

Cognitive Profile Analysis: A Shared Professional Myth

Marley W. Watkins
The Pennsylvania State University

Interpretation of cognitive test profiles is popular in school psychological training and practice but is not supported by research evidence. This special issue contains four studies that provide additional data regarding cognitive test profiles. Consistent with the largely negative research literature, detailed analysis found all four cognitive profile reports lacking reliability, validity, or diagnostic utility; even cognitive profiles composed of composites were psychometrically weak. These results were not surprising because ipsative methods are inferior to normative methods in cognitive assessment. Given this consistent failure of empirical validation, belief in the utility of cognitive test profile interpretations was likened to a shared professional myth and it was recommended that psychologists eschew the application of cognitive test profiles for differential diagnosis and remediation.

Approximately 74% of school psychology training programs place moderate to great emphasis on the use of subtest scores in their individual cognitive assessment courses (Alfonso, Oakland, LaRocca, & Spanakos, 2000) and school psychologists frequently analyze cognitive subtest profiles (Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000). Among their sample of practitioners, for example, Pfeiffer et al. (2000) found that almost 70% reported factor scores or profile analysis to be a useful feature of the Wechsler Intelligence Scale for Children—Third Edition (WISC-III) and 29% reported that they derived specific value from individual WISC-III subtests.

Based on their survey results, Pfeiffer et al. (2000) suggested that critics of cognitive subtest analysis “extend an invitation to the practitioner to learn why practitioners—operating in the complexity of the clinical world—find considerable value in using clinical interpretative techniques such as profile analysis” (p. 384). However, personal explanations are an unreliable validation strategy (Ruscio, 1998) and neither popularity nor longevity of a clinical practice necessarily speaks to its verity. In medicine, for example, practices widespread for centuries were later recognized to have little or no curative effect. Medicine, therefore, provides dramatic examples “of the inadequacies of clinical observation when compared to scientific method” (Ziskin, 1995, p. 53). There is abundant evi-

dence that psychological practice is also vulnerable to the development of beliefs and conjectures that cannot withstand scientific scrutiny (Aspel, Willis, & Faust, 1998; Davidow & Levinson, 1993; Faust, 1986). After an historical review of professional practice, Meehl (1997) concluded that "a clear message of history is that the anecdotal method can deliver wheat or chaff, and it does not enable us to tell which is which" (p. 94). To sift the wheat from the chaff, clinicians must learn when not to rely on intuitive judgments and should adopt a scientific approach to hypothesis testing (Faust, 1986; Tracey & Rounds, 1999).

EXISTING EVIDENCE REGARDING PROFILE ANALYSIS

Program accreditation guidelines of the American Psychological Association (2000) affirm that empirical evidence must support psychological practice. Consequently, "what *exactly* is the evidence" (Loftus, 1998, p. 61) that supports the popular practice of cognitive profile interpretation? A review of the existing professional literature reveals overwhelmingly negative results (Dumont & Willis, 1995; Glutting, McDermott, Konold, Snelbaker, & Watkins, 1998; Kavale & Forness, 1984; Kramer, Henning-Stout, Ullman, & Schnellenberg, 1987; McDermott, Fantuzzo, & Glutting, 1990; Mueller, Dennis, & Short, 1986; Watkins & Kush, 1994).

Despite long-standing assumptions, cognitive profiles have consistently failed to demonstrate utility in predicting students' social and behavioral functioning (Beebe, Pfiffner, & McBurnett, 2000; Glutting et al., 1998; Lipsitz, Dworkin, & Erlenmeyer-Kimling, 1993; McDermott & Glutting, 1997; Riccio, Cohen, Hall, & Ross, 1997). Following a review of the research, Teeter and Korducki (1998) concluded that cognitive profiles are not diagnostically useful.

Cognitive test profiles have also been ineffectual in the diagnosis of learning disabilities. Systematic examination of ACID and SCAD profiles in addition to the Developmental Index and the Profile Variability Index demonstrated that they were little better than chance when used to differentiate children with exceptionalities (Watkins, 1996, 1999; Watkins, Kush, & Glutting, 1997a, b). Beyond diagnosis, cognitive profiles have not provided substantial incremental validity for prediction of academic achievement (Glutting, Youngstrom, Ward, Ward, & Hale, 1997; Hale & Saxe, 1983; Kline, Snyder, Guilmette, & Castellanos, 1992; Watkins & Glutting, 2000; Youngstrom, Kogos, & Glutting, 1999) and have not beneficially contributed via hypothetico-deductive approaches to interpretation (Kramer et al., 1987). A particularly blunt appraisal of the research literature was provided by Kavale and Forness (1984), who concluded that "profile and scatter analysis is not defensible" (p. 136).

