Distinctions Without a Difference:

The Utility of Observed Versus Latent Factors From the WISC-IV in Estimating Reading and Math Achievement on the WIAT-II

Joseph J. Glutting
University of Delaware
Marley W. Watkins
Pennsylvania State University
Timothy R. Konold
Curry School of Education, University of Virginia
Paul A. McDermott
University of Pennsylvania

This study employed observed factor index scores as well as latent ability constructs from the *Wechsler Intelligence Scale for Children–Fourth Edition* (WISC-IV; Wechsler, 2003) in estimating reading and mathematics achievement on the *Wechsler Individual Achievement Test–Second Edition* (WIAT-II; Wechsler, 2002). Participants were the nationally stratified linking sample ($N = 498$) of the WISC-IV and WIAT-II. Observed scores from the WISC-IV were analyzed using hierarchical multiple regression analysis. Although the factor index scores provided a statistically significant increment over the Full Scale IQ, the size of the improvement was too small to be of clinical utility. Observed WISC-IV subtest scores were also subjected to structural equation modeling (SEM) analyses. Subtest scores from the WISC-IV were fit to a general factor (g) and four ability constructs corresponding to factor indexes from the WISC-IV (Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed). For both reading and mathematics, only g (.55 and .77, respectively) and Verbal Comprehension (.37 and .17, respectively) were significant influences. Thus, when using observed scores to predict reading and mathematics achievement, it may only be necessary to consider the Full Scale IQ. In contrast, both g and Verbal Comprehension may be required for explanatory research.

Considerable effort is required to obtain all of the scores in most IQ tests. Presumably, such an investment is made to garner clinically useful information not available from the interpretation of just one omnibus composite. The inherent assumption underlying the interpretation of lower order subtest scores and factor indexes is that they offer practical diagnostic or treatment benefits not available from the general intelligence (g) estimate (Kamphaus, 2001; Kaufman, 1994; Sattler, 2001). Should the analysis of subtest or factor scores fail these premises, their relevance is effectively vitiated.

Subtest analysis has undergone serious challenges over the past 2 decades, both methodologically and empirically. For instance, a series of methodological problems were identified that operate to negate, or equivocate, essentially all research into children's subtest profiles. Prominent among the many limitations is the circular use of subtest profiles for *both* the initial formation of diagnostic groups *and* the subsequent search for profiles that might inherently define or distinguish those groups (Glutting, Watkins, & Youngstrom, 2003; McDermott, Fantuzzo, & Glutting, 1990; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992; Watkins & Kush, 1994). This problem is one of self-selection, which unduly increases the probability of discovering group differences. A second methodological deficiency is the nearly exclusive reliance on clinical samples. In contrast to epidemiological samples that are representative of the population as a whole, classified and referral samples (the majority of whom are subsequently classified) are unrepresentative and are also adversely affected by selection bias (Glutting, McDermott, Konold, Snelbaker, & Watkins, 1998; McDermott et al., 1992; Rutter, 1989). A third shortcoming is the misapplication of base rates, the frequency or percentage of a population identified with a diagnostic pattern (Cureton,

1957; Meehl & Rosen, 1955; Wiggins, 1973). The base rates routinely found in practice are so high that examinees overwhelmingly show an "exceptional" profile—sometimes exceeding 80% of all children in the United States (Glutting, McDermott, Watkins, Kush, & Konold, 1997; Kahana, Youngstrom, & Glutting, 2002). Besides being of dubious value, the high base rates raise a fundamental question: If everyone is exceptional, who then is normal?

The second trend comes from the empirical literature, which over the past 20 years has begun to demonstrate that subtest scores retain limited external validity. Examples of diminished utility include the inability of either individual subtest scores or score patterns to inform the identification of neurological deficits (Watkins, 1996), the diagnosis of learning disabilities (Daley & Nagle,1996; Glutting, McGrath, Kamphaus, & McDermott, 1992; Kavale & Forness, 1984; Kline, Snyder, Guilmette, & Castellanos, 1992; Livingston, Jennings, Reynolds, & Gray, 2003; Maller & McDermott, 1997; McDermott, Goldberg, Watkins, Stanley, & Glutting, in press; Mueller, Dennis, & Short, 1986; Reynolds & Kamphaus, 2003; Smith & Watkins, 2004; Ward, Ward, Hatt, Young, & Mollner, 1995; Watkins, 1999, 2000, 2003, in press; Watkins & Kush, 1994; Watkins, Kush, & Glutting, 1997a, 1997b; Watkins, Kush, & Schaefer, 2002; Watkins & Worrell, 2000), or the classification of behavioral, social, and emotional problems (Beebe, Pfiffner, & McBurnett, 2000; Dumont, Farr, Willis, & Whelley, 1998; Glutting et al., 1992; Glutting et al., 1998; Lipsitz, Dworkin, & Erlenmeyer-Kimling, 1993; McDermott & Glutting, 1997; Reinecke, Beebe, & Stein, 1999; Riccio, Cohen, Hall, & Ross, 1997; Rispens et al., 1997; Teeter & Korducki, 1998). Indeed, nonreactive support for this trend comes from a retrospective review of textbooks on children's intelligence testing (cf. Kamphaus, 1993, 2001; Kaufman, 1979, 1994; Sattler, 1974, 1982, 1988, 1992, 2001). In earlier publications, page after page of empirical studies extolled the importance, and clinical necessity, of interpreting telltale subtest configurations. More recent publications, by contrast, display far fewer affirmative citations, and they lead to one of two conclusions: (a) Empirical support is beginning to wane, or (b) subtest analysis is so universally corroborated that there is no need for referencing. But alas, as demonstrated clearly by the empirical literature just cited, the latter proposition is untrue.

Like most practitioners, we would agree that at least some abilities beyond *g* are clinically relevant. Detterman (2002) indicated that *g* accounts for only 25% to 50% of the variance in achievement, leaving 50% to 75% of the variance to be explained by other constructs. Following this logic, Brody (2002) reported that "no one believes that *g* is the only construct needed to describe individual differences in intelligence" (p. 122).

Factor scores are leading prospects in the provision of information beyond *g*. Factor scores are more valid than conceptual subtest groupings. Unlike the *inductively* derived subtest organizations of Sattler (2001) and Kaufman (1994), factor scores retain considerable construct validity because they are formed *empirically* on the basis of factor analysis. Each fac-

tor score in a test battery also accounts for more variance than that available from individual subtest scores. As a result, factor scores are more reliable than single subtest scores (as per the Spearman-Brown prophecy). Furthermore, because factor scores represent phenomena beyond the sum of method variance, measurement error, and subtest specificity, they potentially escape the myriad drawbacks that beset attempts to interpret subtest scores.

At the same time, the fact that a specific ability is supported by factor analysis does not necessarily mean that the ability has applied diagnostic merit (Briggs & Cheek, 1986). A case in point is the well-known Freedom From Distractibility (FD) factor. It was first described by J. Cohen in 1959 with the original *Wechsler Intelligence Scale for Children* (WISC; Wechsler, 1949). Over the ensuing 45 years, both its internal and its external criterion-related validity became so suspect (Barkley, 1998; M. Cohen, Becker, & Campbell, 1990; Kavale & Forness, 1984; Riccio et al., 1997; Wielkiewicz, 1990) that the factor no longer appears in the most recent WISC, the *Wechsler Intelligence Scale for Children–Fourth Edition* (WISC-IV; Wechsler, 2003). Therefore, it is essential to determine the extent to which factor-based abilities are externally valid, and specifically, to determine whether factor-based abilities provide substantial improvements in predicting important criteria above and beyond levels afforded by *g*.

