognitive Ability (WJ-R COG) and the Woodcock-Johnson Tests of Achievement (WJ-R ACH). The WJ-R COG provides the examiner the option of establishing a Broad Cognitive Ability score based on a 5-subtest Early Development (BCA-EDEV) battery, a 7-subtest Standard (BCA-STD) battery, or a 14-subtest Extended (BCA-EXT) battery. (Examiners may also choose to administer a complete 21-subtest WJ-R COG, but the final 7 supplemental subtests are not included in either the computation of the BCA or the cluster scores. If an examiner uses the Woodcock scoring program [Schrank & Woodcock, 1998] to aid in scoring and interpreting the intra-cognitive discrepancies of the test, only the seven-subtest Standard composite [BCA-STD] is used to establish those discrepancies.) Median internal reliabilities for the subtests range from .75 to .93; reliabilities for clusters range from .81 to .95; and reliability of the BCA-STD is .97. Test-retest reliability values indicate that the scores are highly stable. Total time for administering the seven standard subtests is approximately 30 to 40 minutes, with an additional 40 minutes needed for the supplemental battery (subtests 8 to 14).

Because some cases excluded one or more of the DAS diagnostic subtests, correlations were established using pair-wise deletions followed by Fisher's r to z transformation. Paired comparison, two-tailed t tests with a hypothesized difference of zero were used for score differences.

RESULTS

Table 2 presents the means, standard deviations, and correlations between the DAS and WJ-R COG composite scores. Overall, this referred sample's mean GCA score (94.15) was a significant, $t(80) = -3.51$, $p < .001$, 5.85 points below the normative mean of 100, but the mean BCA-STD score (96.95) was a nonsignificant, $t(80) = -2.08$, $p = .041$, 3.05 points below the normative mean. Of the three DAS clusters, there were significantly lowered Verbal, 96.21, $t(80) = -2.28$, $p = .025$, and Nonverbal Reasoning, 90.88, $t(80) = -5.79$, $p < .0001$, scores, but the Spatial score of 97.79 was not significantly different from the normative mean, $t(80) = -1.25$, $p = .214$.

The DAS GCA score (94.15) was a statistically significant, $t(80) = -2.11$, $p = .038$, 2.8 standard score points lower than the WJ-R COG score (BCA-STD, 96.95). The DAS Verbal and Spatial cluster scores (means of 96.21 and 97.79, respectively) were not significantly different from the BCA-STD score, $t(80) = .741$, $p = .58$ and $t(80) = -.52$, $p = .61$. The DAS Nonverbal Reasoning cluster (90.88) was a statistically significant, $t(80) = 3.98$, $p = .0002$, 6.07 points below the BCA. To estimate the effect size of these differences, gammas (Howell, 1987) were calculated. Gammas, which signify the difference between selected means, are expressed in standard deviation units and ranged from .05 to .40. Using conventions suggested by Cohen (1977), these would indicate low effect size differences between the two measures.

All correlations between the DAS GCA, DAS clusters, and the WJ-R BCA-STD were significant ($p = <.0001$) and ranged from .50 to .65. The DAS Verbal cluster score correlated slightly higher (.64) with the BCA-STD than did the Nonverbal Reasoning and Spatial clusters (.50 and .51, respectively). The DAS GCA score correlated moderately (.65) with the BCA-STD score. This correlation suggests that the DAS and WJ-R COG had approximately 42% shared variance (r^2) and that they assessed some similar skills. However, the amount of variance not shared between the tests indicates that, at least with referred students similar to those in this study, the two tests cannot be used interchangeably.

Similarly, the DAS Verbal cluster shares approximately 41% of the variance with the WJ-R BCA-STD. However, the DAS Nonverbal Reasoning and Spatial clusters have no more than about 26% shared variance with the WJ-R BCA-STD.