EVIDENCE FROM THIS SPECIAL ISSUE

Although the existing research literature suggests that the empirical basis for cognitive profile analysis is inadequate to support competent practice, this spe-

cial issue is a forum for the presentation of additional evidence and research methods for studying profile analysis.

Measurable Biological Substrates

Riccio and Hynd (2000) hypothesized that children with leftward planum asymmetry ($L > R$) will show $VIQ \geq PIQ$ cognitive patterns whereas those children with right planum asymmetry ($R > L$) will be more likely to exhibit $PIQ > VIQ$ profiles. Participants in this study were primarily male (71.2%), Caucasian (96.2%), and right handed (88%). Additionally, participants were identified as (a) normal controls, (b) children with a reading disability, or (c) children with Attention Deficit/Hyperactivity Disorder. Although these physical and disability characteristics probably are related to VIQ - PIQ scores or planum temporale asymmetry, neither research design nor data analyses controlled for their influence. Thus, it is not possible to attribute outcomes unequivocally to the presumptive factors of planum asymmetry and VIQ - PIQ scores. Handedness, ethnicity, gender, or disability of the participants might be plausible alternative explanations for study outcomes.

Of the tests conducted by Riccio and Hynd (2000), the planum asymmetry measures ($L > R$, $R > L$) compared with the cognitive profile ($VIQ \geq PIQ$, $PIQ > VIQ$) are directly related to testing their hypothesis that VIQ - PIQ score differences are diagnostic of planum asymmetry. As recommended by Kessel and Zimmerman (1993), problems of this type should be analyzed with diagnostic utility statistics. Table 1 presents a succinct description of several diagnostic utility statistics and their results when applied to the data of Riccio and Hynd.

For these data, only 54% of the children identified by a $VIQ \geq PIQ$ test profile truly had $L > R$ planum asymmetry and only 33% of the children identified

TABLE 1. Diagnostic Utility Statistics Applied to Planum Asymmetry as Predicted by WISC-III Verbal and Performance IQ Scores

Statistic	Description	Value
Sensitivity	Proportion of participants with a target disorder who are identified by a positive test finding.	.56
Specificity	Proportion of participants free of the disorder who are correctly identified by a negative test result.	.31
Positive Predictive Power	Proportion of participants identified by a positive test score who truly have the target disorder.	.54
Negative Predictive Power	Proportion of participants identified by a negative test score who truly do not have the target disorder.	.33
κ	Proportion of agreement between the test and actual condition of the participants (disordered vs. non-disordered) beyond that accounted for by chance alone.	.12

by a PIQ > VIQ pattern truly had R > L planum asymmetry. Likewise, sensitivity and specificity were low. These statistics indicate that the VIQ-PIQ profile will produce many incorrect diagnostic decisions. This conclusion was not altered even when planum symmetry (L = R) was combined with leftward asymmetry (L > R) in the analysis. Consequently, the hypothesis that VIQ-PIQ profiles diagnostically relate to planum asymmetry was not confirmed and inferences concerning VIQ-PIQ scores and brain functioning are not supported by these data.

Modal WISC-III Profiles

Pritchard, Livingston, Reynolds, and Moses (2000) relied upon modal profile analysis (MPA) to operationalize their evaluation of WISC-III factor index scores. The MPA technique, as described by Skinner and Lei (1980), consists of three stages: (a) two samples are subjected to *Q*-factor analysis to determine the number of shape profiles within each sample; (b) profiles are compared across samples and those that replicate are retained as the final modal profiles; and (c) separate elevation, scatter, and shape parameters are calculated for each participant and the shape parameter is then used to classify new cases into the modal profile that they most closely resemble.

There are, however, several caveats concerning MPA and the study of Pritchard et al. (2000) that suggest their results may not be of appreciable utility. First, measurement experts often discourage use of *Q*-factor analysis to group people because the variance of a case "is fractionated and distributed among several factors" (Tinsley & Tinsley, 1987, p. 415) and "thus the set of factors is not mutually exclusive and exhaustive" (Bailey, 1994, p. 71). Additionally, when many people with few variables are subjected to *Q*-factor analysis, the real level of similarity may be overstated because of chance (Bailey, 1994).

An additional complication of this study is found in the method used by Pritchard et al. (2000) to extract their *Q*-factors. Considerable evidence shows that the eigenvalue > 1.0 criterion is inexact and will inaccurately estimate the number of factors to extract (Zwick & Velicer, 1986). Given the inherent methodological weaknesses of *Q*-factor analysis and the specific methods used in the current investigation, cluster analytic techniques would have been preferable.