Science seeks the simplest explanations of complex facts and then uses those explanations to craft hypotheses that are capable of being disproved (Platt, 1964). More specifically, the law of parsimony states that "what can be explained by fewer principles is needlessly explained by more" (Occam's razor; Jones, 1952, p. 620). The *g* construct satisfies the law of parsimony. It is singular, and more important, the *g*-based score has excellent criterion-related validity (American Psychological Association, Board of Scientific Affairs, 1996; Carroll, 1993; Gottfredson, 1997; Jensen, 1998; Lubinski, 2000; Lubinski & Humphreys, 1997). Consequently, it is imperative that factor-based abilities, to be regarded as useful, must demonstrate greater predictive validity than that obtainable from the lone *g* construct.

Two classes of factor-based variables are available for analysis: latent constructs and observed scores. Researchers oftentimes are interested in understanding theoretical relationships (Reeve, 2004). In such instances, they concentrate on the *latent constructs* underlying factor scores. Latent constructs are perfectly reliable, but they cannot be observed directly. They usually are analyzed through structural equation modeling (SEM), which is a multivariate statistical technique designed to identify relationships among latent variables (i.e., constructs).

Practitioners, on the other hand, are more interested in applied usage. *Observed scores* are the factor-based abilities routinely interpreted during clinical assessments. Psychologists are limited to interpretation of observed scores because latent constructs are not directly observable and are mathematically complex to derive from observed scores (cf. Oh, Glutting, Watkins, Youngstrom, & McDermott, 2004). Examples of observed factor scores are the four indexes in the WISC-

IV: the Verbal Comprehension Index (VCI), the Perceptual Reasoning Index (PRI), the Working Memory Index (WMI), and the Processing Speed Index (PSI). Observed factor scores are not the same as latent constructs, and observed scores clearly contain measurement error (i.e., reliability coefficients less than 1.00).

In either case, it is important to match the level of hypothesis (observed vs. latent variables) with the level of analysis (observed vs. latent variables) to avoid faulty conclusions (Ullman & Bentler, 2003). Some researchers have concentrated on the observed IQs interpreted by practitioners and tested the predictive validity of those scores via hierarchical multiple regression analysis (MRA; Glutting, Youngstrom, Ward, Ward, & Hale, 1997; Ree & Earles, 1991; Ree, Earles, & Treachout, 1994; Youngstrom, Kogos, & Glutting, 1999). In predictive MRA, it is important to demonstrate effects sufficiently large to have meaningful consequences. In other words, when observed ability scores are considered to be interpretable (i.e., to show statistically significant contributions), it is still necessary to demonstrate that their consequences for interpretation are large enough to be clinically relevant (Haynes & Lench, 2003; Hunsley & Meyer, 2003). Other researchers have employed latent constructs from SEM to study the criterion-related validity of intelligence tests (Gustafsson & Balke, 1993; Keith, 1999; Kuusinen & Leskinen, 1988; McGrew, Keith, Flanagan, & Vanderwood, 1997; Oh et al., 2004; Reeve, 2004). In such explanatory analyses, the goal was to demonstrate the effect of IQ constructs on other socially important latent variables.

Surprisingly, no study has attempted to employ both observed scores and latent constructs simultaneously to investigate the criterion-related validity of ability factors. Consequently, this study used both hierarchical MRA and SEM to investigate the relative importance of general versus specific abilities from the WISC-IV in predicting reading and mathematics achievement. The study also avoided reliance on clinical samples by using data from a demographically representative, epidemiological sample of children and adolescents.

Method

Participants and Instruments

All analyses began with standard scores from the linking sample of the WISC-IV and the *Wechsler Individual Achievement Test–Second Edition* (WIAT-II; Wechsler, 2002). Participants with complete data ($N = 498$) ranged in age from 6 years 0 months through 16 years 11 months (see Note). The linking sample was nationally representative within ±5% of the 2000 U.S. Census on the variables of age, gender, race/ethnicity, region of country, and parent education level. See Wechsler (2003) for a complete description of the linking sample and its representativeness of the U.S. child population.

The WISC-IV evaluates abilities among children 6 years 0 months through 16 years 11 months. The battery comprises 16 subtests (Ms = 10, SDs = 3). Of these, 10 are mandatory and

contribute to the formation of four factor-based indexes: the VCI, PRI, WMI, and PSI. Each of the four indexes is expressed as a standard score (Ms = 100, SDs = 15). Ability constructs for the WISC-IV either were based on the observed factor scores or were developed as latent traits from standard scores of the 10 mandatory subtests. The latent constructs were named verbal comprehension (VC), perceptual reasoning (PR), working memory (WM), and processing speed (PS) to distinguish each from its observed-score counterpart. Supplementary subtests were excluded because they are unnecessary for the formation of the WISC-IV's factors.

The WIAT-II contains nine subtests that can be aggregated into four composites: (a) Reading, (b) Mathematics, (c) Written Language, and (d) Oral Language. Like several earlier studies (Glutting, Youngstrom, et al., 1997; Keith, 1999; McGrew et al., 1997; Oh et al., 2004), the current investigation concentrated on outcomes in reading and mathematics. Therefore, either the observed Reading or Mathematics composite served as the dependent variable for the hierarchical MRAs. For the SEM analyses, the Reading and Mathematics constructs were developed from standard scores of the subtests underlying the WIAT-II's Reading (Pseudoword Decoding, Word Reading, and Reading Comprehension) and Mathematics (Numerical Operations and Math Reasoning) composites.

Data Analysis

The contribution of observed scores from the WISC-IV to the prediction of WIAT-II achievement was assessed through a series of hierarchical MRAs. Two different achievement composites (Reading and Mathematics) each served as the dependent measure in one set of regression analyses. The relative contribution of the observed Full Scale IQ (FSIQ) was compared with the four observed factor scores (VCI, PRI, WMI, PSI) through block entry and removal within the hierarchical MRA. Block entry focused on this question: Did any of the four observed factor scores substantially improve the prediction of reading or mathematics achievement above and beyond the contribution made by the parsimonious FSIQ? To explore this issue, the FSIQ entered the regression model by itself in the first block, and the four observed factor scores were entered as a group in the second block. The change in explained achievement variance resulting from the entrance of the second block provided an estimate of the maximum predictive increment attainable through the use of observed factor scores in addition to the FSIQ.

It could be argued that it is inappropriate to perform multiple regression analysis with highly correlated predictors, such as the WISC-IV's FSIQ and its four contributing factor scores (Fiorello, Hale, McGrath, Ryan, & Quinn, 2002). However, this argument ignores differences between explanatory and predictive research: "Predictive research emphasizes practical applications, whereas explanatory research focuses on achieving a theoretical understanding of the phenomenon of interest" (Venter & Maxwell, 2000, p. 152). The current MRA study is clearly predictive, and "there is nothing wrong with

any ordering of blocks [of variables in MRA] as long as the researcher does not use the results for explanatory purposes" (Pedhazur, 1997, p. 228).

