Is the Demise of IQ Interpretation Justified? A Response to Special Issue Authors

Catherine Fiorello

Applied ...

Applied Neuropsychology
2007, Vol. 14, No. 1, 37–51

Is the Demise of IQ Interpretation Justified? A Response to Special Issue Authors

James B. Hale
Philadelphia College of Osteopathic Medicine, Philadelphia, Pennsylvania, USA
Catherine A. Fiorello
Temple University, Philadelphia, Pennsylvania, USA
Jack A. Kavanagh
Loyola University, Chicago, Illinois, USA
James A. Holdnack
Harcourt Assessment, Inc., San Antonio, Texas, USA
Ariel M. Aloe
Florida State University, Tallahassee, Florida, USA

Standardized testing of intellectual and cognitive functioning remains a critical component of psychological assessment despite widespread criticism of the practice. Although most standardized intellectual measures are some of the best tools available to practitioners, opponents of intellectual assessment argue the traditional use of global IQ-achievement discrepancy has little diagnostic utility or treatment validity. It is time to move beyond the academic rhetoric of global intelligence to make standardized intellectual assessment meaningful for individual children. In this paper, we respond to special issue authors by presenting clinical and statistical arguments that support idiographic interpretation of intellectual measures for children with disabilities and variable test profiles, and offer recommendations for practice that demonstrate the clinical utility of such approaches. If practitioners move beyond global IQ interpretation, and methods for objective idiographic interpretation are established, the practice of intellectual assessment will be once again valued and respected among those in clinical and educational practice.

Key words: idiographic, intelligence, IQ, nomothetic, process approach

One of the most notable, influential, and revered neuropsychologists was A. R. Luria. Luria's (1973) enduring approach was highly idiographic, and clinical acumen was something he valued and taught. A review of his works reveals little support for psychometric approaches to understanding brain structure and function, for interpretation of both quantitative and qualitative data are critical for our understanding of the applied neuropsychology of individual differences (Kaplan, 1988). Luria (1979) was especially critical of using standardized measures to determine individual intellectual status, because summative IQ scores represent disparate cognitive constructs and can obscure an individual's neuropsychological status, a position held by many prominent neuropsychologists (Kaplan, 1988; Lezak, 1988). Surely, neuropsychological evaluation requires objective data collection, but it was the inferences drawn from such data that hold the greatest clinical utility in Luria's model. In Luria's worldview, examination of multiple individual characteristics in response to the environmental and assessment demands formed the interpretive basis of understanding underlying neuropsychological processes and

WISC-IV and WIAT-II data were obtained with permission of Harcourt Assessment Inc., WISC-IV Copyright (2003).

Address correspondence to James B. Hale, School Psychology Program, Philadelphia College of Osteopathic Medicine, 4190 City Avenue, Philadelphia, PA 19131. E-mail: jamesha@pcom.edu

integrated brain systems. This was by far more important than focusing on any particular summative outcome, such as global IQ (e.g., Luria, 1979). However, as demonstrated in this special issue, some researchers continue to strongly admonish practitioners to focus on intransient measures such as global intelligence instead of interpreting cognitive and/or neuropsychological processes, even though research has shown such practice has little diagnostic or treatment validity (e.g., Fletcher et al., 2002).

While this entrenched psychometric perspective continues to pervade academic circles, a majority of practitioners working with real children continue to rely on idiographic analysis of intellectual measures rather than global IQ interpretation (Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000). Instead of developing objective methods for evaluating multifactorial intellectual approaches (e.g., Cattell-Horn-Carroll Theory; Flanagan, McGrew, & Ortiz, 2000), and demonstrating the utility of such information using single subject designs to demonstrate diagnostic and treatment validity (e.g., Hale & Fiorello, 2004), advocates of global IQ persist in their attempts to eliminate such practice (e.g., McDermott, Fantuzzo, & Glutting, 1990).

As can be seen in these special issue papers, advocates of global IQ rarely address the plethora of neuropsychological evidence that provides insight into brain functioning, and how changes in brain functioning result from effective interventions (Hale, Kaufman, Naglieri, & Kavale, 2006). We don't have one brain for intelligence tests, another for performance on neuropsychological measures, and a third for acting within our environment, a fact overlooked by those who advocate global IQ interpretation. Perhaps IQ advocates are enticed by large-scale standardization studies that show a positive manifold among intellectual subtests (Kranzler, 1997) that results in an enduring presence of a one-factor solution for intelligence tests (Jensen, 1998). Perhaps it is a reductive need to simplify our understanding of individual differences, as certainly global IQ is a parsimonious predictor of many meaningful outcomes (Gottfredson, 1997). If global IQ is the only meaningful difference, policy can be developed that takes this information into account (e.g., Rushton & Jensen, 2005). Simply put, there is little that practitioners can do to dramatically change global intelligence, because it is relatively stable over time (Carroll, 1997), which suggests social programs designed to ameliorate disparities experienced by underprivileged members of society could be dramatically changed or even eliminated (e.g., Herrnstein & Murray, 1994).

Despite the temptation to engage in a philosophical or politically laden discussion that could mask the important contributions presented in this special issue, we will instead use this opportunity to examine our original findings within the context of each commentary paper, and provide clinical and statistical evidence that addresses their claims. However, it is important to recognize that these debates have raged for some time, and no matter how persuasive the arguments, it is more important to examine the data presented, not the contentious opinions expressed by opposing factions. Whether one believes in the value of global intellectual functioning, or finds it of little relevance in clinical practice, science must dictate our understanding of reality, and temper our ideological fervor.

REPLY TO DANA AND DAWES: DOES SIMULATION REPRESENT REALITY?

The Dana and Dawes (this issue) paper succinctly describes the issues of paramount importance to practitioners: Is global IQ interpretation meaningful given the limited shared variance found for the four WISC-IV Indices, and is idiographic interpretation warranted given that the large amounts of unique variance found for children with learning disabilities (LD), Attention-Deficit-/Hyperactivity Disorder (ADHD), and traumatic brain injury (TBI)? Dana and Dawes suggest that traditional exploratory factor analysis should be used instead of the commonality analysis Fiorello et al. (this issue) employed, and suggest that a one-factor solution representing "*g*" would be found. They go on to argue that the "positive manifold" among subtests suggests "linear combinations of the subtests are so highly correlated with each other that they are nearly indistinguishable, rendering differential weighting unimportant."

As noted by Horn (2007), surely a one factor solution can be found for any set of intellectual subtests, but this does not imply that this solution adequately represents the structure of intellectual functions. In contrast to the Dana and Dawes (this issue) position, differential weighting based on factor analytic findings is theoretically and empirically supported on the Woodcock Johnson-III Tests of Cognitive Abilities (Woodcock, McGrew, & Mather, 2001). An examination of the exploratory and confirmatory factor analyses conducted with the WISC-IV standardization sample (Wechsler, 2003) demonstrates that a four factor Verbal Comprehension (VC), Perceptual Reasoning (PR), Working Memory (WM), and Processing Speed (PS) solution better represents the data than a single factor "*g*" solution, which of course can be found (e.g., Horn, 2007). In addition, three or four factor solutions are the standard findings reported in countries throughout the world (Georgas, Weiss, van de Vijer, & Saklofske, 2003).