Table 2

Means, Standard Deviations, Correlations, and Gammas for DAS/WJ-R Composites

	DAS Composites				WJ-R
	GCA	Verbal	Nonverbal	Spatial	BCA-STD
Mean	94.15† *	96.21†	90.88† **	97.79	96.95
SD	15.02	14.97	14.17	15.90	13.22
r	0.65	0.64	0.50	0.51	
γ	0.19	0.05	0.40	0.06	

Note.— *N* = 81.

All correlations (*r*) between the DAS Composites and WJ-R BCA-STD are significant ($p < .0001$).

The effect sizes (γ) are gammas (Howell, 1987). They signify the difference between means of the DAS and WJ-R in standard deviation units. Cohen (1977) defines effect sizes as small if they are 0.20, medium if they are 0.50, and large if they are 0.80 or higher.

† Indicates significant ($p < .05$) difference from the normative mean of 100.

* Indicates significant ($p < .05$) difference between the BCA and GCA.

** Indicates significant ($p < .001$) difference between the BCA and Nonverbal Reasoning.

Table 3 presents the means and standard deviations for the DAS and WJ-R COG subtests. Because the DAS and the WJ-R COG report subtest scores with different metrics (*T* scores vs. standard scores), the DAS subtest *T* scores were converted to standard scores with a mean of 100 and a standard deviation of 15. Four of the six core subtests—Word Definitions (*Gc*), Matrices (*Gf*), Sequential & Quantitative Reasoning (*Gf*), and Pattern Construction (*Gv*)—and all of the diagnostic subtests—Recall of Digits (*Gsm*) and Recall of Objects-Immediate & Delayed (*Glr*)–of the DAS were significantly lower than the normative mean of 100. Similarities (*Gc*) and Recall of Designs (*Gv*) were not significantly different from the mean. For the WJ-R COG, three of the subtests—Memory for Names (*Glr*), Visual Matching (*Gs*), and Incomplete Words (*Ga*)—were significantly below the normative mean of 100, whereas one subtest—Visual Closure (*Gv*)—was significantly higher than the mean.

Of the five factors (*Gc, Gf, Gv, Gsm,* and *Glr*) represented on both the DAS and WJ-R COG in this study, three (*Gf, Gv,* and *Gsm*) showed significant differences between the two instruments. In each case, the mean WJ-R COG factor score was in the upper half of the average range, and the DAS cluster or subtest score was significantly lower. The differences were sufficiently large (approximately 8 to 10 standard score points) to suggest practical as well as statistical significance. Effect sizes (gammas) were computed for each of the five factors. Overall mean effect size across all subtests was .43, indicating that approximately 67% of the WJ-R COG scores exceeded the mean scores of the DAS. For the three factors (*Gf, Gv,* and *Gsm*) that showed significant differences, effect size ranged from .43 to .77. Using Cohen's classifications, these effects would be considered low to medium. These effect sizes suggest that approximately 67% to 78% of the WJ-R COG subtest scores were above the respective DAS subtest mean for similar *Gf-Gc* constructs. For the remaining two factors, *Gc* and *Glr,* the differences were small and not significant, with effect sizes considered low.

Table 3
Means and Standard Deviations for DAS/WJ-R Broad Domain Subtests

	DAS SS Scores		WJ-R SS Scores		Significance ($p < .01$) between DAS/WJ-R	γ
	Mean	*SD*	*Mean*	*SD*		
Gc						
Word Definition	96.48†	15.32				0.11
Similarities	97.61	14.86				0.07
Picture Vocabulary			98.68	13.78		
Gf						
Matrices	91.06†	14.86				0.77
Sequential & Quantitative	93.48†	13.44				0.60
Analysis and Synthesis			102.54	15.00	Sig	
Gv						
Recall of Designs	100.07	16.84				0.43
Pattern Construction	96.84†	14.34				0.64
Visual Closure			106.51†	15.57	Sig	
Gsm						
Recall of Digits	90.52†	14.11				0.69
Memory for Sentences			100.80	15.98	Sig	
Glr						
Recall of Objects	92.75†	18.86				0.18
Recall of Objects Delayed	90.25†	16.20				0.35
Memory for Names			95.52†	16.62		
Gs						
Speed of Information Processing			Not Administered			
Visual Matching			93.41†	14.72		
Ga						
Incomplete Words			94.15†	13.14		

