School Psychology Review
1998, Vol. 27, No. 4, pp. 599-612

More Ups and Downs of Subtest Analysis: Criterion Validity of the DAS with an Unselected Cohort

Joseph J. Glutting
University of Delaware

Paul A. McDermott
University of Pennsylvania

Timothy R. Konold
University of Virginia

Alisa J. Snelbaker
University of Delaware

Marley W. Watkins
Pennsylvania State University

Abstract: This research evaluated the criterion-related validity of unusual subtest profiles from the Differential Ability Scales (Elliott, 1990). Three methods were used to identify unusual profiles: multivariate-nomothetic, univariate-nomothetic, and univariate-ipsative prevalence rates $\leq 5\%$. Participants were a large cohort ($N = 1{,}200$) stratified, within age levels, in proportion to U.S. Census data on demography (e.g., race, gender, parents' educational levels). From this cohort, children with unusual profiles were identified ($n = 60$) and matched to controls ($n = 60$) by the characteristics listed and overall IQs. The two groups were compared across a variety of external criteria: (a) propensity for placement in special education, (b) three norm-referenced measures of achievement, and (c) six behavioral indices evaluated through standardized teacher ratings. Results showed no group differences across all criteria, regardless of the method used to identify unusual subtest profiles. Findings are discussed in the context of the criterion validity of specific ability constructs.

IQ tests have been criticized almost from their inception (for reviews, see Jensen, 1980; Kamin, 1984). Nevertheless, they remain one of the most popular psychological measures given to children and adolescents (Stinnett, Harvey, & Oehler-Stinnett, 1994; Wilson & Reschly, 1996). The practical justification for IQ testing is, in large part, due to the tests' criterion-related validity.

The substantial relationship between general intelligence and school achievement is perhaps the most documented finding in psychometric psychology and education (American Psychological Association, Board of Scientific Affairs, 1996). This relationship remains invariant, regardless of a child's cultural background, gender, or socioeconomic status (Jensen, 1980; Reynolds & Kaiser, 1990). Likewise, nearly a century of evidence places global ability among the most dominant predictors of the years of formal education children are likely to receive, adults' social status and income, and job performance (Cronbach & Snow, 1977; Hunter, 1983; Jencks, 1972; White, 1982). Global ability also shows significant, but more moderate, criterion validity for personality and social dispositions such as the occurrence of conduct disorders and juvenile delinquency (Kazdin, 1995; Moffitt, Gabrielli, Mednick, & Schulsinger, 1981).

Clinicians who work with children and adolescents tend to find that global scores from intelligence tests have limited relevance. Their alternative is to expound positions advanced in leading textbooks on children's intelligence testing. Therein, ability is viewed as a multi-differentiated construct whose greatest value lies in the extent to which IQ tests enable us to discern individual profiles of specific abilities

Address all correspondence concerning this article to Joseph Glutting, Department of Educational Studies, School of Education, Room 221B Willard Hall, University of Delaware, Newark, DE 19716.

(Kamphaus, 1993; Kaufman, 1994; Sattler, 1988, 1992). Consequently, within this context, subtest analysis is considered to be the most finely grained and sophisticated approach to the discovery of children's differential aptitudes. The goal of subtest analysis is to identify telltale patterns of cognitive strengths and weaknesses that may be important to remediation and/or differential diagnosis.

Validity Issues

Multiple sources of evidence can be used to validate test-score interpretations (Messick, 1989). However, in diagnostic assessment, two types of evidence are primary. Diagnostic, score-based interpretations become valid to the extent that they (a) are associated with a viable *treatment* for individuals suffering from a disorder, or (b) accurately *predict* (either concurrently or in the future) a high probability that an individual will contract a problem or disorder (Cromwell, Blashfield, & Strauss, 1975; Gough, 1971; McDermott, 1981).

Treatment Validity

For some time now, psychologists have been operating as though treatment validity is the most important evidence for intelligence tests. This situation is unfortunate because it occurs at the expense of prediction. Prediction is valuable in its own right because we may never be able to remediate all of the negative circumstances that can impact children's growth and well-being. Moreover, with the exception of findings for global ability, treatment validity remains very much in doubt for more differentiated ability profiles, with research consistently demonstrating few positive outcomes for *multiple* aptitude x treatment interactions (Cronbach & Snow, 1977; Heller, Holtzman, & Messick, 1982; Ysseldyke & Christenson, 1988).