The issue of paramount importance here is construct validity. As noted in their discussion of construct validity, Anastasi and Urbina (1997), "Each construct is developed to explain and organize observed response consistencies... Construct validation requires the gradual accumulation of information from a variety of sources. Any data throwing light on the nature of the trait under consideration, and the conditions affecting its development and manifestations represent appropriate evidence for this validation" (p. 126). It seems clear that the data we present does provide additional validity evidence through commonality analysis, evidence that may not represent large scale factorial studies, but instead represent clinical populations such as LD, ADHD, and TBI, children practitioners commonly see in clinical practice. While data supporting a multifactorial model over a single factor model might not alter the position of Dana and Dawes (this issue) and other academics, because a single factor can still be found (Horn, in press), they could reflect meaningful validity differences that have implications for assessment and intervention (Anastasi & Urbina, 1997), especially in clinical populations.

However, to examine if Dana and Dawes' (this issue) contention is indeed the case for children with learning disabilities (LD; $N = 148$) in the Wechsler Intelligence Scale for Children-IV (WISC-IV)/Wechsler Individual Achievement Test-II (WIAT-II) sample, we used the same principal factor axis factoring method with oblimin rotation as described in Wechsler (2003). As can be seen in Table 1, the results support the four factor VC, PR, WM, PS solution, extracted after 49 iterations (Eigenvalues 3.34, 1.41, 1.12, 1.00), and accounting for 69% of the variance. This is contrasted with a forced single factor solution, which accounted for only 33% of the variance. Although this four factor solution is similar to the one reported in Wechsler (2003), factor loadings are

Table 1. *Comparison of Exploratory Principal Axis Factoring with Direct Oblimin (Oblique) Rotation with Forced Single Factor Solution for Learning Disability Sample*

	VC	PR	WM	PS	Single Factor ("*g*")
Similarities	.81				.70
Vocabulary	.79				.65
Comprehension	.59				.55
Block Design		−.41			.49
Picture Concepts		−.46			.50
Matrix Reasoning		−.95			.52
Digit Span			.68		.42
Letter-Number Sequencing			.38		.50
Coding				.74	.38
Symbol Search				.71	.34

Note. VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Processing Speed.

not as strong, and the Perceptual Reasoning factor showed negative loadings, and negative correlations (range r −.32 to −.43) with other factors, which could reflect hemispheric processing differences (e.g., Riccio, 1998) and explain why children with LD have developmental deficits, not delays (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996) which are amenable to differential intervention strategies (e.g., Hale & Fiorello, 2004).

These factor analytic findings, when combined with examination of the zero-order correlations among subtests (r range .06 to .64), clearly show that the "positive manifold" among subtests is quite limited for children with LD, and that these subtests are hardly "indistinguishable" from each other, as Dana & Dawes (this issue) suggest. In our original paper we showed that there is the requisite specificity necessary for factor-level interpretation by showing extensive unique variance and limited shared variance among the four WISC-IV factors. Maybe these results do not hold "in general," as Dana and Dawes acknowledge, but they are clear for children with LD, ADHD, and TBI. As we note in Fiorello et al. (this issue), and demonstrate in this rebuttal piece, generalizing findings from typical samples to clinical samples is problematic, as validity can be affected when different homogeneous groups are collapsed to form a single large heterogeneous group (Anastasi & Urbina, 1997). However, even for typical children, the factor analytic findings suggest a four factor solution is preferable to a single factor solution (Wechsler, 2003), and when a CHC approach to these data is used, a five factor solution (crystallized, fluid, visual, short-term memory, and processing speed) is obtained (Keith, Fine, Taub, Reynolds, & Kranzler, 2006).

Finally, Dana and Dawes (this issue), are concerned about our truncating global Full Scale IQ (FSIQ) scores between 80 and 120. We have done this in previous studies as well (e.g., Fiorello et al., 2001; Hale et al., 2001), because most studies of children with LD exclude extreme scores to ensure that their samples represent children with ‘‘average’’ intelligence, not those who may be mentally retarded or gifted. Dana and Dawes are correct in noting that truncating global FSIQ does reduce shared variance among the factors, but their findings based on their own simulation data, and our own analysis of real children with LD (see Table 2), do not support their conclusions. Including all FSIQ scores does increase the shared factor variance, but as can be seen in Table 2, the amount is relatively small (7.1%), with 93% of FSIQ variance accounted for by unique (42.2%), two-factor commonalities (31.8%), and three-factor commonalities (7.1%). Certainly, the 7.1% figure found for *real* children with LD has more credibility than the 22.5% reported for the Dana and Dawes *simulation* sample. In addition, even with the Dana and Dawes simulation data, there are appreciable amounts of unique and common variance accounted for by their hypothetical factors below the four-way commonality. Therefore, in analyses using real children and simulation data (which are questionable given the results with real children), we believe there is sufficient evidence that additional information gained through factor level interpretation is warranted and necessary.

Table 2. *WISC-IV FSIQ Commonality Analysis for Total LD Group*

	VC	PR	WM	PS
U_{VC}	.119			
U_{PR}		.123		
U_{WM}			.074	
U_{PS}				.106
C_{VCPR}	.103	.103		
C_{VCWM}	.075		.075	
C_{PRWM}		.040	.040	
C_{PRPS}		.064		.064
C_{WMPS}			.027	.027
C_{VCPRWM}	.098	.098	.098	
C_{VCPRPS}	.038	.038		.038
C_{VCWMPS}	.017		.017	.017
C_{PRWMPS}		.032	.032	.032
$C_{VCPRWMPS}(g)$	.071	.071	.071	.071
Total Explained	.530	.569	.434	.364
Unique Variance	.119	.123	.074	.106
Shared Variance	.411	.446	.360	.258

Note. Commonalities $<.01$ not reported. U = unique variance; C = shared variance; VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Processing Speed.

REPLY TO WATKINS, GLUTTING, AND LEI: MULTICOLLINEARITY IS A SIGNIFICANT PROBLEM

In their treatise of our work, Watkins, Glutting, and Lei (this issue) present hierarchical regression analyses that essentially reflect the Potthoff technique to determine if test bias is present (e.g., DeShon & Alexander, 1996). If the categorical group they constructed (i.e., flat vs. variable profile) or the group by continuous variable (i.e., FSIQ) interaction is significant in the prediction of the dependent variables (WIAT-II Reading and Math Composites), this suggests bias as the predictor does not uniformly predict the outcome (Anastasi & Urbina, 1997). Curiously, Watkins et al. do not call their analyses the Potthoff technique, which is typically used to determine bias for nominal groups such as gender or race. However, unlike the Dana and Dawes paper (this issue), this paper would appear to be strengthened by the use of large representative samples, with much detail offered in an attempt to make the paper seem methodologically sound. At first glance, the reader may be surprised that Watkins et al. (this issue) compare children with flat vs. variable profiles, instead of comparing typical children to those with disabilities, for disability status was the focus of the Fiorello et al. (this issue) analyses. Perhaps they are merely addressing our earlier commonality findings (Fiorello et al., 2001; Hale et al., 2001) of little shared factor variance for children with variable profiles and those with disabilities in the construction of FSIQ. Certainly, our findings cast doubt regarding the utility of FSIQ interpretation for a vast majority of children with disabilities, so addressing the findings in the literature would be of paramount importance. However, and more importantly, perhaps they constructed their design and analyses to ensure the findings supported their position. As we will demonstrate in the following tables and figures, the methodology and findings presented by this academic group are suspect, and their conclusions unwarranted.