Striking lacunae in the Pritchard et al. (2000) report are the failure to generate and apply elevation and scatter parameters in the third stage of the MPA analysis. This omission is inexplicable because one major benefit of MPA is that it allows comparison of the independent contribution of profile elevation, scatter, and shape parameters (Skinner, 1978; Skinner & Lei, 1980). Consequently, the only external validity evidence offered for the four shape-based factor index profiles was that they were not "parallel" with scores from the Wechsler Individual Achievement Test (WIAT). Stronger external validity evidence is required if WISC-III factor index profile shapes are to merit clinical attention.

Fortunately, other researchers have investigated similar questions and their results provide evidence about the external validity of profile elevation, scatter, and shape parameters. Directly applicable is a study conducted by Hale and Saxe (1983), who applied MPA to Wechsler Intelligence Scale for Children-Revised (WISC-R) subtest scores among a referral sample of 269 children. Four modal profiles were formed, cross-validated, and then systematically entered into a regression analysis with elevation and scatter scores to predict academic achievement. Results indicated that elevation information was the single, most important predictor, accounting for 32–38% of the variance in achievement. Scatter scores did not add any incremental predictive accuracy, but shape information accounted for an additional 8% of the variance in achievement. Based upon these results, Hale and Saxe (1983) suggested that the incremental validity of shape information “is not of practical importance” (p. 160).

A similar study was conducted by Kline et al. (1992). These researchers computed numerical indices of profile elevation, scatter, and shape from the WISC-R subtest scores of a referral sample of 146 children. They then evaluated with hierarchical multiple regression whether scatter and shape information had incremental predictive validity beyond elevation metrics in the prediction of academic achievement. Elevation information was again the single, most important predictor of achievement. Scatter scores did not add any incremental predictive accuracy, but shape information accounted for an additional 7–11% of the variance in achievement beyond that accounted for by subtest elevation. Kline and colleagues concluded that “profile shape information adds relatively little unique information, and therefore examiners should not overinterpret particular patterns of scores” (p. 431).

Additional evidence was provided by Watkins and Glutting (2000), who decomposed the WISC-III subtest profiles of 1,118 nonexceptional and 538 exceptional students into elevation, scatter, and shape components and sequentially regressed them onto reading and math achievement scores. Profile elevation was statistically and clinically significant for students with and without exceptionalities. Profile scatter did not aid in the prediction of achievement but profile shape accounted for an additional 5–8% of the variance in achievement measures. Further investigation of shape profiles found that only two contributed to academic prediction: one shape with relatively high verbal subtest scores positively predicted both reading and math achievement and one profile marked by a relatively depressed math subtest score was negatively related to math achievement.

The results reported by Watkins and Glutting (2000) comport closely with those reported by Hale and Saxe (1983) and Kline et al. (1992). Beyond two robust, but somewhat uninformative profiles, WISC-III subtest profile scatter and shape information had inconsequential incremental validity for predicting reading and math achievement for students with and without exceptionalities. Consequently, using elevation information to predict academic achievement is the most parsimonious procedure.

Concerning internal validity, Pritchard et al. (2000) noted that 63% of their

overall WISC-III factor profile variance was accounted for by profile elevation. Profile scatter only accounted for 0.3% of the variance in WISC-III index scores, whereas 37% of the WISC-III variance was because of differences in profile shape plus measurement error (they could not statistically disentangle error and shape variance). Thus, profile elevation was the primary internal influence on their profiles. Pritchard and colleagues (2000) agreed that these internal variance estimates were congruent with those reported by other researchers, but asserted that this “does not imply that profile shape is of little utility” (p. 416). However, the consistent lack of support from incremental predictive validity reports, combined with weak internal validity, does provide strong evidence that profile scatter and shape have little utility.

Profile Analysis with the Cognitive Assessment System

Naglieri (2000) argued that profile analysis as ordinarily practiced with subtests is uninformative because most cognitive instruments are not based on strong theory and because subtests possess inadequate reliability and validity for the task. Naglieri suggested, however, that profile analysis could be conducted fruitfully with the Cognitive Assessment System (CAS; Naglieri & Das, 1997). This instrument is based on Planning, Attention, Simultaneous, Successive (PASS) cognitive theory, which provides the putative strong theoretical foundation. Naglieri indicated that CAS factor scores, rather than subtest scores, provide sufficient reliability and validity for profiles to be used to make diagnostic and intervention decisions.