Predictive/Concurrent Validity

The importance of prediction is well-established for profile validity. For instance, experts in clinical assessment routinely encourage profile analysis as a mechanism for generating hypotheses (i.e., predictions) about how processing strengths and weaknesses observed in subtest profiles are likely to impact children's achievement and personal adjustment (Kamphaus, 1993; Kaufman, 1994; Sattler, 1992). Similarly, profile analysis is regarded as essential to the differential diagnosis (i.e., prediction) of certain disabilities covered in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) (DSM-IV) (American Psychiatric Association, 1994). Specifically, DSM-IV criteria for mental retardation state:

> When there is significant scatter in the subtest scores, the profile of strengths and weaknesses, rather than the mathematically derived full-scale IQ, will more accurately reflect the person's learning abilities. When there is a marked discrepancy across verbal and performance scores, averaging to obtain a full-scale IQ score can be misleading. (p. 40)

Previous Research on Subtest Profiles

The diagnostic and/or predictive validity of IQ subtest profiles has been investigated repeatedly with researchers attempting to establish relationships between subtest profiles and meaningful external criteria. Thus, direct comparisons of groups of diversely diagnosed children have lead some researchers to conclude that subtest profiles are helpful in differentiating among groups who are emotionally, mentally, and learning impaired and among groups experiencing specific forms of underachievement, conduct disorders, and attention-deficit/hyperactivity disorders (Bowers et al., 1992; Plante & Sykora, 1994; Prifitera & Dersh, 1993; Rourke & Strang, 1984; Schwean, Saklofske, Yackulic, & Quinn, 1993; Teeter & Smith, 1993; Wielkiewicz, 1990; Wielkiewicz & Daood, 1993). In contrast, other researchers have found subtest profiles to be ineffective in predicting criterion variables, and in identifying children with recognized exceptionalities (Glutting, & Bear, 1989; Glutting, McGrath, Kamphaus, & McDermott, 1992; Humphries & Bone, 1993; Kavale & Forness, 1984; Kline, Snyder, Guilmette, & Castellanos, 1992; Kramer, Henning-Stout, Ullman, & Schnellenberg, 1987; Mueller, Dennis, & Short, 1986; Watkins & Kush, 1994).

That review conveys the strong impression that there is equivocal validity to hypotheses generated on the basis of subtest profiles. The problem is that the review pays no attention to the *quality* of inquiry underlying the various studies. Across a series of recent investigations, we identified several methodological problems that operate to negate or render uninterpretable nearly all research on children's ability profiles (Glutting, Konold, McDermott, Watkins, & Kush, in press; McDermott, Fantuzzo, & Glutting, 1990; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992). Among the methodological problems we identified, two predominate.

Repercussions of Circular Evidence and Inverse Probabilities

Circular Reasoning

The first methodological issue 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. This problem is one of self-selection and it is a limitation that even undergraduate textbooks on research methodology warn against. The remedy is to begin with unselected cohorts (i.e., representative samples, a proportion of which may be receiving special education), identify children with and without unusual subtest profiles, and subsequently compare their performance on external criteria (cf. Glutting et al., in press).

Inverse Probabilities

The second problem relates to inverse probabilities. Two scenarios will help illuminate this confound. The first is the customary method used to validate subtest profiles. Here, children are selected for study when they are known to exhibit a theoretically interesting disorder (e.g., a certain subtype of learning disability). Membership in the target group is presumed to be associated with unusual subtest variation. This group is then contrasted against either children with another disorder (e.g., a type of emotional disturbance) or those who are free from recognized problems (i.e., children who function normally).

Alternatively, the typical situation facing clinicians is just the opposite: determining the probability that a referred (but unclassified) child has a certain disorder *given* that he or she obtained an unusual subtest profile. Under most circumstances, the two scenarios are not equivalent, and empirical studies demonstrate that the latter, inverse probability, can have a profound, negative impact on the validity of personality tests (Elwood, 1993; Sines, 1966; Wiggins, 1973). As remarkable as it may appear, an unusual profile configuration can discriminate between previously identified groups, yet be incapable of concurrently estimating who has a personality disorder!