Before examining the problems associated with the Watkins et al. (this issue) analyses, we will first examine their support of hierarchical regression over our use of regression commonality analysis in the construction of FSIQ and prediction of achievement outcomes. Their creative use of quotations and references may be convincing to some readers, yet misinformed. Most statisticians now recognize the old adage "correlation does not equal causation" is, according to Cohen and Cohen (1983) "…well intentioned, but grossly misleading. Causation manifests itself in correlation, and its analysis can only proceed through the systematic analysis of correlation and regression" (p. 15). This is especially true if regression findings are replicated using similar designs, analyses, and independent samples (e.g., Draper & Smith, 1998; Johnson, 2001), as we have done in Fiorello et al. (this issue; see also Fiorello et al., 2001) and Hale et al. (2001). The replicated results suggest explanatory conclusions may be warranted, as is the case with the use of partial correlations when determining simultaneous linear regression equations in structural equation modeling (Anastasi & Urbina, 1997), yet statistical debate persists about whether these models are indeed "causal" (e.g., Bentler, 1988; Pearl & Verma, 1991; Freedman, 1997). Despite the semantic concerns of Watkins et al., our replication of findings across instruments and samples with independent data sets provides the critical evidence necessary to re-examine previously held beliefs about the construct validity of these measures (e.g., Cronbach & Meehl, 1955), especially the belief that global IQ is the only score worth examining on the WISC-IV.

It is our view that Pedhazur's (1997) comments regarding the use of commonality analysis for explanatory purposes is based on two premises ignored by Watkins et al. (this issue). First, Pedhazur (1997) argued that large numbers of independent variables used in single study regression equations will result in many higher-order commonalities that lead to interpretation difficulties. Clearly, this is not the case in Fiorello et al. (this issue, 2001) and Hale et al. (2001), where the common components have been interpreted and replicated across samples and studies. Second, Pedhazur stated that negative commonalities, or suppressor effects, are common when large numbers of independent variables are used in single study regression equations. As we will show later, suppressor effects occur when we collapse disparate subtest scores into global ones, and this is especially true for children with LD, where almost half (41%) of the reading variance explained is lost when we interpret global IQ as compared to component scores. Additionally, Hale et al. (2001) used commonality analysis to predict achievement outcomes, which is preferred over hierarchical regression when multicollinearity of predictors is present (Pedhazur, 1997). The positive manifold among subtests (e.g., Kranzler, 1997), which is so often used as a rationale to support global IQ interpretation, suggests multicollinearity among the subtests, and argues against the use of hierarchical regression.

Watkins et al. (this issue) suggest hierarchical regression is preferred over commonality analysis, but Hale et al. (2001) show these "incremental validity" studies are clearly misleading because the predictors are highly collinear—they are made of the same variance. The same subtests are used to compute the factor scores and the FSIQ score, so if you reverse the order of independent variable entry (i.e., factors first, FSIQ last), the opposite results are found. A simple demonstration of this significant statistical problem is presented in Table 3 using the WISC-IV/WIAT-II standardization data (Wechsler, 2003). In the first example, which reflects the techniques used by Glutting et al. (1997), the FSIQ is entered first in the prediction of the Reading Composite, accounting for 53% of the variance, with little remaining variance left for the factors (2%). But the opposite is true if you enter the factors first, as they account for 55% of the Reading Composite variance, followed by FSIQ, which accounts for virtually no (.2%) variance. You can't have completely *opposite* results with the *same* data set and statistical analysis! Although beta weights are commonly reported in multiple regression (Cohen & Cohen, 1983), these values are typically ignored in the "incremental validity" studies (e.g., Glutting et al., 1997), and their instability clearly reflects significant multicollinearity (e.g., Pedhazur, 1997). It is clear that commonality analysis is the correct method for predicting dependent variables—not hierarchical regression—when predictors are highly collinear (Pedhazur, 1997). If these authors wish to cite these dubious "incremental validity" papers, they should report the statistical fact that order of independent variable entry determines whether the FSIQ means *everything* (FSIQ entered first, followed by factors/subtests/profiles) or *nothing* (factors/subtests/profiles entered first, followed by FSIQ).

Table 3. *Hierarchical Regression Results for WIAT-II Reading Composite*

Step/Predictors	Beta	t	p	R^2 Equation	Change R^2
			FSIQ First Step/Factors Second Step		
Step 1. Enter FSIQ				.528	—
FSIQ	.727	31.85	<.001		
Step 2. Enter VC, PR, WM, PS Factors				.547	.019
FSIQ	.792	1.78	.075		
VC	.091	.54	.591		
PR	−.160	−1.01	.315		
WM	.042	.34	.734		
PS	−.057	−.06	.644		
			Factors First Step/FSIQ Second Step		
Step 1. Enter VC, PR, WM, PS Factors				.546	—
VC	.389	13.56	<.001		
PR	.119	4.07	<.001		
WM	.253	8.67	<.001		
PS	.158	5.89	<.001		
Step 2. Enter FSIQ				.547	.002
VC	.091	.54	.591		
PR	−.160	−1.01	.315		
WM	.042	.34	.734		
PS	−.057	−.06	.644		
FSIQ	.792	1.78	.075		

Note. FSIQ = Full Scale Intelligence Quotient; VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Processing Speed.