It could also be argued that it is inappropriate to partial global ability (the FSIQ) prior to letting the observed factors predict achievement. Correspondingly, the hierarchical strategy would be reversed from the one used here (i.e., one would examine the effect of the factor scores and then let the FSIQ predict achievement). The strategy has some intuitive appeal, and it has been employed on occasion (Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001). However, to sustain such logic, we would have to repeal scientific law. That is, psychologists would be compelled to accept the novel notion that if many things essentially account for no more, or only marginally more, predictive variance than that accounted for by merely one thing (global ability), we should adopt the *less* parsimonious system.

All latent-trait models were completed through the Analysis of Moment Structures (AMOS; Arbuckle & Wothke, 1999) program using maximum likelihood estimation on covariance matrices. The left sides of Figures 1 and 2 provide a graphic representation of the WISC-IV measurement models that were investigated. The mandatory 10 subtests are enclosed in rectangles, and the four first-order factors and the *g* construct are enclosed in ellipses. This hierarchical configuration posits that the four first-order dimensions directly influence one or more of the underlying measured subtests, as indicated by the single-headed arrows. In turn, these four first-order factors are directly influenced by the overall ability construct (i.e., *g*). Respectively, the *r*s and *u*s enclosed in circles depict residual and unique variances in the measured and latent variables that are not accounted for by the higher order factors. Scaling of the latent variables was accomplished by setting a single path so that each assumed the scale of that variable. We investigated separate achievement models for Reading (Figure 1) and Mathematics (Figure 2).

Several nested structural models were employed to investigate the relative contributions of the WISC-IV's first-order factors, beyond *g*, in predicting children's latent achievement in reading and mathematics. The same model-comparison strategy was employed separately for each achievement domain. In both instances, a structural path linking the second-order *g* factor was specified to influence children's achievement factors. Subsequent models then estimated paths from each of the four first-order WISC-IV factors to the latent achievement variable of interest. For instance, VC was the first WISC-IV factor to be considered beyond *g*. The path was retained if VC resulted in a better statistical fit, as judged by the chi-square difference test. A path from the next first-order factor (i.e., PR) was then added to the model. If no statistical improvement was obtained in the model fit, the path was dropped before moving on to consider the next factor.

Multiple measures of fit, each developed under a somewhat different theoretical framework and focusing on a different aspect of fit, exist for evaluating the quality of measurement models (cf. Browne & Cudeck, 1993; Hu & Bentler, 1995). For this reason, it is generally recommended that multiple measures of fit be considered (Tanaka, 1993). Given the well-known problems with chi-square (χ^2) as a stand-alone measure of fit (Hu & Bentler, 1995; Kaplan, 1990), use of this statistic was limited to testing differences (χ^2_D) between competing models. In addition, the goodness of fit index (GFI), adjusted goodness of fit index (AGFI), Tucker-Lewis index (TLI), comparative fit index (CFI), and root mean square error of approximation (RMSEA) are reported for each model. The GFI is similar to a squared multiple correlation in that it provides a measure of the amount of variance/covariance that can be explained by the model. The AGFI, by contrast, is analogous to a squared multiple correlation corrected for model complexity. Thus, the AGFI is useful for comparing competing models. The TLI and CFI provide measures of model fit by comparing a given hypothesized model with a null model that assumes no relationship among the observed variables (Kranzler & Keith, 1999). These four measures generally range between 0.0 and 1.0, with values larger than .90 and .95 reflecting good and excellent fits, respectively, to the data (Bentler & Bonett, 1980; Marsh, Ellis, Heubeck, Parada, & Richards, 2005). Alternatively, smaller RMSEA values support better fitting models. Here values of .05 or less are generally taken to indicate good fit, although values of .08 or below are considered reasonable (Browne & Cudeck, 1993).

Results

Outcomes are presented separately according to the variables under consideration (observed scores, latent constructs). Means and standard deviations are presented in Table 1 for all measured variables in the WISC-IV and WIAT-II. As expected, scores on both instruments were normally distributed.

Observed Scores

Table 2 provides improvements obtained by entering the four observed factor indexes into the hierarchical MRA after the FSIQ was entered at the first step. The change in explained achievement variance resulting from entrance of the second block provided an estimate of the maximum predictive increment attainable through the use of factor scores in addition to the FSIQ. Table 2 also presents the unique contribution of each of the four factor scores (VCI, PRI, WMI, and PSI) when all variables were simultaneously included in the regression equation. These values are equivalent to the overall change in the model R^2 if the given variable (e.g., the VCI) was entered last into the equation after the contribution of the other predictors (e.g., the FSIQ, PRI, WMI, and PSI). Alternatively, they can be thought of as squared part correlations.

The FSIQ by itself explained 60.2% of the variance in the observed Reading composite and 59.7% of the variance in the Mathematics composite. As a group, the four factors ex-
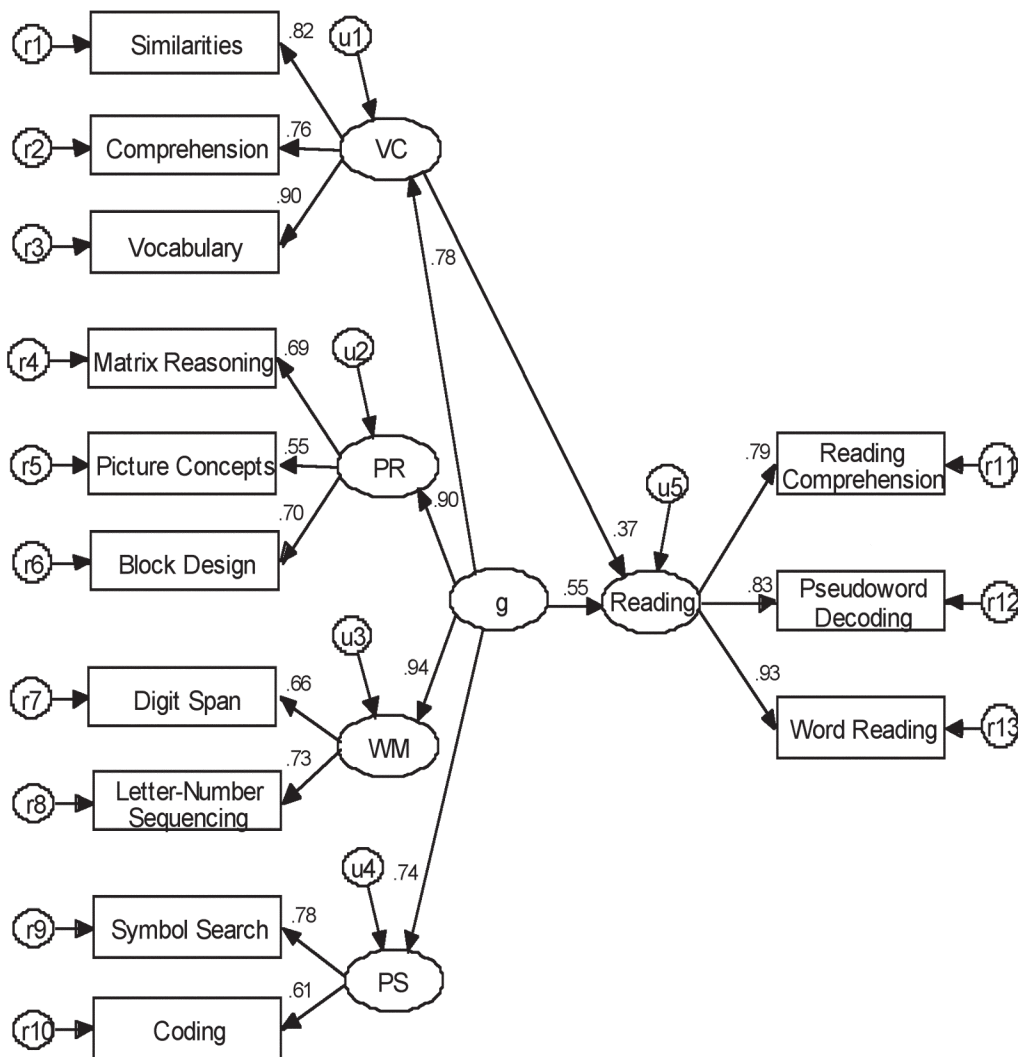
FIGURE 1. Standardized coefficients for *Wechsler Intelligence Scale for Children–Fourth Edition* (WISC-IV; Wechsler, 2003) structural influences on the *Wechsler Individual Achievement Test–Second Edition* (WIAT-II; Wechsler, 2002) Reading Factor.