By Naglieri's (2000) standards, the WISC-III is not amenable to profile analysis because the Wechsler scales were developed without a sound theoretical rationale and its subtests lack sufficient reliability and construct validity. Although not founded in theory, the WISC-III produces factor scores that provide enhanced construct validity and exhibit incremental reliability in comparison to its subtests (i.e., M factor reliability = .896, M subtest reliability = .796). Thus, a test of Naglieri's assertion that profile analysis works uniquely with the CAS can be conducted by applying his profile analysis methods to the WISC-III.

Naglieri (2000) distinguished three interrelated procedures to define cognitive profiles. First, Relative Weakness (RW) profiles rely on ipsative assessment methods that have been advanced as rich alternatives to normative assessment and are popularized in leading assessment texts (Kaufman, 1994; Sattler, 1992). Second, Cognitive Weakness (CW) methods require that at least one of the CAS factor scores be below a specified cut score and an RW exist concurrently. Finally, Cognitive Weakness and Academic Weakness (CWA) methods require that an RW and CW exist in addition to a low normative score on an academic achievement test.

Table 2 provides the results of applying Naglieri's (2000) profile methods (with cut scores of 80) to the 1,118 nationally representative students in the WISC-III/WIAT linking sample. These results can be compared with CAS out-

TABLE 2. WISC-III and WIAT Standard Score Means for Students with a Relative Weakness, Cognitive Weakness, Cognitive and Academic Weakness, Low IQ Score, Low IQ Score and Low Academic Score, Low Achievement, and Ability-Achievement Discrepancy in Reading, Math, or Spelling.

	RW	CW	CWAW	IQ	IQ-AW	AW	A-AD
<i>N</i>	299	112	45	119	69	204	53
Percent	26.7	10.0	4.0	10.6	6.2	18.2	4.7
Verbal Comprehension	98.5	86.0	77.9	74.8	71.5	81.6	107.2
Perceptual Organization	101.2	89.1	86.9	74.6	73.5	88.1	116.8
Freedom from Distractibility	101.2	88.0	81.0	80.7	77.7	84.4	99.9
Perceptual Speed	104.2	96.0	97.7	86.1	84.4	91.6	105.8
Full Scale IQ	100.0	86.1	80.6	72.2	69.8	82.4	111.3
Basic Reading	101.8	93.2	84.7	85.5	79.5	86.0	93.0
Reading Comprehension	101.4	92.4	83.1	83.9	79.0	84.2	96.2
Numerical Operations	101.3	92.2	82.0	84.0	76.5	84.8	91.2
Math Reasoning	102.2	90.6	80.0	82.5	76.7	84.6	98.5
Spelling	101.5	93.9	84.2	87.0	80.6	85.9	89.2

Note: *N* = 1,118. A-AD, Ability-Achievement Discrepancy in Reading, Math, or Spelling; AW, Low Achievement; CW, Cognitive Weakness; CWAW, Cognitive and Academic Weakness; IQ, Low IQ Score; IQ-AW, Low IQ Score and Low Academic Score; RW, Relative Weakness.

comes presented by Naglieri in his Tables 3 and 4. Naglieri found that 42%, 13.5%, and 6.5% of his participants exhibited RW, CW, and CWAW patterns, respectively, whereas the corresponding percentages from the WISC-III sample were 26.7%, 10%, and 4%, respectively. The presence or absence of an RW or CW correlated with academic achievement at .01 and .21, respectively, in the WISC-III sample and .01 and .30, respectively, in the CAS sample. As with the CAS, IQ and achievement scores steadily declined across RW, CW, and CWAW selections (*M* FSIQ = 100, 86.1, and 80.6, respectively; *M* achievement = 101.6, 92.5, and 82.8, respectively). Results from the CAS and WISC-III are similar in magnitude and are conceptually identical.

Naglieri (2000) also applied RW and CW methods to a sample of 144 students enrolled in special education programs and concluded that "children with cognitive weaknesses are likely to be similar to those children independently placed in special educational settings" (p. 432). To test further the unique specificity of RW and CW methods to the CAS, they were applied to the WISC-III among a sample of 537 students enrolled in special education programs (LD = 468, ED = 40, MR = 29) in three suburban, southwestern school districts. These students were extracted from the sample of Watkins (1996) contingent upon complete data for all WISC-III factor indices and achievement test composite scores.

As with the CAS, χ^2 tests found that RW did not significantly differentiate students placed in special education from those placed in regular education settings, whereas both CW and CWAW produced significantly different propor-

tions between groups. However, tests of proportions are inappropriate for determining diagnostic utility (Elwood, 1993). The diagnostic utility statistics presented in Table 3 provide a more accurate picture. It is apparent from scrutiny of these statistics that neither the CAS nor the WISC-III are accurate in identifying special education cases. The CAS appears to be somewhat more efficient in negative predictive power but low base rates make all predictions risky (Meehl & Rosen, 1955). For example, κ coefficients (Cohen, 1960) of .03 to .22 indicated "poor" agreement (Cicchetti, 1994) between the tests and actual diagnostic category of the students beyond that accounted for by chance alone.