The present Watkins et al. (this issue) study is also fraught with methodological problems, and careful analysis of their design and statistics demonstrates the limitations of their findings. First, it is unclear why Watkins et al. enter FSIQ into the regression equation that purports to examine whether a variable versus a flat profile differentially predicts academic performance, and why they did not focus on children with disabilities—the focus of the Fiorello et al. (this issue) paper. The answers in part can be found in Tables 4 and 5. As can be seen in Table 4, the most variance in the prediction of achievement domains is accounted for by the subtests, followed by the factors, and the least amount of variance is accounted for by FSIQ, with almost half (41%) of the WIAT-II Reading Composite variance lost when one uses FSIQ as the predictor as compared to subtests. This difference is less for the flat and variable profile groups, but still consistent. If the 10 subtests are used to compute the VC, PR, WM, and PS Indices, and the FSIQ, the reader should wonder why there is such a dramatic loss of achievement variance when different combinations of predictors are used. They should

Table 4. *WIAT-II Reading and Math Variance Accounted for by Different Predictor Combinations for Standardization and LD Samples*

	Standardization group						LD Group	
	Total		Flat profile		Variable profile			
	R^2	% Change	R^2	% Change	R^2	% Change	R^2	% Change
				Reading Composite				
10 Core Subtests	.57	—	.55	—	.58	—	.32	—
4 Factor Indexes	.55	−4%	.51	−7%	.56	−3%	.28	−13%
Full Scale IQ	.53	−7%	.50	−9%	.54	−7%	.19	−41%
				Mathematics Composite				
10 Core Subtests	.58	—	.61	—	.57	—	.46	—
4 Factor Indexes	.57	−2%	.58	−5%	.56	−2%	.42	−9%
Full Scale IQ	.56	−3%	.58	−5%	.56	−2%	.40	−13%

Table 5. *Total Sample, Flat Profile, and Variable Profile Descriptive Statistics*

	Matched			Standardization		
	Total	**Flat**	**Variable**	**Total**	**Flat**	**Variable**
FSIQ						
M	100.89	100.89	100.89	100.33	101.43	100.12
SD	12.89	12.91	12.91	14.25	13.54	14.38
VC						
M	100.54	100.42	100.65	99.59	100.35	99.44
SD	11.55	10.73	12.37	13.93	11.73	14.32
PR						
M	100.05	100.20	99.90	99.94	99.79	99.97
SD	12.09	10.05	13.87	13.70	10.93	14.24
WM						
M	99.56	100.12	99.00	99.53	100.22	93.38
SD	12.44	10.47	14.16	14.40	10.98	15.02
PS						
M	101.52	100.75	102.29	100.38	100.24	100.41
SD	12.97	10.80	14.84	14.49	11.08	15.11
READ						
M	100.20	102.96	97.64	99.47	102.42	98.87
SD	15.39	16.44	13.94	16.11	16.40	15.99
MATH						
M	102.00	102.14	101.86	100.70	101.36	100.56
SD	15.53	15.48	15.63	16.88	15.90	17.09

Note. FSIQ = Full Scale Intelligence Quotient; VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Proces-Processing Speed; READ = Reading Composite; MATH = Math Composite.

account for the same amount of variance. This is a clear sign that there are suppressor effects or negative commonalities when one looks at global scores as opposed to subtest scores, and this finding is dramatic for the LD group. Why is variance lost when the FSIQ is used? Simply put, if one variable is a positive predictor of an outcome, and the other a negative predictor, and they are combined into a single score, less dependent variable (e.g., reading achievement) variance will be accounted for by the global score, and in this case, for children with LD, the loss of achievement variance when FSIQ is used as the predictor is significant.

Another methodological issue has to do with matching the groups on FSIQ, as can be seen in Table 5. Readers will notice that there is less variance in the matched sample than the total standardization sample, and less variance in the FSIQ as opposed to the four factor Indices and achievement composites (which is reflective of the profile variability). By restricting the range of the independent variable, Watkins et al. restrict the variance of corresponding values of a dependent variable (e.g., Chambers, 1986), in an effort to decrease power and increase likelihood of a Type II error. But more importantly, by matching on FSIQ, there is no difference between the flat and variable profile groups in the FSIQ predictor. There is no need to enter FSIQ into the regression equation since the effects of FSIQ are already accounted for, and hence Watkins et al. are double controlling for FSIQ when they enter it first into the regression equation. Since both groups have equal FSIQ, predictive differences between the groups are not attributable to FSIQ, and similar prediction equations will result, again decreasing power and increasingly the likelihood of a Type II error.

Another methodological issue is creating a categorical variable (flat vs. variable groups) from essentially a continuous one (profile variability), and it is well known that this particular practice reduces score variability. Watkins et al. (this issue) use this variable construction to reduce power and increase likelihood of a Type II error, but this is not uncommon practice for this academic group. For instance, Watkins and Canivez (2004) take a continuous variable (profile variability), turn it into a nominal one (agreement vs. disagreement), and then use a nonparametric statistic that is typically used to judge interobserver agreement, one that is controversial for even observations because it is overly conservative. Of course, they show that there is little "agreement" regarding profile variability over time, and conclude that profile analysis and subtest interpretation should be avoided. However, while the Watkins and Canivez (2004) design and analyses ensure reduced power and increased likelihood of a Type II error, they are not statistically flawed as is the case with the hierarchical regression approaches described earlier, or the present Watkins et al. (this issue) paper.

Although these problems (i.e., using FSIQ instead of factor/subtest predictors and the resultant loss of achievement variance; comparing flat and variable profile groups when disability is the focus; FSIQ matching to reduce variability and power; entering FSIQ into the regression equations after controlling for group differences; creating a categorical variable for continuous level data) might convince the reader of the impropriety of the Watkins et al. (this issue) results, probably the single most significant problem in their analysis can be found in the fact that the FSIQ and Index variability are not statistically independent—FSIQ is a linear function of the four factor Index scores (or even a quadratic or cubic function, depending on the sample). Not only is between-group variability limited by the use of FSIQ, but the design and analyses create variability in calculating the flat and variable profile group variable, and then remove it by using FSIQ as the predictor. This also explains why the interaction term is clearly inappropriate. An individual with a FSIQ of 40, obtained from a sum of scaled scores of 10, has a maximum discrepancy between any two indices of five (i.e., PS and WM would be identical in this case). The maximum degree of discrepancy obtained is also not the same for each pairwise comparison for a specific sum of squares total. For an individual with a FSIQ of 90, obtained from a sum of squares to 88, if VCI = 100, PRI = 86, and WMI = 94, then PSI has to be 88. Therefore, the variability between scores does not have local independence at the specific level of FSIQ.

Local independence is required for nearly all statistical procedures including regression techniques. The Potthoff technique is appropriate for comparing nominal group differences suggestive of bias (DeShon & Alexander, 1996), not for groups created and destroyed using the same variance, where the interaction will surely eliminate the possibility the null will be rejected. Including FSIQ and either the four Index scores or indication of variability within Indexes, renders the statistical tests invalid—a more serious violation of regression assumptions than even the obvious communality issues. An explicit demonstration of these problems can be found in Tables 6 and 7, which examines whether bias is found for the FSIQ in the prediction of achievement domains, even after matching on FSIQ. Note how the group effect is significant for the Reading Composite for flat verses variable profiles (N = 236), and for the Reading Composite and Math Composite for a comparison of the

Table 6. *Hierarchical Regression for Matched Flat and Variable Profile Standardization Groups*

Step/Predictors	Beta	t	p	R^2 Equation
		WIAT-II Reading Composite		
Step 1				.491
FSIQ	.679	14.55	<.001	
Significance Group	−.183	−3.91	<.001	
Step 2				.493
FSIQ	.726	11.06	<.001	
Significance Group	.185	.51	.614	
FSIQ*Significance Group	−.374	−1.01	.313	
		WIAT-II Math Composite		
Step 1				.483
Full Scale IQ	.695	15.18	<.001	
Significance Group	−.009	−.19	.843	
Step 2				.486
FSIQ	.751	11.61	<.001	
Significance Group	.429	1.19	.236	
FSIQ*Significance Group	−.134	−1.22	.223	

Note. FSIQ = Full Scale Intelligence Quotient.