Sines (1966) anticipated the dual problems of circular reasoning and probabilities. He suggested that we validate psychological tests in just the reverse of their usual order. In other words, rather than our near exclusive concentration on exceptional samples, children from the general population would *first* be classified on the basis of their obtained score configurations (i.e., groups with unusual vs. common profiles). *Thereafter*, the groups would be compared across a variety of criteria, including placement in specific criterion groups (e.g., learning disabled, conduct disordered, normal functioning, etc.).

Likewise, in a previous special issue of the *School Psychology Review*, Glutting, McDermott, Watkins, Kush, and Konold (1997) recommended that future research on IQ subtest profiles employ concomitant use of (a) *heterogenous samples* (i.e., unselected cohorts com-prising children from special education *and* regular education) and (b) *criterion-related methodology* (i.e., either concurrent or longitu-dinal research designs). Implementation of these two procedures is the only effective remedy to the circular reasoning, inverse probabilities, and host of other methodological limitations that limit current inquiry on ability profiles.

An example will assist in clarifying the importance of employing the two proposed methods. Assume, for instance, that an unselected cohort is administered a diagnostic medical test. The test identifies one group as being HIV positive. These individuals are compared to a control group matched on background characteristics and followed longitudinally. Results from the study would inevitably show that the diagnostic sign (i.e., test result) has clinical utility because it *predicts* AIDs (the criterion) and the unfortunate consequences of this disorder and that it did so regardless whether or not individuals were referred for medical services.

Only one study of subtest profiles has avoided the dual problems of circular reasoning and inverse probabilities. Glutting et al. (1992) investigated subtest scores from the Kaufman Assessment Battery for Children (K-ABC) (Kaufman & Kaufman, 1983). They ascertained if an unselected cohort with unusual profiles was more likely to receive special education than a control group without unusual profiles. Results raised serious concerns about beliefs in the superiority of specific to general ability constructs. Children with learning disabilities (LD) and emotional disturbance (ED) were no more likely to show unusual subtest configurations than the control group.

The study by Glutting et al. (1992) is noteworthy because it avoided the two most prominent confounds that affect research on ability patterns. At the same time, the study was limited in a number of respects. First, it was confined to examining subtest validity from a single IQ test, the K-ABC. Second, it evaluated the contribution of subtest profiles to the differential diagnosis of LD and ED and did not explore if profiles were able to predict performance on other important criteria such as children's academic achievement or personal-social adjustment. Third, unusual subtest profiles were identified solely according to a multivariate method. More popular approaches to subtest analysis (ipsative score comparisons and univariate prevalence-base rates) were not employed.

Current Study

The Differential Ability Scales (DAS; Elliott, 1990) presents evidence that it measures a broader array of abilities than most other intelligence batteries. The DAS also contains subtests with the highest specificities of any ability measure (Elliott, 1990). High levels of reliable, specific variance are a precursor to identifying accurately differences between subtest profiles (Flanagan, Andrews, & Genshaft, 1997; McDermott et al., 1992). According to Elliott (1990), "The DAS was designed primarily as a profile test. That is, it should yield reliable, focused, and interpretable scores at the cluster or subtest level" (p. 385). Moreover, two recent studies demonstrated that the DAS is able to identify distinct subtest profiles in individuals and groups of children with LD (Kercher & Sandoval, 1991; McIntosh & Gridley, 1993).

In light of those findings, some researchers have asserted that DAS subtest scores have utility for the purposes of prediction and/or differential diagnosis (cf. Elliott, 1997). The present research examines the criterion validity of unusual subtest profiles from the DAS. It does so in the context of a large, stratified sample (i.e., an unselected cohort). Subgroups were identified with and without unusual subtest profiles. The subgroups were then evaluated on multiple external criteria. As indicated by Sines (1966), we are *not* interested in learning whether children with known disorders have elevated rates of unusual subtest profiles; rather we are interested in knowing if unusual subtest profiles signal abnormal clinical status (e.g., placement in programs for learning disabilities, emotional disturbance, etc.). Furthermore, the study attempts to answer the question, Do unusual subtest profiles portend adverse consequences across two of the most important outcomes of child development and well-being—performance on standardized measures of academic achievement and teacher ratings of classroom adjustment and behavior?