For comparison purposes, four additional methods of identifying exceptional students were applied to the WISC-III samples: (a) a low IQ (IQ) was defined as a FSIQ < 80; (b) a low IQ and low achievement (IQ-AW) was comprised of FSIQ < 80 and an achievement test composite score < 80; (c) low achievement was defined by an achievement test composite score < 80; and (d) an ability-achievement discrepancy (A-AD) was composed of an achievement test composite score that was more than 1.96 standard errors of estimate from a predicted level of achievement given the FSIQ score (i.e., a regression criterion). Tables 2 and 3 illustrate that these additional identification methods demonstrated improved diagnostic accuracy as they became more selective, but none was precise enough to be applied with confidence to individual students. This observation emphasizes the public policy rather than scientific foundation of the categorical special education system (Stanovich, 1999).

TABLE 3. Positive Predictive Power (PPP), Negative Predictive Power, Sensitivity, Specificity, and Kappa of the Cognitive Assessment System and Wechsler Intelligence Scale for Children-Third Edition for Relative Weakness, Cognitive Weakness, Cognitive and Academic Weakness, Low IQ Score, Low IQ and Low Academic Score, Low Achievement, and Ability-Achievement Discrepancy in Reading, Math, or Spelling

	PPP	NPP	Sensitivity	Specificity	κ
CAS					
Relative Weakness	.11	.92	.51	.58	.03
Cognitive Weakness	.19	.93	.32	.87	.14
Cognitive and Academic Weakness	.30	.93	.28	.94	.22
WISC-III					
Relative Weakness	.30	.67	.24	.73	.03
Cognitive Weakness	.45	.69	.17	.90	.09
Cognitive and Academic Weakness	.63	.70	.14	.96	.13
Low IQ	.54	.72	.26	.89	.18
Low IQ and Low Achievement	.66	.72	.25	.94	.22
Low Achievement	.67	.88	.76	.82	.56
Ability-Achievement Discrepancy	.80	.77	.40	.95	.40

Note: CAS, Cognitive Assessment System; NPP, Negative Predictive Power; PPP, Positive Predictive Power; WISC-III, Wechsler Intelligence Scale for Children-Revised.

These CAS/WISC-III comparisons neither confirm the hypothesis that Naglieri's (2000) profile methods perform uniquely with the CAS, nor do they support the contention that these profile methods identify exceptional children with accuracy. Given the research of McDermott et al. (1990) on ipsative subtest profiles, these disconfirmations were not unexpected. Specifically, when ipsative assessment methods are used with cognitive subtests and composites, they remove the bulk of variance contributed by the general intelligence factor (g) and consequently lose much of their power to predict academic achievement (Jensen, 1992; McDermott et al., 1990; McDermott & Glutting, 1997).

Accordingly, CAS ipsative scores did not forecast achievement test scores (i.e., $r = .01$). Conversely, when a low CAS factor score was required to become a member of the CW group, some variation because of general intelligence (g) was returned to the prediction (i.e., $r = .30$). By requiring both low achievement and low ability scores, the CWA grouping added an additional, unpredictable amount of g variance because the low ability test score tapped g and the low achievement score reflected an accumulation of previous g -demanding learning (Jensen, 1998). This progressive infusion of g is also clearly illustrated by the WISC-III/WIAT data, where the presence of an RW, CW, or CWA was correlated with mean achievement at .01, .21, and .27, respectively. Thus, "No" is the correct answer to Naglieri's query, "Can profile analysis of test scores work?"

CFA

Stanton and Reynolds (2000) applied a variant of CFA (von Eye, Spiel, & Wood, 1996) to the WISC-R normative sample. This method of dichotomizing continuous WISC-R subtest scores was ipsative because subtests were subtracted from mean composite scores and thereby transformed into person-relative metrics from their original population-relative metric (McDermott et al., 1990). For example, two hypothetical students' normative (population-relative) and ipsative (person-relative) WISC-R verbal scores are displayed in Table 4. Students One and Two share an identical ipsative profile (coded 000001), whereas their normative scores are very different. This ipsative dichotomization ignores normative elevation information.