Table 7. *Hierarchical Regression for Matched Standardization and Learning Disabled Groups*

Step/Predictors	Beta	*t*	*p*	R^2 Equation
		WIAT-II Reading Composite		
Step 1				.288
FSIQ	.400	5.26	< .001	
SPED Group	−.347	−4.55	< .001	
Step 2				.288
FSIQ	.426	3.84	< .001	
SPED Group	−.118	−.17	.869	
FSIQ*SPED Group	−.230	−.32	.750	
		WIAT-II Math Composite		
Step 1				.331
FSIQ	.528	7.16	< .001	
SPED Group	−.224	−3.04	.003	
Step 2				.333
FSIQ	.486	4.65	< .001	
SPED Group	−.602	−.89	.376	
FSIQ*SPED Group	.382	.56	.576	

Note. FSIQ = Full Scale Intelligence Quotient; SPED = Special Education.

standardization group vs. children with LD matched on FSIQ and profile variability ($N = 128$; *FSIQ* $M = 96.36$, $SD = 10.57$; Variable Profile $n = 51$ in each group). This significance is eliminated when the interaction term is entered at Step 2. Note the dramatic changes in beta weights for the grouping variable at Step 2, but the FSIQ beta weights are not as dramatically affected, and actually are strengthened by eliminating the variability inherent in the grouping variables and their interactions. Despite it being common to report beta weights in regression (Cohen & Cohen, 1983), Watkins et al. conveniently omit such information, as has been done in the "incremental validity" studies (e.g., Glutting et al., 1997), because it would clue readers in to their statistical trickery. Without the interaction term, these findings demonstrate that the more variable the profile, and the more a child has LD, the more biased FSIQ is in the prediction of reading achievement, and the more a child has a LD, the more biased FSIQ is in the prediction of math achievement as well (for further discussion of bias, see DeShon & Alexander, 1996).

Table 8. *Repeated Measures Results for Significance Group Status Matched on Full Scale IQ*

Source	*df*	*SS*	*MS*	*F*	*p*	Power
Repeated	1.89	381.05	200.86	2.62	.077	.506
Repeat*SIG	1.89	1068.09	563.02	7.34	.001	.928
Error	443.91	34058.61	76.72			

Note. SIG = Significance Group (Flat Profile vs. Variable Profile).

One final problem to address in the Watkins et al. (this issue) paper has to do with the use of multiple regression instead of the more appropriate repeated measures general linear model (GLM) approach, because these children were administered the WISC-IV FSIQ and the WIAT-II Reading and Math Composites at the same time. Examination of concurrent—not predictive—validity should occur, so the GLM procedure is the appropriate analysis, with results depicted in Tables 8 and 9, and Figures 1 and 2. As can be seen, there are interaction effects for both the flat vs. variable group, and the matched typical vs. LD group comparisons, revealing FSIQ bias for both analyses. Post-hoc ANOVAs revealed that while there was a significant difference ($F(1,235) = 7.76$, $p = .008$) between flat and variable profiles for the Reading Composite (see Figure 1), there was no Math Composite difference. However, when comparing typical children and those with LD, matched for FSIQ and profile variability, there was a significant difference for both the Reading Composite ($F(1,124) = 18.14$, $p < .001$) and the Math

Table 9. *Repeated Measures Results for Special Education Status Matched on Full Scale IQ*

Source	*df*	*SS*	*MS*	*F*	*p*	Power
Repeat	1.96	2764.94	1414.63	17.95	<.001	1.000
Repeat* SPED	1.96	1193.94	610.86	7.75	.001	.945
Error	238.45	18792.69	78.81			

Note. SPED = Special Education Group (Typical vs. Learning Disabled).

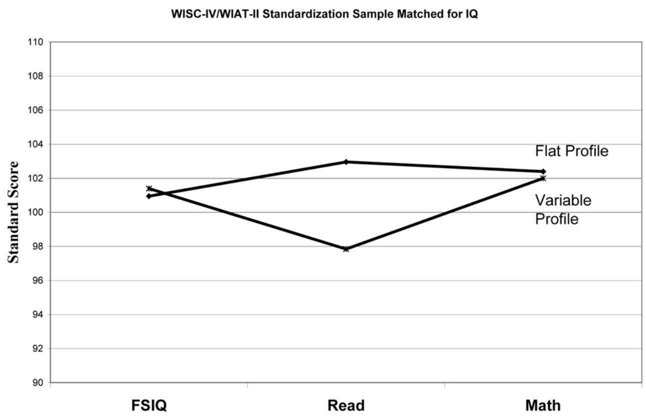

Figure 1. Interaction of repeated measures by significance group.

Composite ($F(1,124) = 6.86$, $p = .010$), confirming the regression findings presented earlier that FSIQ does not equally predict achievement across samples. These results provide little support for Watkins et al. (this issue) "unambiguous" contention that the FSIQ was a "robust predictor of academic achievement regardless of significant score variability for participants with and without disabilities" or that FSIQ is the better predictor of achievement than other derived intelligence test scores, especially given the results reported in Table 4. While FSIQ may be a "parsimonious" predictor of achievement, it is neither the most effective predictor derived from intellectual tests, nor is it unbiased with regard to profile variability and disability status.

REPLY TO FAUST: SOME SPECIFIC THOUGHTS ON GLOBAL INTELLIGENCE

Faust (this issue) discusses a number of issues that warrant attention and several points that support our position. Faust notes that science progresses if "contrary evidence that ought to count" is examined, and through the clinical and empirical evidence presented in this special issue, we have provided the impetus for progress in intelligence test interpretation. Much evidence exists to support "g," yet one can argue that its existence is largely irrelevant (e.g., Horn, 2007). Surely FSIQ, however it is constructed, will continue to predict outcomes. As Faust points out, the issue is not one of clinical groups, or variable or flat profiles as Watkins et al. suggest, but rather it is how variable performance manifests itself in the child's daily functioning, and whether this affects intervention choice and efficacy. Variability is not an answer - it merely leads to additional questions the practitioner must address in clinical practice.

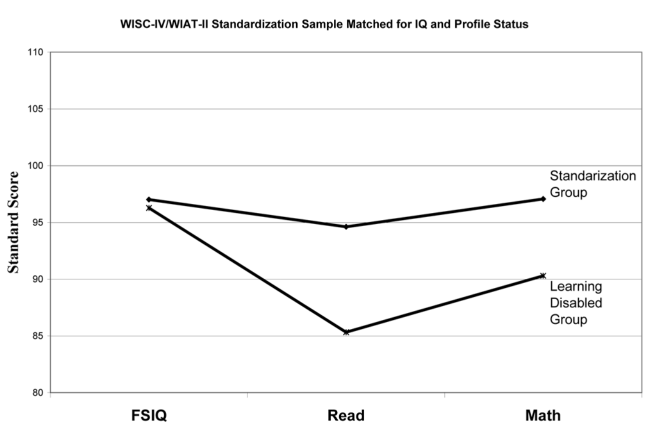

Figure 2. Interaction of repeated measures by special education group.