Note.—All Standard Scores (SS) are based on a mean of 100 and standard deviation of 15.
For the DAS, *T* scores were converted to Standard Scores.
† Indicates significant ($p < .05$) difference from the normative mean of 100.
The effect sizes (γ) are gammas (Howell, 1987). They signify the differences between means of the DAS and the WJ-R in standard deviation units. Cohen (1977) defines effect sizes as small if they are 0.20, medium if they are 0.50, and large if they are 0.80 or higher.

Table 4 presents the significant correlations between the DAS and the WJ-R COG subtests. Among the broad *Gf-Gc* constructs, the correlations between DAS core subtests and WJ-R COG subtests show the following pattern: The DAS *Gc* subtests (Word Definitions and Similarities) correlated with the Picture Vocabulary (*Gc*) subtest .53 and .62, respectively. The DAS *Gf* subtests (Matrices and Sequential & Quantitative Reasoning) correlated .37 and .54, respectively, with the WJ-R Analysis-Synthesis (*Gf*) subtest. Interestingly, the DAS *Gv* subtests (Recall of Designs and Pattern Construction) did not correlate highly with the WJ-R Visual Closure (*Gv*). These two DAS *Gv* subtests actually correlated highest (both .44) with the WJ-R Analysis-Synthesis (*Gf*) subtest. In fact, there was no significant correlation between the Recall of Designs and the Visual Closure subtest. The DAS Recall of Digits (*Gsm*) correlated highest (.62) with the Memory for Sentences (*Gsm*) subtest, whereas the Recall of Objects-Immediate and -Delayed subtests, both measures of *Glr,* revealed no significant correlations with Memory for Names nor any of the other six WJ-R subtests.

Table 4
Significant Correlations for DAS/WJ-R Subtests

DAS	(*Gc*) PVoc	(*Gf*) AnSy	(*Gv*) Vcl	WJ-R (*Gsm*) Ms	(*Glr*) Mn	(*Gs*) Vm	(*Ga*) Incw
Gc							
Word Definition	.53	.51		.43	.35	.23	
Similarities	.62	.39		.46			.23
Gf							
Matrices	.40	.37		.27		.29	
Sequential & Quantitative	.32	.54		.25		.34	
Gv							
Recall of Designs	.34	.44		.31			
Pattern Construction	.36	.44	.22		.32		
Gsm							
Recall of Digits				.62			.26
Glr							
Recall of Objects							
Recall of Objects Delayed							

Note.—*N* = 81 except: Recall of Digits (*n* = 63); Recall of Objects (*n* = 60)
Correlations are based on pairwise comparisons.
Only correlations at $p < .05$ are reported.
Correlations in double-lined boxes indicate the overlap of broad and narrow *Gf-Gc* domains.
Correlations in dotted-lined boxes indicate the remaining broad *Gf-Gc* domains.
Correlations in single-lined boxes indicate the remaining narrow *Gf-Gc* domains.

Comparison of the correlations for the narrow domains of the *Gf-Gc* construct showed both some expected and surprising relationships. The DAS Word Definitions and Similarities and the WJ-R COG Memory for Sentences and Picture Vocabulary subtests, all measures of Language Development, correlated moderately well (range = .43 to .62). The DAS Word Definitions and Similarities subtests and the WJ-R COG Picture Vocabulary subtest, measures of Lexical Knowledge, also correlated moderately well (.53 to .62).