After creating ipsatized subtest profiles, Stanton and Reynolds (2000) compared the relative frequency of each profile within normative and exceptional student populations to identify those profiles that occurred more often among students with learning disabilities. Using this methodology, two types or patterns were determined to be associated with learning disabilities (LDs): (a) one with relatively depressed scores on Arithmetic, Digit Span, and Coding subtests; and (b) a second with relatively low Arithmetic and Coding scores. Stanton and Reynolds noted that these patterns are consistent with subtest groupings proposed by Kaufman (1979). However, previous empirical investigations have demonstrated that these patterns are inaccurate diagnostic indicators of LD (Watkins, 1996, 1999; Watkins et al. 1997a, 1997b).

TABLE 4. Normative (Population-Relative) and Ipsative (Person-Relative) WISC-R Verbal Subtest Scores for Two Hypothetical Students

Subtest	Student One			Student Two		
	Norm Score	Ipsative Score	Ipsative Code	Norm Score	Ipsative Score	Ipsative Code
Information	11	+1	0	17	+2	0
Similarities	11	+1	0	16	+1	0
Arithmetic	11	+1	0	15	0	0
Vocabulary	11	+1	0	17	+2	0
Comprehension	11	+1	0	15	0	0
Digit Span	5	-5	1	10	-5	1
Mean	10	0	—	15	0	—

Moreover, reliance on differential prevalence rates is insufficient for determining diagnostic utility (Elwood, 1993). Diagnostic decisions must be informed, at a minimum, by knowledge of the sensitivity, specificity, positive predictive power, and negative predictive power of a test (Zarin & Earls, 1993). Sensitivity (.02-.03 probability of accurately identifying a student with LD) and positive predictive power (.23-.40 probability that a student with a subtest profile is really LD) were low for the two types of WISC-R profiles identified by Stanton and Reynolds (2000). A positive test profile was more likely to misdiagnose a normative group student than to identify properly a student with LD. Conversely, specificity and negative predictive power rates were high ($\geq .93$), indicating that a negative test profile was likely to diagnose normative group students. However, blind categorization of all 2,374 students as non-exceptional, regardless of their subtest profile, would accurately diagnose all 2,200 normative group students for an overall correct classification rate of 93%. Thus, identification of non-exceptional students is easily accomplished by following base rates and not substantially improved by knowledge of subtest profiles. Overall, agreement between subtest profiles and the actual diagnostic category of the participants beyond that accounted for by chance was only .04. These diagnostic statistics do not inspire confidence and suggest that the subtest profiles should not be interpreted (Mossman & Somoza, 1991).

As previously described, transformation of normative to ipsative scores is problematic. Jensen (1992), for example, concluded that subtest analysis is "practically worthless" (p. 276) because ipsative scores, by definition, remove general ability variance. Jensen's assessment was corroborated by McDermott et al. (1990), who demonstrated that the ipsatization of WISC-R scores produced a loss of almost 60% of that test's reliable variance. Thus, the low estimate of α reliability (.59) reported by Stanton and Reynolds (2000) could have been anticipated.

Given the limitations of ipsative scores, it can be hypothesized that the temporal stability of significant subtest profiles formed by ipsatization will be seri-

ously compromised. This was investigated with a national sample of 277 students (73% male, 27% female) with LD who were administered the WISC-III twice as part of special education eligibility evaluations (M time between tests = 2.9 years). Students were selected from the sample of Watkins and Canivez (2000) contingent upon diagnosis of LD and complete scores on 11 subtests. The WISC-III subtest scores earned by 277 students with LD at both test and retest were categorized into one of the 30 profiles listed in Table 2 of Stanton and Reynolds (2000). Those students whose WISC-III subtests did not fit one of the 30 profiles were placed into an "other" category to achieve universal categorization.

These subtest profiles were extremely unstable over time. The κ statistic revealed that there was only 12% agreement beyond chance across test and retest. For the two significant profile types identified by Stanton and Reynolds (2000), partial κ statistics (Fleiss, 1971) were $-.02$ and $.01$. This "poor" level of agreement (Cicchetti, 1994) on ipsative subtest profiles was in contrast with a dichotomization where $FSIQ < 85$ were coded as 0 and $FSIQ \geq 85$ were coded as 1. When $FSIQ$ test and retest scores were categorized in this manner, a κ of $.65$ resulted, and the temporal stability of ipsative subtest profiles, as hypothesized, was severely compromised.

Although admitting that their profile configurations demonstrated low reliability, Stanton and Reynolds (2000) asserted that this "does not provide a convincing argument against their clinical use." However, low internal consistency, poor temporal stability, and weak diagnostic efficiency do provide strong evidence that these ipsative subtest profiles have little utility.