In contrast to traditional psychometric approaches that focus on end products, examining a child's performance in terms of specific cognitive processes not only provides necessary insight into what the test actually measures (i.e., construct validity), but this information can be used to tailor interventions as well (Anastasi & Urbina, 1997). Take three children who obtain a mildly deficient FSIQ of 69. The first child has a VC of 96, a PR of 61, a WM of 80, and a PS of 56. The second child has a VC of 57, a PR of 104, a WM of 56, and a PS of 73. For the third child, a VC of 65, a PR of 69, a WM of 80, and a PS of 88 are obtained. If FSIQ is the only score worth interpreting, all three children would be classified as mildly deficient with concomitant academic delays. What if you knew the first child was the son of highly educated parents and was an A student until he had a TBI that resulted in multiple social and academic problems, with his grade point subsequently dropping to a low C– average? What if you knew the second child lived in a Spanish-speaking home, and preferred to speak Spanish with adults and peers, but was required to speak English at school? Would it help to know that the third child was diagnosed with fetal alcohol syndrome at birth and received intensive early intervention services, but he made few academic gains? According to the global IQ-only perspective, all three would be placed in an educable mentally retarded program with the same curriculum and interventions because the same level of academic impairment would be predicted.

Faust (this issue) argues that parsimonious actuarial models often work as well as or better than more complex models if one is making dichotomous decisions where the criterion can be accurately identified. Although this would be relevant for broad classification purposes (e.g., special education eligibility), clinical decisions are not (or at least should not be) straightforward dichotomies, so simple actuarial models are likely to perform poorly in clinical practice. When a criterion is poorly defined or differentiated depending upon criterion specificity (e.g., learning disabilities), actuarial models cannot yield better prediction of outcomes than the accuracy/agreement that defines the criterion, which is a chief obstacle in accurate decision-making (Anastasi & Urbina, 1997). Focusing on global IQ to calculate ability-achievement discrepancies for LD identification has received widespread criticism (see Fuchs, Mock, Morgan, & Young, 2003; Hale, Kaufman, Naglieri, & Kavale, 2006) from those who support intellectual and neuropsychological assessment and those who abhor it, and the response-to-intervention model that is offered in replacement is also fraught with methodological problems (see Fuchs et al., 2003; Kavale, Kaufman, Naglieri, & Hale, 2005). Diagnosis of children with LD requires determination of a deficit in the basic psychological processes in the presence of cognitive integrities (Hale et al., 2006; Kavale et al., 2005). This, in essence, is the definition of specific learning disabilities under the Individuals with Disabilities. Improvement Act of 2004 (34 C.F.R. 300.7). The focus should be on cognitive and neuropsychological processes, not the content domains typically identified through factor analysis (e.g., Anastasi & Urbina, 1997), if we are to correctly identify children with LD. As a result, global FSIQ does not address the definition of LD (Hale et al., 2006), and ability-achievement discrepancy does little to inform intervention (Fletcher et al., 2002). We should heed Faust's recommendation to avoid spending "inordinate time...attending to variables that provide little or no help in predicting which intervention will yield the best outcome," and avoid global IQ overemphasis in psychological evaluations. Instead, researchers and practitioners should focus on how the extensive cognitive and/or neuropsychological literature can be used to guide our interpretation of standardized tests, and how this information can become meaningful for intervention (Hale & Fiorello, 2004).

We have never claimed that intellectual assessment is irrelevant; in fact, we have vociferously argued the opposite position (e.g., Fiorello et al., 2001; Fiorello et al., 2006; Hale & Fiorello, 2004; Hale & Fiorello, 2001; Hale et al., 2005; Hale et al., 2004; Hale et al., 2003; Hale et al., 2001; Hale et al., 2006; Kavale et al., 2005). We have not argued that global IQ is meaningless, as it surely represents the summative average of many disparate cognitive functions (e.g., Lezak, 1995; Luria, 1979). Certainly, FSIQ can be helpful in identifying children with mental retardation and those who are gifted, provided that their level and pattern of intellectual performance are examined within the context of other diagnostic information. We have not argued that variability among indexes is diagnostic of any particular disorder (although see Calhoun & Mayes, 2005; Mayes & Calhoun, 2006; Mayes & Calhoun, 2004). When score variability or disability are present, our findings suggest the construct validity of FSIQ is diminished, as a global score can only reflect the measurement integrity of subcomponent tasks used to derive it (e.g., Anastasi & Urbina, 1997). However, rather than focus on FSIQ construct validity, variability begs the question as to what factors contribute to inconsistent performance, and how this variability is useful in making classification and intervention recommendations. This is the challenge we seek and Faust lauds. It is also important to note that we do not support subtest analysis in its traditional sense, as was suggested by Dana and Dawes (this issue). It would be nice if all children would perform comparably on all intellectual subtests—then we could interpret FSIQ. Or if they at least performed comparably on the WISC-IV VC, PR, WM, and PS Indices, then we could interpret the factors. If scores are consistent, then global score interpretation makes sense, because the results are consistent across intellectual domains. However, reliability found in large standardization samples for measures should not be confused with intrapersonal performance consistency. Some children do not perform equally well on subcomponent measures (i.e., subtests), and this variability can be due to a variety of factors (see Sattler, 2001), which we believe suggests there may be utility in examining intellectual functions below the global IQ. Notice we say that idiographic analysis *may*—not *does*—have utility, because ample evidence suggests idiographic interpretation can be problematic because subtest scores are by definition less reliable, at least in typical populations (e.g., McDermott et al., 1990).

We concur with several authors, including those in this special issue, that subtest profile analysis is problematic unless concurrent, ecological, and treatment validity are established (Hale & Fiorello, 2004). The clinical need for objective confirmation of a child's purported cognitive strengths and deficits is why we developed the Cognitive Hypothesis Testing model (CHT; Hale & Fiorello, 2004), which uses the latest neuropsychological theory and empirical findings to guide test interpretation. The CHT model is designed to help practitioners validate or refute profile findings, interpret profiles within the context of the child's natural environment, and ensure that profiles are considered tentative until additional critical validity evidence is obtained. We see the intellectual/cognitive measure as a screening tool for developing hypotheses about cognitive strengths and deficits. If these hypotheses are substantiated with additional evidence, and demonstrate treatment validity, then the evidence Faust requires would be documented, and the complex model subsequently explored would be deemed worthy of the clinician's time and effort.