As expected, the DAS Sequential & Quantitative Reasoning subtest correlated better with the WJ-R COG Analysis-Synthesis subtest (.54) than did the DAS Matrices subtest with Analysis-Synthesis (.37). Matrices is a measure of Induction, whereas the DAS Sequential & Quantitative Reasoning and WJ-R COG Analysis-Synthesis are measures of Sequential Reasoning. In fact, the DAS Matrices subtest correlation with the WJ-R COG Picture Vocabulary subtest (.40) was similar to the Matrices correlation with the WJ-R COG Analysis-Synthesis subtest (.37), but Matrices produced low, but significant correlations with the WJ-R COG Memory for Sentences (.27) and Visual Matching (.29) subtests. Finally, the DAS Recall of Digits and the WJ-R COG Memory for Sentences subtests, both measures of Memory Span, correlated moderately well (.62).

DISCUSSION

The results from this sample of referred children indicate that the WJ-R COG BCA-STD and DAS GCA scores correlated significantly, but moderately, for this sample. There was a significant correlation between the overall ability scores, but only about 42% of the variance in the abilities being assessed was shared. Although the composite scores from these two batteries can be interpreted as indicators of the same *Gf-Gc* abilities (McGrew & Flanagan, 1998), there is no reason to assume that they are identical for referred students similar to those in this study. This study suggests that different scores for the same *Gf-Gc* abilities will often be found.

Though not surprising, a lack of equivalence between two popular intelligence tests in a sample of referred students should raise disturbing concerns for examiners and agencies that persist in applying rigid statistical formulae for the identification of disabilities. Students who would otherwise be eligible for services may be denied those services, and other students may be erroneously labeled as having disabilities.

Bracken (1988) offered "Ten Psychometric Reasons Why Similar Tests Produce Dissimilar Results." Several of his reasons may have contributed to these dissimilar results:

Many WJ-R subtests do appear to have much steeper "Item Gradients" than most DAS subtests, such that raw score differences of a single point cause relatively large changes in derived scores. That apparent difference between the instruments might have contributed to the differences found here. The "Norm Table Layout" of the WJ-R is as fine-grained as any currently available test, with separate norms for each month of age (or of grade placement). The DAS, on the other hand, uses the more common conventions of 3-month or 6-month spans so students at the bottom end of the age range on a page in the norms tables receive slightly depressed scores, and students at the top end receive slightly inflated scores. Most important of Bracken's (1988) reasons were probably "Skill Differences Assessed Across Tests" and "Content Differences Across Tests."

Also, although both the DAS and WJ-R incorporate aspects of the Rasch Model of item-response theory (IRT), each test approaches the issue of ceilings and discontinuation differently. The DAS assesses ability by having the examiner test within predetermined blocks of items. Typically, as long as a child passes three items in the block and fails three items within the block, the examiner discontinues testing. There is no need for a child to fail consecutive items or even the final items. In contrast, on the WJ-R COG, ceilings are typically established when a child has failed six consecutive items, with the further caveat that the child must be administered an entire page of stimuli, failing the final item.

The Pattern Construction and Visual Closure differences may be the result of the narrow abilities being measured (Visualization vs. Closure Speed). Pattern Construction requires the manual manipulation of blocks and provides scores based on speed. Visual Closure requires the child to name distorted pictures.