Method

Participants

The sample comprised all children (N = 1,200) who participated in the validation effort of the Adjustment Scales for Children and Adolescents (McDermott, Marston, & Stott, 1993). The sample was designed to represent the population of all noninstitutionalized children ages 6 years, 0 months through 17 years, 11 months residing in the United States at the beginning of the 1990 decade. Children were obtained from 201 school systems located in 70 U.S. Census statistical areas (metropolitan, suburban, rural) across four regions of the country (Northeast, Midwest, South, West). Within each age level, the sample conformed to parameters of the 1990 U.S. Census (U.S. Department of Commerce, 1990) for the variables of gender, race, grade level, geographic region, and mothers' and fathers' educational attainments. Detailed descriptions of the sample and its conformity to Census projections are provided by McDermott (1993).

Instruments

Predictor

Differential Ability Scales (DAS; Elliott, 1990). The DAS is a cognitive assessment battery designed to be individually administered to children between the ages of 2-1/2 through 17 years. It was developed primarily as a profile test and, as such, is stated to provide "specific information about children's strengths and weaknesses across a range of cognitive domains" (Elliott, 1990, p. 1).

The normative sample encompassed 3,475 preschoolers, children, and adolescents. Individuals were selected according to a stratified quota system, including 175 preschoolers at each half-year level between the ages of 2 years, 6 months through 4 years, 11 months, and 200 children and adolescents at each year level between the ages of 5 years, 0 months and 17 years, 11 months. There were equal numbers of males and females per level. Quotas for distributions of children's race, education level of parents, geographic region, and educational placements (regular vs. special education) were arranged to approximate distributions identified in the 1986 U.S. Census (U.S. Bureau of the Census, 1986).

The DAS is a hierarchically structured test that spans 17 cognitive ability subtests, which form the base of the hierarchy, and which, theoretically, reflect separate mental abilities. Depending upon the age of the child, some subtests are labeled *core* subtests and can be grouped to form cluster scores that reflect more general variants of ability (e.g., Verbal, Nonverbal Reasoning, & Spatial Ability). Other subtests, labeled *diagnostic* subtests, have higher amounts of subtest specificity that assist in individual interpretations and make differences between subtest scores more meaningful (Elliott, 1990).

The cluster scores have been measured to have mean internal consistency reliabilities ranging from .88 to .92. The internal consistency reliability of the GCA for the school-aged population was .95. Test-retest values for the DAS indicate that the scores are highly stable: the clusters ranged from .83 to .90 and the GCA's test-retest reliability was .93 (Elliott, 1990).

Beginning at age 6 years, 0 months and continuing through the test's upper limit of 17 years, 11 months, children are administered the same nine subtests: 6 core subtests (Pattern Construction, Recall of Designs, Word Definitions, Matrices, Similarities, Sequential and Quantitative Reasoning) and three diagnostic subtests (Recall of Digits, Recall of Objects, Speed of Information Processing). In the present study, each participant's profile was composed of his or her standard scores (M = 50, SD = 10) on the nine cognitive subtests contained in the school-age version of the DAS.

Criteria

Achievement measures. Because psychologists frequently compare ability and achievement scores, the DAS was co-normed with three measures of achievement: Word Reading, Basic Number Skills, and Spelling. Standard scores (M = 100, SD = 15) from these scales were employed as the criteria of academic performance. The mean internal consistency reliabilities for the achievement tests are .87 for Basic Number Skills and .92 for both Word Reading and Spelling (Elliott, 1990).