As Faust (this issue) suggests, complex models would be warranted if evidence emerged that there are different disorder subtypes, and they respond differentially to interventions. As can be seen in Table 10, there is a great deal of shared factor variance in the prediction of reading and math domains for those with flat profiles, but little for children with variable profiles and those with LD, where unique and lower level commonalities become prevalent. For the reading (see Table 11) and math (see Table 12) commonality analysis for the entire WISC-IV/WIAT-II (Wechsler, 2003) LD sample (to address Dana & Dawes', this issue, concerns about FSIQ truncation), there appears to be sufficient specificity for interpretation of the factor Indices, with very little shared variance among all four factors (less than 5%) in the prediction of academic domains. In addition, there are differences in the factor predictors between achievement domains, such as the need for more Perceptual Reasoning or novel problem solving skills during mathematics performance.

These commonality findings demonstrate differential relationships between factors and outcomes for children with LD, a key element of establishing factorial construct validity (Anastasi & Urbina, 1997). As Cronbach and Meehl (1955) admonished researchers long ago, previously held convictions about a test's construct validity must be altered to account for new disconfirming evidence. We now have considerable evidence (e.g., Fiorello et al., 2001; Georgas et al., 2003; Hale et al., 2001; Keith et al., 2006; Wechsler, 2003) that a multifactorial model better represents the WISC-IV standardization and especially the clinical group data than a single factor model, which certainly can be found (e.g., Horn, 2007). If this clinically relevant information is examined within the context of a comprehensive evaluation of cognitive and neuropsychological processes, informed academic and psychosocial interventions will result, especially if findings are validated using single subject designs to document treatment efficacy (e.g., Hale & Fiorello, 2004). FSIQ does little to enhance intervention development, because it is merely a composite of these disparate predictors (e.g., Kaplan, 1988; Lezak, 1988; Luria, 1979). One

Table 10. *WIAT-II Reading and Math Commonality Analysis for Flat Profile ($n = 122$), variable profile ($n = 657$), and LD ($n = 148$) groups*

	Standardization Group				LD Group	
	Flat Profile		Variable Profile			
	R^2	% Total	R^2	% Total	R^2	% Total
	Reading Composite					
Unique	.01	2%	.18	32%	.20	69%
Two-Way Commonality	.04	8%	.16	28%	.05	17%
Three-Way Commonality	.07	14%	.14	25%	.03	10%
Four-Way Commonality (*g*)	.39	76%	.09	16%	.01	3%
	Mathematics Composite					
Unique	.02	3%	.17	30%	.22	52%
Two-Way Commonality	.04	7%	.15	27%	.12	29%
Three-Way Commonality	.08	14%	.14	25%	.06	14%
Four-Way Commonality (*g*)	.45	76%	.10	18%	.02	5%

Table 11. *WIAT-II Reading Composite Commonality Analysis for Total LD Group*

	VC	PR	WM	PS
U_{VC}	.153			
U_{PR}		.006		
U_{WM}			.012	
U_{PS}				.029
C_{VCWM}	.053		.053	
C_{VCPRWM}	.011	.011	.011	
C_{VCPRPS}	.011	.011		.011
$C_{VCPRWMPS}$ (g)	.012	.012	.012	.012
Total Explained	.241	.030	.091	.054
Unique Variance	.153	.006	.012	.029
Shared Variance	.088	.024	.079	.025

Note. Commonalities < .01 not reported. U = Unique Variance; C = Shared Variance; VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Processing Speed.

should never accept the null hypothesis about cognitive-intervention associations, as Faust (this issue) suggests we do; rather, we must continue to systematically explore cognitive and neuropsychological models using single subject designs to determine if these data can inform intervention efforts (e.g., Braden & Kratochwill, 1997). Since neuropsychological deficits—not delays—are likely to cause many disorders such as LD and ADHD (e.g., Francis et al., 1996; Castellanos et al., 2002), and certainly TBI, with hundreds of cognitive, neuropsychological, and neuroimaging studies demonstrating that psychological processes associated with achievement competency and psychosocial functioning can inform intervention (see Berninger, 2002; Hale & Fiorello, 2004; Mather & Jaffe, 2002; Naglieri & Pickering, 2003; Naglieri, 2003; Semrud-Clikeman, 2005), it makes more sense to begin exploration of more complex models—models that will be absent in research and practice if global IQ is the only relevant metric in the field of applied neuropsychology.

Table 12. *WIAT-II Math Composite Commonality Analysis for Total LD Group*

	VC	PR	WM	PS
U_{VC}	.123			
U_{PR}		.038		
U_{WM}			.011	
U_{PS}				.043
C_{VCPR}	.053	.053		
C_{VCWM}	.016		.016	
C_{PRPS}		.024		.024
C_{WMPS}			.013	.013
C_{VCPRWM}	.026	.026	.026	
C_{VCPRPS}	.022	.022		.022
C_{PRWMPS}		.010	.010	.010
$C_{VCPRWMPS}$ (g)	.024	.024	.024	.024
Total Explained	.274	.202	.108	.146
Unique Variance	.123	.038	.011	.043
Shared Variance	.151	.164	.097	.103

Note. Commonalities <.01 not reported. U = Unique Variance; C = Shared Variance; VC = Verbal Comprehension; PR = Perceptual Reasoning; WM = Working Memory; PS = Processing Speed.

REFERENCES

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Bentler, P. M. (1988). Causal modeling via structural equation systems. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed.). (pp. 317–335). New York: Plenum.

Berninger, V. (2002). *Brain literacy for educators and psychologists*. New York: Academic Press.

Braden, J. P., & Kratochwill, T. R. (1997). Treatment utility of assessment: Myths and realitites. *School Psychology Review*, *26*, 475–485.

Calhoun, S. L., & Mayes, S. D. (2005). Processing speed in children with clinical disorders. *Psychology in the Schools*, *42*, 333–343.

Carroll, J. B. (1997). Psychometrics, intelligence, and public perception. *Intelligence*, *24*, 25–52.

Castellanos, F. X., Lee, P. P., Sharp, W., Jeffries, N. O., Greenstein, D. K., Clasen, L. S., et al. (2002). Developmental trajectories of brain volume abnormalities in children and adolescents with attention-deficit/hyperactivity disorder. *Journal of the American Medical Association*, *288*, 1740–1748.

Chambers, R. L. (1986). Outlier robust finite population estimation. *Journal of the American Statistical Association*, *81*, 1063–1069.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

Dana J., & Dawes, R. (2007). Comment on Fiorello et al. "Interpreting intelligence test results for children with disabilities: Is global intelligence relevant?" *Applied Neuropsychology*, *14*, 21–25.

DeShon, R. P., & Alexander, R. A. (1996). Alternative procedures for testing regression slope homogeneity when group error variances are unequal. *Psychological Methods*, *1*, 261–277.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: Wiley-Interscience.

Faust, D. (2007). Some global and specific thoughts about some global and specific issues. *Applied Neuropsychology*, *14*, 26–36.