plained an additional 1.8% of the variance in the Reading composite and 0.3% of the variance in the Mathematics composite. Although the factors as a group made statistically significant improvements in the prediction of the reading and mathematics criteria, the magnitude of increment was small according to standards offered by J. Cohen (1988; R^2 = .03 for a small effect, versus R^2 = .10 for a medium effect, or R^2 = .30 for a large effect). Importantly, none of the specific factor scores uniquely augmented, by even 2%, the explained variance in either reading or mathematics achievement.

There was a high degree of multicollinearity (i.e., redundancy) among the predictors as a consequence of the FSIQ being drawn in nearly its entirety from the same 10 subtests as the factor scores. However, such redundancy also is inherent in the scores psychologists interpret. SEM, by contrast, more accurately evaluates the true effects of one latent variable on another. Applying the dichotomy of Pedhazur (1997), the MRA analyses were predictive, whereas the SEM analyses were explanatory.

Latent Variables

Results of the nested model comparisons that were conducted on the two achievement constructs were similar in demonstrating that only VC influenced student achievement beyond *g*. The sections that follow consider each of the two achievement models under investigation in turn.

Reading. The first model that we investigated consisted of a single structural path linking the WISC-IV's latent *g* construct to the WIAT-II latent variable of Reading. The overall fit of this model was good, with GFI, AGFI, TLI, and CFI val-
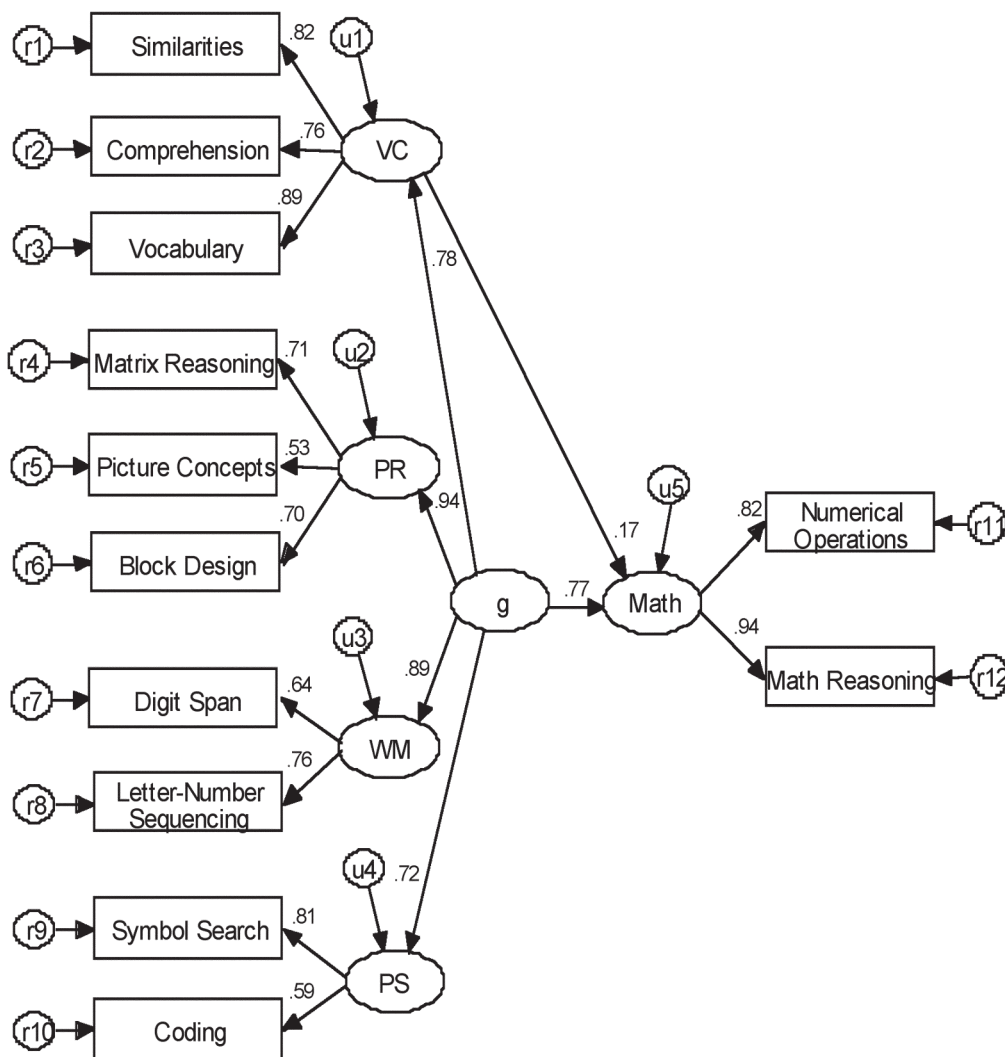
FIGURE 2. Standardized coefficients for *Wechsler Intelligence Scale for Children–Fourth Edition* (WISC-IV; Wechsler, 2003) structural influences on the *Wechsler Individual Achievement Test–Second Edition* (WIAT-II; Wechsler, 2002) Mathematics Factor.

ues in excess of .90, and RMSEAs less than .08 (see Table 3). The second iteration of this model freed the path from VC to Reading. This was a statistically better fitting model than the previous model that had only a path between *g* and Reading, $\chi^2_D(1) = 19.83$, $p < .05$. In addition, all other measures of fit remained good to excellent (see Table 3). Subsequent models that attempted to link PR, $\chi^2_D(1) = 0.49$, $p > .05$; WM, $\chi^2_D(1) = 0.45$, $p > .05$; and PS, $\chi^2_D(1) = 0.0$, $p > .05$; to Reading failed to demonstrate statistically better fits in comparison with the model that included paths from *g* and VC. As a result, only the *g* and VC factors were found to influence Reading in terms of model fit and parsimony.

Standardized values for this final Reading model are presented in Figure 1. Measurement models for the WISC-IV and WIAT-II Reading constructs were favorable. All factor load-ings were statistically significant and ranged from a low of .55 for Picture Concepts to a high of .90 for Vocabulary on the WISC-IV, and from a low of .79 for Reading Comprehension to a high of .93 for Word Reading on the WIAT-II Reading factor. The squared multiple correlations (SMCs; not shown in Figure 1) were also favorable, ranging from a low of .30 for Picture Concepts to a high of .81 for Vocabulary across the WISC-IV subtests, and a low of .63 for Reading Comprehension to a high of .86 for Word Reading across the Reading factor scales. Standardized parameter estimates linking *g* to the four underlying first-order factors were moderate to large, with values ranging from .74 to .94. The SMCs for these factor scores ranged from a low of .54 for PS to a high of .88 for WM, indicating that an appreciable amount of factor score variance is accounted for by *g*.