Classroom behavior measures. The Adjustment Scales for Children and Adolescents (ASCA) (McDermott et al., 1993) is an objective, behavioral assessment device completed by children's teachers. Its total normative sample contained 1,400 5- through 17-year-old children stratified according to 1990 U.S. Census data on the variables of age, gender, academic level, ethnicity, handicapping condition, community size, region of the country, and parents' education levels. Standard scores on the ASCA are expressed as t scores (M =50, SD = 10). Exploratory and confirmatory components analysis of the standardization sample uncovered 6 core scales: Attention-Deficit/Hyperactivity, Solitary Aggressive (Provocative), Solitary Aggressive (Impulsive), Oppositional Defiant, Diffident, and Avoidant. The ASCA also yields two omnibus dimensions: Overreactivity (obtained by adding item scores from the first 4 core scales) and Underreactivity (based on item scores from the last 2 core scales). The 6 core ASCA scores were used as criteria. The 2 omnibus indexes were not used because they are additive components from the 6 core syndromes, and thereby, would be redundant during multivariate statistical analyses.

The core scales have been measured to have internal consistencies ranging from .71 to .86 for the total standardization population. Interrater reliability for these scales ranged from .65 to .85, and test-retest reliability ranged from .66 to .91. The overactivity scale had an internal consistency of .92 for the total standardization population and interrater and test-retest reliabilities of .81 and .75, respectively. The underactivity scale had an internal consistency of .82 for the total standardization population and interrater and test-retest reliabilities of .84 and .79, respectively. Thus, the ASCA's core syndromes and adjustment scales facilitate congruent and stable assessments (McDermott et al., 1993).

Classification status. Project staff involved in the ASCA validation effort also established the proportion of all participants (N = 1,200) who were identified by multidisciplinary child study teams as learning disabled (LD) (2.9%), emotionally disturbed (ED) (1.3%), mentally retarded (MR) (1.0%), speech and language impaired (SLI) (1.8%), and Other (0.4%). The category of Other includes all disability conditions that appeared in the sample less than 1.0% of the time. The proportion of the sample in special education (7.4%) is comparable to levels reported nationally (6.6%) (U.S. Department of Commerce, 1990).

Procedures

Assessments. Prior to being rated on the ASCA, participants were evaluated on the ability and achievement portions of the DAS. The DAS was administered by field coordinators and region supervisors, each of whom was trained with the DAS in national workshops and submitted two accurate practice cases to project staff. Classroom teachers rated children on the ASCA within two months of the DAS. At that time, information was obtained regarding special education services received by children and their respective, specific classifications.

Identifying unusual profiles. Scholars in intelligence testing recommend that the examination of subtest scatter extend beyond the presence or absence of statistically significant differences (Kamphaus, 1993; Kaufman, 1994; Sattler, 1988, 1992). Statistical significance is a necessary, but insufficient, condition for determining abnormality because as many as 40% of an IQ test's standardization sample will show at least one statistically significant subtest deviation (Glutting, Konold, McDermott, Watkins, & Kush, in press; Glutting, McDermott, & Konold, 1997; Kaufman, 1979). The common response to elevated rates of *exceptional* subtest profiles has been to encourage examiners to compare and contrast either nomothetic or ipsatized subtest scores to distributions of univariate prevalence rates.

Univariate-nomothetic analyses customarily begin by subtracting a child's lowest standardized subtest score from his or her highest standardized subtest score. The resulting difference is compared to cumulative percentages reported for the test's standardization sample, and evidence of an unusual (i.e., infrequent) prevalence of the obtained discrepancy is determined. Univariate-ipsative analyses employ an identical procedure; however, children's standardized subtest scores are ipsatized first according to deviations from their own average (i.e., mean) subtest score. Both procedures are univariate because only one difference is derived even though two subtest scores are used.

In reality, all univariate methods are inadequate to analyze groups of subtest scores because profile analysis requires *multiple* dependent comparisons. Measurement specialists have recognized for more than five decades that profiles are integrated sets of test scores that require appropriate hypotheses and statistical treatments (Cattell, 1949; Horst, 1941; Mosel & Roberts, 1954). However, it was not until the last decade that multivariate methods were used to group children according to the level and shape of their ability profiles. Normative taxonomies of the most common, multivariate subtest profiles have been developed for standardization samples from a number of individually administered IQ tests, including the DAS (Holland & McDermott, 1996). The advantages of comparing subtest scores to a multivariate normative taxonomy are twofold: (a) the methodology accounts for the strength and pattern of correlations among subtest scores, and (b) it accounts for both linear and *nonlinear* aspects of profiles because profiles are doubly defined according to level (position toward the upper, central, or lower region of the ability continuum) and shape (the pattern of peaks and valleys across subtest scores).