Fiorello, C. A., Hale, J. B., McGrath, M., Ryan, K., & Quinn, S. (2001). IQ interpretation for children with flat and variable test profiles. *Learning and Individual Differences*, *13*, 115–125.

Fiorello, C. A., Hale, J. B., & Snyder, L. E. (2006). Cognitive hypothesis testing and response to intervention for children with reading problems. *Psychology in the Schools*, *43*, 835–853.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary interpretive approach*. Boston, MA: Allyn & Bacon.

Fletcher, J. M., Lyon, G. R., Barnes, M., Stuebing, K. K., Francis, D. J., Olson, R. K., et al. (2002). Classification of learning disabilities: An evidence-based evaluation. In R. Bradley, I. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 185–250). Mahwah, NJ: Lawrence Erlbaum.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental delay versus deficit models of reading disability: A longitudinal, individual growth curve analysis. *Journal of Educational Psychology*, *88*, 3–17.

Freedman, D. A. (1997). From association to causation via regression. *Advances in applied mathematics*, *18*, 59–110.

Fuchs, D., Mock, D., Morgan, P., & Young, C. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications fro the learning Disabilities Construct. *Learning Disabilities Research & Practice*, *18*, 157–171.

Georgas, J., Weiss, L. G., van de Vijer, F. J. R., & Saklofske, D. H. (2003). *Culture and children's intelligence*. San Francisco, CA: Academic Press.

Glutting, J. J., Youngstrom, E. A., Ward, T., Ward, S., & Hale, R. L. (1997). Incremental efficacy of WISC-III factor scores in predicting achievement: What do they tell us? *Psychological Assessment*, *9*, 295–301.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*, 79–132.

Hale, J. B., & Fiorello, C. A. (2001). Beyond the academic rhetoric of g: Intelligence testing guidelines for practitioners. *The School Psychologist*, *55*(4), 113–139.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. New York: Guilford Press.

Hale, J. B., Fiorello, C. A., & Brown, L. (2005). Determining medication treatment effects using teacher ratings and classroom observations of children with ADHD: Does neuropsychological impairment matter? *Educational and Child Psychology*, *22*, 39–61.

Hale, J. B., Fiorello, C. A., Bertin, M., & Sherman, R. (2003). Predicting math competency through neuropsychological interpretation of WISC-III variance components. *Journal of Psychoeducational Assessment*, *21*, 358–380.

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Hoeppner, J. B., & Gaither, R. A. (2001). WISC-III predictors of academic achievement for children with learning disabilities: Are global and factor scores comparable? *School Psychology Quarterly*, *16*, 31–55.

Hale, J. B., Hoeppner, J. B., & Fiorello, C. A. (2002). Analyzing Digit Span components for assessment of attention processes. *Journal of Psychoeducational Assessment*, *20*, 128–143.

Hale, J. B., Kaufman, A. S., Naglieri, J. A., & Kavale, K. A. (2006). Implementation of IDEA: Integrating response to intervention and cognitive assessment methods. *Psychology in the Schools*, *43*, 753–770.

Hale, J. B., Naglieri, J. A., Kaufman, A. S., & Kavale, K. A. (2004). Specific learning disability classification in the new Individuals with Disabilities Education Act: The danger of good ideas. *The School Psychologist*, *58*, 6–13.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Horn, J. L. (2007). Understanding human intelligence: Where have we come since Spearman? In R. Cudeck & R. MacCallum (Eds.), *Factor analysis at 100. Historical developments and future directions.* Florence, KY: Taylor & Francis.

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

Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher*, *30*, 3–13.

Kaplan, E. (1988). A process approach to neuropsychological assessment. In T. Boll & B. K. Bryant (Eds.), *Clinical neuropsychology and brain function: Research, measurement, and practice* (pp. 129–231). Washington, DC: American Psychological Association.

Kavale, K. A., Kaufman, A., Naglieri, J., & Hale, J. (2005) Changing procedures for identifying learning disabilities: The danger of poorly supported ideas. *The School Psychologist*, *59*, 15–25.

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

Kranzler, J. H. (1997). Educational and policy issues related to the use and interpretation of intelligence tests in the schools. *School Psychology Review*, *26*, 150–162.

Lezak, M. D. (1988). IQ:RIP. *Journal of Clinical and Experimental Neuropsychology*, *10*, 351–361.

Luria, A. R. (1973). *The working brain*. Baltimore, MD: Penguin Books.

Luria, A. R. (1979). *The making of mind: A personal account of Soviet psychology*. Cambridge, MA: Harvard University Press.

Mather, N., & Jaffe, L. E. (2002). *Woodcock Johnson-III: Reports, recommendations, and strategies*. New York: Wiley.

Mayes, S. D., & Calhoun, S. L. (2004). Similarities and differences in WISC-III profiles: Support for subtest analysis in clinical referrals. *Clinical Neuropsychologist*, *18*, 559–572.

Mayes, S. D., & Calhoun, S. L. (2006). WISC-IV and WISC-III profiles in children with ADHD. *Journal of Attention Disorders*, *9*, 486–493.

McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment*, *8*, 290–302.

Naglieri, J. A. (2003). Current advances in assessment and intervention for children with learning disabilities. In T. E. Scruggs and M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities*

Volume 16: Identification and assessment (pp. 163–190). New York: JAI.

Naglieri, J. A., & Pickering, E. (2003). *Helping children learn: Intervention handouts for use in school and at home*. Baltimore, MD: Brookes.

Pearl, J., & Verma, T. (1991). A theory of inferred causation. In J. A. Allen, R. Fikes, & E. Sandewall (Eds.), Principles of knowledge representation and reasoning: Proceedings of the second international conference (pp. 441–452). San Mateo, CA: Morgan-Kaufmann.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Orlando, FL: Holt, Rinehart & Winston.

Pfeiffer, S. I., Reddy, L. A., Kletzel, J. E., Schmelzer, E. R., & Boyer, L. M. (2000). The practitioner's view of IQ testing and profile analysis. *School Psychology Quarterly*, *15*, 376–385.

Riccio, C. A. (1998, August). *There measurable biological substrates to Verbal-Performance differences in Wechsler scores*. Paper presented at the American Psychological Association Annual Convention, San Francisco, CA.

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive Ability. *Psychology, Public Policy, and Law*, *11*, 235–294.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications*. San Diego, CA: Author.

Semrud-Clikeman, M. (2005). Neuropsychological aspects for evaluating learning disabilities. *Journal of Learning Disabilities*, *38*, 563–568.

Watkins, M. W., & Canivez, G. L. (2004). Temporal stability of WISC-III subtest composite: Strengths and weaknesses. *Psychological Assessment*, *16*, 133–138.

Watkins, M. W., Glutting, J. J., & Lei, P.-W. (2007). Validity of the full-scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology*, *14*, 13–20.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—fourth edition (WISC-IV) manual*. San Antonio, TX: The Psychological Corporation.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson Psycho-Educational Battery Third Edition (WJ III)*. Itasca, IL: Riverside.