TABLE 1. Descriptive Statistics for the WISC-IV and WIAT-II

Scale	*M*	*SD*
WISC-IV		
FSIQ	100.11	15.10
VCI	99.60	15.09
PRI	98.99	14.25
WMI	100.74	14.88
PSI	99.26	14.69
Similarities	9.97	2.70
Comprehension	10.34	2.82
Vocabulary	10.12	3.01
Matrix Reasoning	9.92	2.88
Picture Concepts	10.04	2.96
Block Design	9.84	2.84
Digit Span	10.03	2.89
Letter–Number Sequencing	10.11	2.96
Symbol Search	10.19	2.96
Coding	10.25	2.93
WIAT-II		
Reading composite	100.16	16.81
Mathematics composite	101.28	17.26
Pseudoword Decoding	101.41	14.31
Word Reading	101.32	15.41
Reading Comprehension	100.85	15.92
Numerical Operations	102.26	15.84
Math Reasoning	101.47	15.72

Note. WISC-IV = *Wechsler Intelligence Scale for Children–Fourth Edition* (Wechsler, 2003); WIAT-II = *Wechsler Individual Achievement Test–Second Edition* (Wechsler, 2002); FSIQ = Full Scale IQ; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index;, PSI = Processing Speed Index.

The paths of primary substantive interest in this model were those linking *g* and VC to Reading. Though both were statistically significant, *g* clearly had more influence on Reading than the VC as gauged by the standardized coefficients presented in Figure 1. Path coefficients are standardized to *M* = 0.0 and *SD* = 1.0 and are interpreted in terms of *SD* units. For example, the .37 path from VC to Reading means that for each *SD* increase in VC, performance on the Reading construct would increase by .37 *SD*. In comparison, a 1 *SD* increase in *g* would increase Reading by .55 *SD*. According to the rough guidelines provided by Kline (2005), the effect size of VC was medium, whereas the effect size of *g* was large. Jointly, those two factors accounted for 75% of the variance in Reading.

Mathematics. As in the strategy we employed for understanding the influences of Reading, we first investigated a model consisting of a single structural path linking the WISC-IV's latent *g* construct to the WIAT-II latent Mathematics variable. The overall fit of this model was excellent, with GFI,

AGFI, TLI, and CFI values in excess of .95, and a RMSEA less than .05 (see Table 3). The second model freed the path from VC to Mathematics, which was a statistically better fitting model, $\chi^2_D(1) = 4.36$, $p < .05$. In addition, all other fit indexes continued to be excellent (see Table 3). Subsequent models that attempted to link PR, $\chi^2_D(1) = 3.45$, $p > .05$; WM, $\chi^2_D(1) = 2.52$, $p > .05$; and PS, $\chi^2_D(1) = 0.14$, $p > .05$; to Mathematics failed to demonstrate statistically better fitting models in comparison with the model that included paths from *g* and VC. Thus, as with Reading, only the *g* and VC factors were found to influence Mathematics in terms of model fit and parsimony.

Standardized values for the final Mathematics model are presented in Figure 2. The measurement model for the WIAT-II Mathematics construct was favorable. All factor loadings were statistically significant and ranged from a low of .82 for Numerical Operations to a high of .94 for Math Reasoning. The SMCs were also favorable, ranging from a low of .67 for Numerical Operations to a high of .89 for Math Reasoning. The paths of primary substantive interest in this model were those linking the *g* and VC factors to Mathematics. Both were again statistically significant; however, the discrepancy between the influence of *g* and VC was even more pronounced when the outcome construct of interest was Mathematics (see Figure 2 for standardized coefficients). The effect size of VC (.17) could be categorized as small to medium, whereas the effect size of *g* (.77) was large (Kline, 2005). Jointly, these two factors accounted for 81% of the variance in Mathematics.

Discussion

The current study makes clear-cut distinctions between the applied versus theoretical utility of factor-based abilities. It does so in the context of estimating reading and mathematics achievement scores. IQ tests are also currently used to determine eligibility for special education, for example by identifing specific learning disabilities (LD) through IQ–achievement discrepancies. Professionals, however, might have many legitimate reservations about using IQ–achievement discrepancies to diagnose LD (Aaron, 1997; Fletcher et al., 1998; Fuchs, Fuchs, & Speece, 2002; Siegel, 1998; Vellutino, Scanlon, & Lyon, 2000).

The purpose of this study was to examine both predictive and explanatory relationships between ability and achievement using the linking sample of the WISC-IV and WIAT-II. MRA analyses were predictive, whereas SEM analyses were explanatory (Pedhazur, 1997). The MRA analyses tested hypotheses about observed variables, whereas the SEM analyses tested hypotheses about latent variables (Ullman & Bentler, 2003). At the observed variable level, the FSIQ accounted for the bulk of variance (approximately 60%) in both reading and mathematics composite scores. Although the factor-score indexes provided a statistically significant increment over the FSIQ, the size of improvement was too small to be of clini-

TABLE 2. Incremental Contribution of Observed WISC-IV Factor Scores in Predicting Reading and Mathematics Composites on the WIAT-II

Predictor	Reading		Mathematics	
	Variance (%)	Increment[a] (%)	Variance (%)	Increment[a] (%)
FSIQ	60.2*	60.2*	59.7*	59.7*
Four factors (df = 4)[b]	62.0*	1.8*	60.0*	0.3*
VCI			0.2	0.0
PRI			0.0	0.0
WMI			0.1	0.0
PSI			0.0	0.0

Note. WISC-IV = *Wechsler Intelligence Scale for Children–Fourth Edition* (Wechsler, 2003); WIAT-II = *Wechsler Individual Achievement Test–Second Edition* (Wechsler, 2002); FSIQ = Full Scale IQ; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index.
[a]Unless indicated otherwise, all unique contributions are squared part correlations, equivalent to the change in R^2 if this variable were entered last in a block entry regression. [b]Partialing out FSIQ.
*p = .001.

TABLE 3. Structural Model Fit Statistics for the Prediction of WIAT-II Reading and Mathematics Achievement Constructs From WISC-IV Ability Constructs

	χ^2	df	GFI	AGFI	TLI	CFI	RMSEA
Reading							
FSIQ	211.69	60	.94	.91	.94	.95	.071
FSIQ, VC	191.86	59	.94	.91	.95	.96	.067
FSIQ, VC, PR	191.37	58	.94	.91	.95	.96	.068
FSIQ, VC, WM	191.41	58	.94	.91	.95	.96	.068
FSIQ, VC, PS	191.86	58	.94	.91	.95	.96	.068
Mathematics							
FSIQ	84.78	49	.97	.96	.98	.99	.038
FSIQ, VC	80.42	48	.97	.96	.98	.99	.037
FSIQ, VC, PR	76.97	47	.98	.96	.99	.99	.036
FSIQ, VC, WM	77.90	47	.98	.96	.99	.99	.036
FSIQ, VC, PS	80.28	47	.97	.96	.98	.99	.038

Note. WISC-IV = *Wechsler Intelligence Scale for Children–Fourth Edition* (Wechsler, 2003); WIAT-II = *Wechsler Individual Achievement Test–Second Edition* (Wechsler, 2002); GFI = goodness of fit index; AGFI = adjusted goodness of fit index; TLI = Tucker-Lewis index; CFI = comparative fit index; RMSEA = root mean square error of approximation; FSIQ = Full Scale IQ; VC = Verbal Comprehension factor; PR = Perceptual Reasoning factor; WM = Working Memory factor; PS = Processing Speed factor.

cal utility (no single index provided an increment greater than 1%). The current findings for observed factor scores from the WISC-IV align well with previous epidemiological studies from both the United States and Europe that showed specific cognitive abilities add little or nothing to prediction beyond the contribution made by *g* (Jencks et al., 1979; Ree et al., 1994; Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003; Schmidt & Hunter, 1998; Thorndike, 1986).