If we were to accept *Gf-Gc* theory as a useful, transferable method of analyzing abilities measured by, or task demands of, various cognitive tests and were to apply the "cross-battery approach," we would need to determine the minor, as well as the major, factor loadings and tasks demands of each subtest (McGrew & Flanagan, 1998). As discussed by Bracken (1988) and shown by the present study and others (e.g., Prewett & Matavich, 1994), minor, incidental aspects of tests may contribute to significant differences between tests intended to assess the same cognitive factor. The proposed *Gf-Gc* constructs represent very broad domains. Composites or cluster scores that seem to measure the same broad ability may be based on different mixes of narrow abilities (McGrew & Flanagan, 1998). When subtests subsumed under the same broad ability label actually do not correlate at levels of practical or even statistical significance, examiners must be very cautious in their assumption that the tests are in fact measuring the constructs attributed to them. No subtest, much less a factor score based on more than one subtest, is likely to be a "pure" measure of any *Gf-Gc* factor. Incidental contributions of other task demands or abilities, irrelevant to the average student, may alter the nature of a test for a particular individual, especially one with a specific learning disability. For example, the DAS Recall of Designs is intended to measure spatial, visual ability (*Gv*) but may reflect attention or memory difficulties for a student with Attention-Deficit/Hyperactivity Disorder. Again, the DAS Sequential and Quantitative Reasoning and WJ-R Analysis-Synthesis subtests are both measures of fluid reasoning (*Gf*) that require mathematical thinking. However, only the DAS subtest requires the actual calculation skills taught in school. Consequently, a student retained in grade might be at a disadvantage on the DAS, but not on the WJ-R subtest. Some of the unanticipated patterns of mean differences and correlations among DAS and WJ-R COG scores with the present sample of referred students suggest that examiners must realize that a subtest or factor score may not always reflect the ability it is supposed to measure.

Because of the proposal for "cross-battery" approaches to test interpretation (Carroll, 1993; Flanagan & McGrew, 1997; McGrew & Flanagan, 1998), further research into how clinical samples perform on different tests is needed. Ultimately, studies are needed that compare the factor structures and intercorrelations of all the major intelligence tests with very large, nationally representative samples and large samples of students with various clearly defined disabilities. From the present, small, referred sample, it seems likely we would find different patterns of subtest intercorrelations for various groups of students.

Another important concern in cross-battery interpretation is the administration of individual subtests taken out of the context and sequence in which they were given during standardization. Most manuals caution that it is important to adhere strictly to both directions and sequence of administration. Currently, it is unknown the extent to which subtests taken out of the context of their entire test will measure the same abilities. Perhaps new tests should be normed with subtests administered in counterbalanced or random order in anticipation of the uses recommended by McGrew and Flanagan (1998).

REFERENCES

Bracken, B. A. (1988). Ten psychometric reasons why similar tests produce dissimilar results. *Journal of School Psychology, 26*, 155–166.

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

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38,* 592.

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* New York: Academic Press.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). San Diego, CA: Academic Press.

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

Elliott, C. D. (1997). The Differential Ability Scales. In D P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 183–208). New York: Guilford Press.

Flanagan D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 314–325). New York: Guilford Press.

Horn, J. L. (1985). Remodeling old models of intelligence: *Gf-Gc* theory. In B. B. Wolman (Ed.), *Handbook of intelligence* (pp. 267–300). New York: Wiley.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57,* 253–270.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: *Gc-Gf* theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 53–91). New York: Guilford Press.

Howell, D. C. (1987). *Statistical methods for psychology* (2nd ed.). Boston: PWS-Kent.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III.* New York: Wiley.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive *Gf-Gc* framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 151–179). New York: Guilford Press.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment.* Boston: Allyn & Bacon.

New Hampshire Department of Education Special Education Bureau. (1988). *Standards for the education of handicapped students.* Concord, NH: Department of Education.

Prewett, P. N., & Matavich, M. A. (1994). A comparison of referred students' performance on the WISC-III and the Stanford-Binet Intelligence Scale: Fourth Edition. *Journal of Psychoeducational Assessment, 12,* 42—48.

Schrank, F. A., & Woodcock, R. W. (1998). *Woodcock Scoring and Interpretive Program* [Computer software]. Itasca, IL: Riverside.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale: Fourth Edition.* Chicago: Riverside.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231–258.

Woodcock, R. W. (1997). The Woodcock-Johnson Tests of Cognitive Ability-Revised. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests and Issues* (pp. 230–246). New York: Guilford Press.

Woodcock, R. W., & Johnson, M B. (1989). *Woodcock-Johnson Psycho-educational Battery-Revised.* Chicago: Riverside.