For an elaboration of mathematical and psychometric benefits of the multivariate approach in the context of the WISC-III and DAS case studies, consult Glutting, McDermott, Watkins, et al. (1997), in a special issue of the *School Psychology Review*. Glutting and his colleagues also present step-by-step directions for the WISC-III and a case study for clinicians who wish to incorporate the multivariate method into their assessments. Holland and McDermott (1996) present a case study for the DAS that uses the multivariate method. As with the first two methods, the multivariate-nomothetic method is prevalence based.

All three methods were employed to identify unusual subtest profiles: univariate nomothetic, ipsative nomothetic, and multivariate nomothetic.

Determination of the percentage of unusual profiles was based on cut scores derived from the DAS standardization sample using a 5% prevalence rate. The 5% prevalence criterion approximates differences greater than 1-1/2 (i.e., 1.65) standard deviations beyond the mean population expectancy and is consistent with the abnormality standard established in other studies (Glutting et al., in press; Glutting et al., 1992; McDermott, Glutting, Jones, Watkins, & Kush, 1989; Watkins & Kush, 1994). Thus, for each method, 60 children were identified with unusual subtest profiles (i.e., 1,200 x .05 = 60). These children were then matched to an equal number of controls without unusual subtest configurations ($n = 60$) on the background characteristics of age, race, gender, parent education levels, and overall IQs from the DAS (i.e., General Cognitive Ability scores).

Results

Once the constituent groups had been formed (i.e., unusual vs. common subtest profiles), the extent to which the groups differed in terms of criterion outcomes was investigated. Three sets of analyses were undertaken. The first addressed differences in classification status. The second focused upon achievement. The third compared classroom behaviors. The analyses are presented according to the method used to identify unusual subtest profiles (multivariate nomothetic, univariate nomothetic, ipsative nomothetic) because group compositions varied with each method.

Multivariate Nomothetic

It was hypothesized that children with unusual subtest profiles would be characterized by a greater incidence of special education classifications, lower achievement levels, and higher (i.e., more aberrant) classroom behaviors. Table 1 compares the criterion performance of groups identified through the multivariate nomothetic method. The categorical nature of the exceptionality status criterion (LD, ED, MR, SLI, Other, Not Classified) required a 2 x 6 chi-square analysis to test for differences in group rates. Using a medium effect size (i.e., "an effect likely to be visible to the naked eye" [Cohen, 1992, p. 156]), a $p < .05$, and a sample size of 120, a priori power was .72 for the chi-square analysis. In other words, there was a 72% probability of finding a significant difference in classification rates if such a difference truly exists in the population. Results from the chi-square analysis indicate that children with unusual subtest profiles were no more likely to experience elevated rates of exceptionality classifications than matched controls ($X^2 = 3.24$, $df[5]$, $p <$ ns).[1]

Both univariate (t test) and multivariate (discriminant function) analyses were conducted on the achievement and behavioral criteria. The univariate analyses were made more sensitive by not correcting for the number of simultaneous contrasts. Nevertheless, all univariate results failed to support hypotheses. No significant differences emerged between groups on the three DAS achievement criteria or the six ASCA behavioral criteria (all ps < ns).

Discriminant function analysis provides a better indication of criterion validity because it uses the best possible combination of variables to differentiate between groups. The multivariate analyses were completed separately for the DAS and ASCA criteria. Using a medium effect size, a two-tailed $p < .05$, and a sample size of 120, a priori power was .92 for the DAS analysis and .95 the ASCA analysis which indicated that there was a 95% probability of finding a significant difference in behavior ratings using the ASCA if such differences exist in the population. As with the univariate comparisons, results from the discriminant function analyses failed to support inferences that children with unusual subtest profiles would display lower achievement levels (Wilk's $\lambda = .994$, F 0.23, $df[3, 114]$, $p <$ ns) or more pathological classroom behaviors (Wilk's $\lambda = .979$, F 0.43, $df[6, 115]$, $p <$ ns).