At the latent variable level, only *g* and VC significantly influenced the reading and mathematics achievement constructs. For both reading and mathematics, *g* exhibited a large effect size (.55 and .77, respectively). In contrast, VC had a

medium effect on reading (.37) and a small-to-medium effect on mathematics (.17). Thus, both *g* and VC explained reading and mathematics achievement, but *g* was the more powerful explanatory construct. Current findings with the WISC-IV and WIAT-II are consonant with SEM outcomes obtained by Keith (1999) and McGrew et al. (1997) with the Woodcock-Johnson–Revised and Oh et al. (2004) with the *Wechsler Intelligence Scale for Children–Third Edition* (WISC-III; Wechsler, 1991). Similar conclusions were reached by Kuusinen and Leskinen (1988) as well as by Gustafsson and Balke (1993) with other measures of ability and achievement. When general and specific ability constructs are compared with general and spe-

cific achievement constructs, *g* usually accounts for the largest proportion of variance in achievement. However, additional variance in narrow achievement domains may be explained by specific cognitive constructs, especially at higher *g* levels (Lubinski, 2000).

The current MRA results are relatively equivalent to the 2% to 4% increments found by Youngstrom et al. (1999) with the *Differential Ability Scales* (Elliott, 1990), but lower than the 5% to 16% increments found previously for factor scores with the linking sample of the WISC-III and original WIAT (Glutting, Youngstrom, et al., 1997). Therefore, observed factors scores from the WISC-IV appear to be even *less* clinically relevant in predicting children's reading and mathematics achievement than was the case with the earlier WISC-III. Nevertheless, all of these results are consistent with Thorndike's (1985) observation that roughly 85% to 90% of predictable variance in criterion variables is accounted for by the single general score from an ability test battery.

Applied Implications

Results from the SEM analyses suggest that psychologists must go beyond *g* to meaningfully understand children's latent cognitive abilities. At the same time, psychologists should *not* give equal weight to all WISC-IV constructs. For example, when attempting to explain children's reading achievement on the WIAT-II, psychologists should limit interpretations to just two constructs: *g* and VC. No increase in explanatory power will be obtained from the PR, WM, or PS constructs. Likewise, when explaining children's mathematics achievement, psychologists should confine interpretations to just *g* and VC. Thus, current results strongly indicate that psychologists should look no further than the WISC-III constructs of *g* and VC when attempting to explain achievement in two of the most crucial areas of education: reading and mathematics.

Practitioners can directly apply current findings from the MRA analyses, unlike the results from the SEM analyses, to their day-to-day assessments. For example, when examining for reading or mathematics problems, psychologists would limit interpretations to just the FSIQ. Alternatively, whereas practitioners may want to apply results from the SEM analyses and interpret both the FSIQ and the VCI, results show that including even just the observed VCI (and ignoring the WISC-IV's other three factor scores) is likely to lead to overinterpretation.

To understand why overinterpretation is probable, psychologists must recognize that the observed scores obtained during clinical assessments are *very* different than the latent traits (i.e., constructs) derived by SEM. Observed scores are standard scores, such as the FSIQ, index scores, and subtest scores in the WISC-IV. SEM, on the other hand, provides results that are best interpreted as relationships among *pure constructs measured without error*. SEM is a good method for testing theory but it is less satisfactory for direct, diagnostic applications. The observed scores employed by psychologists contain measurement error, whereas latent SEM traits do not (i.e., reliability coefficients = 1.00). Basing diagnostic decisions on theoretically pure constructs is very difficult in practice. In fact, we previously demonstrated the following: (a) The constructs from SEM *rank children differently* than observed scores, and children's relative position on factor-based constructs (e.g., VC) can be radically different than their standing on corresponding observed factor scores (the VCI); (b) construct scores are not readily available to psychologists; and (c) although it is possible to estimate construct scores, the calculations are difficult and laborious (cf. Oh et al., 2004, for an example). Therefore, one of the most important findings here is that *psychologists cannot directly apply results from SEM*. Observed scores must first be converted to construct scores before outcomes can be translated into practical, everyday use. This situation holds not only for ability and achievement tests but for *all* SEM findings, regardless of whether analyses are directed to personality variables (e.g., parent, teacher, and self-reports of psychopathology), neuropsychological test scores, results from memory experiments, or data from similar sources.

Some psychologists have recommended interpretation of the factor indexes over the FSIQ when predicting academic achievement (Weiss, Saklofske, & Prifitera, 2005; Williams, Weiss, & Rolfhus, 2003). To the contrary, current results reveal that psychologists need to interpret only the FSIQ to predict performance in reading and mathematics. This is because the WISC-IV factor indexes do not substantially increment predictive validity beyond the FSIQ. There may be several reasons for the weak contribution of factor scores. First, performance on any subtest reflects a mixture of method variance, error variance, and construct representation from general, broad, and narrow abilities (Carroll, 1993). For example, a number of subtests may be summed to create an omnibus IQ, but a proportion of that score's nonerror variance will be contributed by narrow and broad abilities in addition to general ability (McClain, 1996). Thus, the omnibus IQ measure will contain some variance from the lower order constructs that will take precedence in hierarchical predictive situations. This distinction between *g* and the FSIQ was described by Colom, Abad, Garcia, and Juan-Espinosa (2002) as the difference between "general intelligence" and "intelligence in general." Second, as with the accumulation of true score variance across items (Cronbach, 1951), "broad abilities account for a relatively small proportion of variance in specific tasks but for a substantial proportion of the variance in scores that are aggregated over several tasks" (Gustafsson & Undheim, 1996, p. 205). Thus, the FSIQ formed by summing over subscale scores is powerfully affected by the *g* factor (Lubinski & Dawis, 1992). For example, Gustafsson and Undheim (1996) found that 71% of the total variance of the WISC-III FSIQ was due to the *g* factor. Finally, measurement error and unique variance components may obfuscate relationships between obtained scores that were apparent in analyses of constructs purged of measurement error (i.e., SEM).

NOTE

REFERENCES

Aaron, P. G. (1997). The impending demise of the discrepancy formula. *Review of Educational Research, 67,* 461–502.

American Psychological Association, Board of Scientific Affairs. (1996). Intelligence: Knowns and unknowns. *The American Psychologist, 51,* 77–101.

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide.* Chicago: Small Waters.

Barkley, R. A. (1998). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (2nd ed.). New York: Guilford Press.

Beebe, D. W., Pfiffner, L. J., & McBurnett, K. (2000). Evaluation of the validity of the Wechsler Intelligence Scale for Children–Third edition comprehension and picture arrangement subtests as measures of social intelligence. *Psychological Assessment, 12,* 97–101.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality, 54,* 106–148.

Brody, N. (2002). *g* and the one–many problem: Is one enough? In *The nature of intelligence* (Novartis Foundation Symposium 233; pp. 122–135). New York: Wiley.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

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

Cohen, J. (1959). The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *Journal of Consulting and Clinical Psychology, 23,* 285–299.

Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, M., Becker, M. G., & Campbell, R. (1990). Relationships among four methods of assessment of children with attention deficit–hyperactivity disorder. *Journal of School Psychology, 28,* 189–202.

Colom, R., Abad, F. J., Garcia, L. F., & Juan-Espinosa, M. (2002). Education, Wechsler's full scale IQ, and *g. Intelligence, 30,* 449–462.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334.

Cureton, E. E. (1957). Recipe for a cookbook. *Psychological Bulletin, 54,* 494–497.

Daley, C. E., & Nagle, R. J. (1996). Relevance of WISC-III indicators for assessment of learning disabilities. *Journal of Psychoeducational Assessment, 14,* 320–333.

Detterman, D. K. (2002). General intelligence: Cognitive and biological explanations. In *The nature of intelligence* (Novartis Foundation Symposium 233; pp. 223–243). New York: Wiley.

Dumont, R., Farr, L. P., Willis, J. O., & Whelley, P. (1998). 30-second interval performance on the Coding subtest of the WISC-III: Further evidence of WISC folklore? *Psychology in the Schools, 35,* 111–117.

Elliott, C. D. (1990). *Differential ability scales.* San Antonio, TX: Psychological Corp.

Fiorello, C. A., Hale, J. B., McGrath, M., Ryan, K., & Quinn, S. (2002). IQ interpretation for children with flat and variable test profiles. *Learning and Individual Differences, 13,* 115–125.

Fletcher, J. M., Francis, D. J., Shaywitz, S. E., Lyon, G. R., Foorman, B. R., Stuebing, K. K., et al. (1998). Intelligent testing and the discrepancy model for children with learning disabilities. *Learning Disabilities Research & Practice, 13,* 186–203.

Fuchs, L. S., Fuchs, D., & Speece, D. L. (2002). Treatment validity as a unifying construct for identifying learning disabilities. *Learning Disability Quarterly, 25,* 33–45.

Glutting, J. J., McDermott, P. A., Konold, T. R., Snelbaker, A. J., & Watkins, M. W. (1998). More ups and downs of subtest analysis: Criterion validity of the DAS with an unselected cohort. *School Psychology Review, 27,* 599–612.

Glutting, J. J., McDermott, P. A., Watkins, M. W., Kush, J. C., & Konold, T. R. (1997). The base rate problem and its consequences for interpreting children's ability profiles. *School Psychology Review, 26,* 176–188.

Glutting, J. J., McGrath, E. A., Kamphaus, R. W., & McDermott, P. A. (1992). Taxonomy and validity of subtest profiles on the Kaufman Assessment Battery for Children. *The Journal of Special Education, 26,* 85–115.

Glutting, J. J., Watkins, M. W., & Youngstrom, E. A. (2003). Multifactored and cross-battery ability assessments: Are they worth the effort? In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment: Vol. 1. Intelligence and achievement* (2nd ed., pp. 343–373). New York: Guilford Press.

Glutting, J. J., Youngstrom, E. A., Ward, T., Ward, S., & Hale, R. (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). Intelligence and social policy. *Intelligence, 24,* 288–320.

Gustafsson, J. E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28,* 407–434.

Gustafsson, J. E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 186–242). New York: Macmillan.

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.

Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment, 15,* 456–466.

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment, 15,* 446–455.

Jencks, C., Bartlett, S., Corcoran, M., Crouse, J., Eaglesfield, D., Jackson, G., et al. (1979). *Who gets ahead? The determinants of economic success in America.* New York: Basic Books.

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

Jones, W. T. (1952). *A history of western philosophy.* New York: Harcourt, Brace.

Kahana, S. Y., Youngstrom, E. A., & Glutting, J. J. (2002). Factor and subtest discrepancies on the Differential Abilities Scales: Examining prevalence and validity in predicting academic achievement. *Assessment, 9,* 82–93.

Kamphaus, R. W. (1993). *Clinical assessment of child and adolescent intelligence.* Boston: Allyn & Bacon.

Kamphaus, R. W. (2001). *Clinical assessment of child and adolescent intelligence* (2nd ed.). Boston: Allyn & Bacon.

Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research, 25,* 137–155.

Kaufman, A. S. (179). *Intelligent testing with the WISC-R.* New York: Wiley.

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

Kavale, K. A., & Forness, S. R. (1984). A meta-analysis of the validity of

Wechsler scale profiles and recategorizations: Patterns or parodies? *Learning Disabilities Quarterly, 7,* 136–156.

Keith, T. Z. (1999). Effects of general and specific abilities on student achievement: Similarities and differences across ethnic groups. *School Psychology Quarterly, 14,* 239–262.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.

Kline, R. B., Snyder, J., Guilmette, S., & Castellanos, M. (1992). Relative usefulness of elevation, variability, and shape information from WISC-R, K-ABC, and Fourth Edition Stanford-Binet profiles in predicting achievement. *Psychological Assessment, 4,* 426–432.

Kranzler, J. H., & Keith, T. Z. (1999). Independent confirmatory factor analysis of the Cognitive Assessment System (CAS): What does the CAS measure? *School Psychology Review, 28,* 117–144.

Kuusinen, J., & Leskinen, E. (1988). Latent structure analysis of longitudinal data on relations between intellectual abilities and school achievements. *Multivariate Behavioral Research, 23,* 103–118.

Lipsitz, J. D., Dworkin, R. H., & Erlenmeyer-Kimling, L. (1993). Wechsler comprehension and picture arrangement subtest and social adjustment. *Psychological Assessment, 5,* 430–437.

Livingston, R. B., Jennings, E., Reynolds, C. R., & Gray, R. M. (2003). Multivariate analyses of the profile stability of intelligence tests for IQs, low to very low for subtest analyses. *Archives of Clinical Neuropsychology, 18,* 487–507.

Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "Sinking shafts at a few critical points." *Annual Review of Psychology, 51,* 405–444.

Lubinski, D., & Dawis, R. V. (1992). Aptitudes, skills, and proficiencies. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 3, pp. 1–59). Palo Alto, CA: Consulting Psychology Press.

Lubinski, D., & Humphreys, L. G. (1997). Incorporating general intelligence into epidemiology and the social sciences. *Intelligence, 24,* 159–201.

Maller, S. J., & McDermott, P. A. (1997). WAIS-R profile analysis for college students with learning disabilities. *School Psychology Review, 26,* 575–585.

Marsh, H. W., Ellis, L. A., Heubeck, B. G., Parada, R. H., & Richards, G. (2005). A short version of the Self Description Questionnaire II: Operationalizing criteria for short-form evaluation with new applications of confirmatory factor analyses. *Psychological Assessment, 17,* 81–102.

McClain, A. L. (1996). Hierarchical analytic methods that yield different perspectives on dynamics: Aids to interpretation. *Advances in Social Science Methodology, 4,* 229–240.

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

McDermott, P. A., Fantuzzo, J. W., Glutting, J. J., Watkins, M. W., & Baggaley, A. R. (1992). Illusions of meaning in the ipsative assessment of children's abilities. *The Journal of Special Education, 25,* 504–526.

McDermott, P. A., & Glutting, J. J. (1997). Informing stylistic learning behavior, disposition, and achievement through ability subtest patterns: Practical implications for test interpretation [Special issue]. *School Psychology Review, 26,* 163–175.