CONCLUSION

Interpretation of cognitive test profiles is popular in school psychology training and practice but is not supported by research evidence. This special issue contains four studies that provide additional data regarding cognitive test profiles. Interestingly, only Stanton and Reynolds (2000) chose to analyze cognitive subtest profiles. The other three studies examined profiles composed of more reliable cognitive composites. Even among advocates of cognitive profile analysis, it appears that factors or other composites are preferred over subtests.

Detailed analyses found all four cognitive profile reports lacking in terms of reliability, validity, or diagnostic utility. This conclusion is consistent with the largely negative research literature. Even cognitive profiles composed of factor scores were psychometrically weak. These results were not surprising because ipsative assessment is not effective when a general factor accounts for much of the variance of a test (Thorndike, 1994).

Suggestions that unreliable and invalid cognitive test profiles somehow become useful when integrated informally and subjectively with a complex variety of other assessment data are suggested despite an overwhelmingly negative research literature (Dawes, Faust, & Meehl, 1989). It is well established that clini-

cians are most susceptible to decision errors in exactly this type of situation (Aspel et al., 1998; Davidow & Levinson, 1993; Faust, 1986; Grove & Meehl, 1996). As summarized by Faust (1990), this "common belief in the capacity to perform complex configural analysis and data integration might thus be appropriately described as a shared professional myth" (p. 478).

Given these skeptical conclusions, it is fitting to reiterate Weiner's (1989) admonition that psychologists should: "(a) know what their tests can do and (b) act accordingly" (p. 829). This paper described what cognitive tests cannot do: Cognitive ability profiles are not particularly useful for diagnosis, prediction of academic performance, or differential intervention. Fortunately, there is also ample evidence about what cognitive tests can do (Neisser et al., 1996). For example, the substantial relation between general intelligence and school persistence and achievement is one of the best documented findings in psychology and education (Jensen, 1998). Beyond that, general cognitive ability is the best single predictor of job performance across a wide range of occupations and professions (Hunter, 1986).

Although it is appropriate for researchers to continue investigating complex patterns of aptitudes (Reynolds, 1988), scientific psychological practice cannot be sustained by clinical conjectures and personal anecdotes that have consistently failed empirical validation. Consequently, psychologists should eschew interpretation of cognitive test profiles and must accept that they are acting in opposition to the scientific evidence if they engage in this practice.

REFERENCES

- Alfonso, V. C., Oakland, T. D., LaRocca, R., & Spanakos, A. (2000). The course on individual cognitive assessment. *School Psychology Review, 29*, 52-64.
- American Psychological Association. (2000). *Guidelines and principles for accreditation of programs in professional psychology*. Washington, DC: Author.
- Aspel, A. D., Willis, W. G., & Faust, D. (1998). School psychologists' diagnostic decision-making processes: Objective-subjective discrepancies. *Journal of School Psychology, 36*, 137-149.
- Bailey, Q. E. D. (1994). *Typologies and taxonomies: An introduction to classification techniques*. Thousand Oaks, CA: Sage.
- 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.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290.
- Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.
- Davidow, J., & Levinson, E. M. (1993). Heuristic principles and cognitive bias in decision making: Implications for assessment in school psychology. *Psychology in the Schools, 30*, 351-361.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668-1674.
- Dumont, R., & Willis, J. O. (1995). Intrasubtest scatter on the WISC-III for various clinical samples vs. the standardization sample: An examination of WISC folklore. *Journal of Psychoeducational Assessment, 13*, 271-285.

- Elwood, R. W. (1993). Psychological tests and clinical discriminations: Beginning to address the base rate problem. *Clinical Psychology Review, 13*, 409–419.
- Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice, 17*, 420–430.
- Faust, D. (1990). Data integration in legal evaluations: Can clinicians deliver on their premises? *Behavioral Sciences and the Law, 7*, 469–483.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*, 378–382.
- 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., 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? *Journal of Psychological Assessment, 9*, 295–301.
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2*, 293–323.
- Hale, R. L., & Saxe, J. E. (1983). Profile analysis of the Wechsler Intelligence Scale for Children—Revised. *Journal of Psychoeducational Assessment, 1*, 155–162.
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340–362.
- Jensen, A. R. (1992). Commentary: Vehicles of *g*. *Psychological Science, 3*, 275–278.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Kaufman, A. S. (1979). *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.
- Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self-administered questionnaires: Extent of the problem, recommendations for standardized presentation of results, and implications for the peer review process. *Psychological Assessment, 5*, 395–399.
- 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.
- Kramer, J. J., Henning-Stout, M., Ullman, D. P., & Schnellberg, R. P. (1987). The viability of scatter analysis on the WISC-R and the SBIS: Examining a vestige. *Journal of Psychoeducational Assessment, 5*, 37–47.
- Lipsitz, J. D., Dworkin, R. H., & Erlenmeyer-Kimling, L. (1993). Wechsler Comprehension and Picture Arrangement subtests and social adjustment. *Psychological Assessment, 5*, 430–437.
- Loftus, E. (1998). Who is the cat that curiosity killed? *Skeptical Inquirer, 22*, 60–61.
- 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.
- McDermott, P. A., & Glutting, J. J. (1997). Informing stylistic learning behavior, disposition, and achievement through ability subtests—Or, more illusions of meaning? *School Psychology Review, 26*, 163–175.
- Meehl, P. E. (1997). Credentialed persons, credentialed knowledge. *Clinical Psychology: Science and Practice, 4*, 91–98.
- Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194–216.
- Mossman, D., & Somoza, E. (1991). Neuropsychiatric decision making: The role of disorder prevalence in diagnostic testing. *Journal of Neuropsychiatry and Clinical Neurosciences, 3*, 84–88.