Univariate Nomothetic and Univariate Ipsative

Criterion comparisons were repeated independently for the two univariate methods (univariate nomothetic, univariate ipsative).[2] Here too, with but one exception, all results failed to support suppositions of criterion validity. The one exception occurred for the univariate-nomothetic method of identifying unusual subtest profiles. However, this difference was the opposite of expectations. A univariate analysis revealed that children with unusual subtest profiles actually obtained more adjusted levels of Solitary Aggressive (Impulsive) behaviors (t score $M = 47.4$ for the unusual profile group vs. 50.2 for the common profile group) ($t = -2.51$, $df[120]$, p

Table 1
Percentages of Exceptionality Status and Scores on Academic and Behavioral Criteria for Groups Identified Through the Multivariate-Nomothethic Method

Criterion variable	Unusual subtest profiles		Common subtest profiles	
	Percentage[a]		Percentage	
Exceptionality status				
Learning disabled	3.3		3.3	
Emotionally disurbed	3.3		3.3	
Mentally retarded	1.7		0.0	
Speech and language impaired	3.3		1.7	
Other impairments	0.0		0.0	
Not classified	88.4		91.7	
	M	*SD*	*M*	*SD*
Academic achievement				
Word reading	102.7[b]	20.1	102.8	15.2
Basic number skills	102.8	19.1	101.8	15.3
Spelling	101.8	18.7	102.6	17.3
Classroom behavior				
Attention-deficit/hyperactivity	49.2[c]	9.9	49.2	9.7
Solitary aggression (provocative)	49.5	9.2	47.7	7.6
Solitary aggression (impulsive)	48.0	5.4	48.1	4.8
Oppositional defiance	49.8	10.0	49.2	9.1
Diffidence	51.0	9.7	49.9	9.5
Avoidance	50.0	10.2	50.4	9.9

Note. Table values are rounded at the first decimal for conventient presentation; n = 60 for unusual profile group; *n* = 60 for common profile group.
[a]The sum of percentages within each column is 100%.
[b]The population standard score $M = 100$ and $SD = 15$.
[c]The population standard score $M = 50$ and $SD = 10$.

< .02). The opposite difference is probably trivial because it evaporated upon a subsequent multivariate, discriminant function analysis (Wilk's $\lambda = .928$, F 1.48, df [6, 115], $p <$ ns).

Discussion

The interpretation of subtest profiles has a long and rich history in the field of children's intelligence testing. The process began nearly a half century ago when Wechsler introduced his first test of childhood intelligence in 1949. Since then, and based mainly on inductive theory and reasoning, interpretations have been offered on more than 100 patterns of subtest variation from Wechsler's tests as well as from other individually administered measures of children's abilities (Bannatyne, 1974; Delaney & Hopkins, 1987; Glasser & Zimmerman, 1967; Guilford, 1967; Kamphaus & Reynolds, 1987; McGrew, 1986;

Selz & Reitan, 1979; Wechsler, 1991; Wechsler & Jaros, 1965). This tradition of interpretive relevance continues to advanced as the standard of good practice in textbooks on children's intelligence testing (Kamphaus, 1993; Kaufman, 1994; Sattler, 1988, 1992), and the process has become more convenient with the advent of commercial microcomputer programs (e.g., Psychological Corporation, 1994).

Notwithstanding appreciable popularity, subtest profiles have not been well-researched and there is accumulating evidence that nearly all of the obtained outcomes are confounded (Glutting et al., in press; McDermott, Fantuzzo, & Glutting, 1990; McDermott et al., 1992). The present investigation avoided two of the more prominent confounds affecting research on children's subtest profiles: the circularity issue and the problem of inverse probabilities. It did so by classifying an unselected cohort according to its obtained DAS subtest configurations (i.e., groups with and without unusual subtest profiles). Thereafter, the two groups were compared and contrasted across a variety of criteria.