McDermott, P. A., Goldberg, M. M., Watkins, M. W., Stanley, J. L., & Glutting, J. J. (in press). A nationwide epidemiological modeling study of learning disabilities: Risk, protection, and unintended impact. *Journal of Learning Disabilities.*

McGrew, K. S., Keith, T. Z., Flanagan, D. P., & Vanderwood, M. (1997). Beyond *g*: The impact of Gf-Gc specific cognitive ability research on the future use and interpretation of intelligence tests in the schools. *School Psychology Review, 26,* 189–201.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, and cutting scores. *Psychological Bulletin, 52,* 194–216.

Mueller, H. H., Dennis, S. S., & Short, R. H. (1986). A meta-exploration of

WISC-R factor score profiles as a function of diagnosis and intellectual level. *Canadian Journal of School Psychology, 2,* 21–43.

Oh, H. J., Glutting, J. J., Watkins, M. W., Youngstrom, E. A., & McDermott, P. A. (2004). Correct interpretation of latent versus observed abilities: Implications from structural equation modeling applied to the WISC-III and WIAT linking sample. *The Journal of Special Education, 38,* 159–173.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York: Harcourt Brace.

Platt, J. R. (1964). Strong inference. *Science, 146,* 347–352.

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than *g. Personnel Psychology, 44,* 321–331.

Ree, M. J., Earles, J. A., & Treachout, M. S. (1994). Predicting job performance: Not much more than *g. The Journal of Applied Psychology, 79,* 518–524.

Reeve, C. L. (2004). Differential ability antecedents of general and specific dimensions of declarative knowledge: More than *g. Intelligence, 32,* 621–652.

Reinecke, M. A., Beebe, D. W., & Stein, M. A. (1999). The third factor of the WISC-III: It's (probably) not freedom from distractibility. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 322–328.

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds intellectual assessment scales.* Lutz, FL: Psychological Assessment Resources.

Riccio, C. A., Cohen, M. J., Hall, J., & Ross, C. M. (1997). The third and fourth factors of the WISC-III: What they don't measure. *Journal of Psychoeducational Assessment, 15,* 27–39.

Rispens, J., Swaab, H., van den Oord, E. J., Cohen-Kettenis, P., van Engeland, H., & van Yperen, T. (1997). WISC profiles in child psychiatric diagnosis: Sense or nonsense? *Journal of the American Academy of Child and Adolescent Psychiatry, 36,* 1587–1594.

Rutter, M. (1989). Isle of Wright revisited: Twenty-five years of child psychiatric epidemiology. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 633–653.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & de Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology, 56,* 573–605.

Sattler, J. M. (1974). *Assessment of children's intelligence* (Revised reprint). Philadelphia: W. B. Saunders.

Sattler, J. M. (1982). *Assessment of children's intelligence and special abilities* (2nd ed.). Boston: Allyn & Bacon.

Sattler, J. M. (1988). *Assessment of children* (3rd ed.). San Diego: Author.

Sattler, J. M. (1992). *Assessment of children* (3rd ed., revised and updated). San Diego: Author.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego: Author.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262–274.

Siegel, L. S. (1998). The discrepancy formula: Its use and abuse. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 123–135). Timonium, MD: York Press.

Smith, C. B., & Watkins, M. W. (2004). Diagnostic utility of the Bannatyne WISC-III pattern. *Learning Disabilities Research & Practice, 19,* 49–56.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. S. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10–39). Newbury Park, CA: Sage.

Teeter, P. A., & Korducki, R. (1998). Assessment of emotionally disturbed children with the WISC-III. In A. Prifitera & D. H. Saklofske (Eds.), *WISC-III clinical use and interpretation: Scientist–practitioner perspectives* (pp. 119–138). New York: Academic Press.

Thorndike, R. L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior, 29,* 332–339.

Ullman, J. B., & Bentler, P. M. (2003). Structural equation modeling. In J. A.

Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (Vol. 2, pp. 607–634). Hoboken, NJ: Wiley.

Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ–achievement discrepancy definition for reading disability. *Journal of Learning Disabilities, 33,* 223–238.

Venter, A., & Maxwell, S. E. (2000). Issues in the use and application of multiple regression analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied statistics and mathematical modeling* (pp. 151–182). New York: Academic Press.

Ward, S. B., Ward, T. J., Hatt, C. V., Young, D. L., & Mollner, N. R. (1995). The incidence and utility of the ACID, ACIDS, and SCAD profiles in a referred population. *Psychology in the Schools, 32,* 267–276.

Watkins, M. W. (1996). Diagnostic utility of the WISC-III developmental index as a predictor of learning disabilities. *Journal of Learning Disabilities, 29,* 305–312.

Watkins, M. W. (1999). Diagnostic utility of WISC-III subtest variability among students with learning disabilities. *Canadian Journal of School Psychology, 15,* 11–20.

Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly, 15,* 465–479.

Watkins, M. W. (2003). IQ subtest analysis: Clinical acumen or clinical illusion. *Scientific Review of Mental Health Practice, 2,* 118–141.

Watkins, M. W. (in press). Diagnostic validity of Wechsler subtest scatter. *Learning Disabilities: A Contemporary Journal.*

Watkins, M. W., & Kush, J. C. (1994). WISC-R subtest analysis of variance: The right way, the wrong way, or no way? *School Psychology Review, 23,* 640–651.

Watkins, M. W., Kush, J. C., & Glutting, J. J. (1997a). Discriminant and predictive validity of the WISC-III ACID profile among children with learning disabilities. *Psychology in the Schools, 34,* 309–319.

Watkins, M. W., Kush, J. C., & Glutting, J. J. (1997b). Prevalence and diagnostic utility of the WISC-III SCAD profile among children with disabilities. *School Psychology Quarterly, 12,* 235–248.

Watkins, M. W., Kush, J. C., & Schaefer, B. A. (2002). Diagnostic utility of the learning disability index. *Journal of Learning Disabilities, 35,* 98–103.

Watkins, M. W., & Worrell, F. C. (2000). Diagnostic utility of the number of WISC-III subtests deviating from mean performance among students with learning disabilities. *Psychology in the Schools, 37,* 29–41.

Wechsler, D. (1949). *Wechsler intelligence scale for children.* New York: Psychological Corp.

Wechsler, D. (1991). *Wechsler intelligence scale for children* (3rd ed.). San Antonio, TX: Psychological Corp.

Wechsler, D. (2002). *Wechsler individual achievement test* (2nd ed.). San Antonio, TX: Psychological Corp.

Wechsler, D. (2003). *Wechsler intelligence scale for children* (4th ed.). San Antonio, TX: Psychological Corp.

Weiss, L. G., Saklofske, D. H., & Prifitera, A. (2005). Interpreting the WISC-IV index scores. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist–practitioner perspectives* (pp. 71–100). New York: Elsevier Academic Press.

Wielkiewicz, R. M. (1990). Interpreting low scores on the WISC-R third factor: It's more than distractibility. *Psychological Assessment, 2,* 91–97.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment.* Reading, MA: Addison-Wesley.

Williams, P. E., Weiss, L. G., & Rolfhus, E. L. (2003). *Clinical validity* (WISC-IV Technical Report No. 3). San Antonio, TX: Psychological Corp.

Youngstrom, E. A., Kogos, J. L., & Glutting, J. J. (1999). Incremental efficacy of Differential Ability Scales factor scores in predicting individual achievement criteria. *School Psychology Quarterly, 14,* 26–39.