- 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.
- Naglieri, J. A. (2000). Can profile analysis of ability test scores work? An illustration using the PASS theory and CAS with an unselected cohort. *School Psychology Quarterly, 15*, 419-433.
- Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System*. Chicago: Riverside.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77-101.
- 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.
- Pritchard, D. A., Livingston, R. B., Reynolds, C. R., & Moses, J. A. (2000). Modal profiles for the WISC-III. *School Psychology Quarterly, 15*, 400-418.
- Reynolds, C. R. (1988). Putting the individual into aptitude-treatment interaction. *Exceptional Children, 54*, 324-331.
- 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.
- Riccio, C. A., & Hynd, G. W. (2000). Measurable biological substrates to verbal-performance differences in Wechsler scores. *School Psychology Quarterly, 15*, 386-399.
- Ruscio, J. (1998). The perils of post-hockery. *Skeptical Inquirer, 22*, 44-48.
- Sattler, J. M. (1992). *Assessment of children: Revised and updated third edition*. San Diego, CA: Author.
- Skinner, H. A. (1978). Differentiating the contribution of elevation, scatter, and shape in profile similarity. *Educational and Psychological Measurement, 38*, 297-308.
- Skinner, H. A., & Lei, H. (1980). Modal profile analysis: A computer program for classification research. *Educational and Psychological Measurement, 40*, 769-772.
- Stanovich, K. E. (1999). The sociopsychometrics of learning disabilities. *Journal of Learning Disabilities, 32*, 350-361.
- Stanton, H. C., & Reynolds, C. R. (2000). Configural frequency analysis as a method of determining Wechsler profile types. *School Psychology Quarterly, 15*, 434-448.
- 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. (1994). *g. Intelligence, 19*, 145-155.
- Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology, 34*, 414-424.
- Tracey, T. J., & Rounds, J. (1999). Inference and attribution errors in test interpretation. In J. W. Lichtenberg & R. K. Goodyear (Eds.), *Scientist-practitioner perspectives on test interpretation* (pp. 113-131). Boston: Allyn & Bacon.
- von Eye, A., Spiel, C., & Wood, P. K. (1996). Configural frequency analysis in applied psychology research. *Applied Psychology: An International Review, 45*, 301-352.
- 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., & Canivez, G. L. (2000, August). *Temporal stability of the WISC-III factor structure among disabled students*. Poster session presented at the 108th annual meeting of the American Psychological Association, Washington, DC.
- Watkins, M. W., & Glutting, J. J. (2000). Incremental validity of WISC-III profile elevation, scatter, and shape information for predicting reading and math achievement. *Psychological Assessment*.

- Watkins, M. W., & Kush, J. C. (1994). WISC-R subtest analysis: 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). 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., & Glutting, J. J. (1997b). Discriminant and predictive validity of the WISC-III ACID profile among children with learning disabilities. *Psychology in the Schools, 34*, 309–319.
- Weiner, I. B. (1989). On competence and ethicality in psychodiagnostic assessment. *Journal of Personality Assessment, 53*, 827–831.
- Youngstrom, E. A., Kogos, J. L., & Glutting, J. J. (1999). Incremental efficacy of differential ability scales factors in predicting individual achievement criteria. *School Psychology Quarterly, 14*, 26–39.
- Zarin, D. A., & Earls, F. (1993). Diagnostic decision making in psychiatry. *American Journal of Psychiatry, 150*, 197–206.
- Ziskin, J. (1995). *Coping with psychiatric and psychological testimony* (5th ed.). Los Angeles: Law and Psychology Press.
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432–442.

Action Editor: Timothy Z. Keith

Acceptance Date: August 31, 2000