Diagnostic Validity of Subtest Profiles

Perhaps the most important criterion was classification status because the classifications themselves carry certain predictions and expectations about what will happen in lieu of effective interventions. Neither statistically significant nor noticeable differences emerged between children with and without unusual subtest profiles in their classifications as LD, ED, MR, SLI, and Other Impaired. These outcomes illustrate a fundamental point: whereas unusual subtest profiles may be of some help in distinguishing between groups with known disorders (e.g., children previously identified as LD and as normal controls), they offer precious little validity for the typical situation that occurs during psychodiagnostic assessments. In other words, results show that when clinicians encounter an unusual DAS profile in their daily practice, they simply will be unable to determine with greater probability than mere chance, whether the child being evaluated is experiencing a learning disability, an emotional disturbance, a particular level of mental retardation, a speech and language impairment, or is free of psychopathology. Therefore, unusual subtest profiles from the DAS have no meaningful criterion validity for differential diagnosis or decision making.

Validity of Subtest Profiles for Hypothesis Testing

The role and function of unusual subtest profiles extend beyond their implied diagnostic precision. Authorities in intelligence testing maintain that irregular subtest variation is crucial to the development of hypotheses about which, among the many, specific cognitions are likely to impact children's achievement and psychosocial adjustment (Kamphaus, 1993; Kaufman, 1994; Sattler, 1988, 1992). Consequently, this study also examined whether unusual subtest profiles from the DAS predict adverse outcomes across several concurrent criteria of academic achievement and behavioral deportment. Results from both univariate and multivariate statistical analyses were uniform in demonstrating that academic and behavior problems were not related to the presence of unusual DAS profiles. Children with unusual subtest profiles, relative to matched controls, showed commensurate performance in the achievement domains of Word Reading, Basic Number Skills, and Spelling. Likewise, children with or without unusual subtest profiles showed equivalent adjustment on the standardized, teacher-rated variables of Attention-Deficit/Hyperactivity, Solitary Aggression (Provocative), Solitary Aggression (Impulsive), Oppositional Defiance, Diffidence, and Avoidance.

Negative results are not unique to the DAS. Glutting et al. (1992) avoided circular methods and inverse probabilities with the K-ABC. Their study was limited to identifying unusual subtest profiles according to the multivariate nomothetic method. Nevertheless, they found that unusual K-ABC subtest configurations were not associated with placement in special education for children previously classified as LD or ED.

Equally important, findings from the current study remained invariant, regardless of the specific method used to identify unusual subtest variation (multivariate nomothetic, univariate nomothetic, univariate ipsative). The current results were consistent across multiple sets of external criteria and across multiple methods used to identify unusual subtest profiles. These findings, in conjunction with those from our earlier study of the K-ABC, combine to raise serious concerns about whether the putative multiple intelligences measured by subtest scores have any validity beyond what can be gained through the analysis of general intelligence or factor deviation quotients.

Importance of Rival Hypotheses

The principle issue behind the current study is that there should be criterion differences between unselected cohorts with and without unusual subtest profiles. Alternatively, it could be argued that the obtained, nonsignificant findings are really nothing more than a weak attempt to prove the null hypothesis of no differences in the general population.

We did not try to prove the null hypothesis; we are not under any obligation to do so. Stated in its most simple terms, the logic of science has become reversed in the case of subtest analysis. We believe three reasons underlie the reversal: (a) subtest analysis is so widespread that practitioners are lulled into thinking it must have validity; (b) the process has such a long, rich history in the field of individual appraisal, and it remains such a mainstay at graduate training programs in school psychology and clinical child psychology, that professionals are inclined to believe subtest analysis works; and (c) psychologists mistakenly assume reasonable evidence already exists for its use and, therefore, definitive evidence must be presented *against* subtest analysis before they will stop (i.e., the null hypothesis must be "proven").

To the contrary, as highlighted throughout this article and elsewhere (Glutting et al., in press; McDermott et al., 1990), the substantial methodological problems surrounding earlier outcome studies of subtest profiles raise grave concerns about their validity. More importantly, the misperception of having to "prove" the null hypothesis becomes apparent when one considers that science actually places the burden of proof on *advocates* of subtest analysis (as well as on advocates of *any* scientific and/or diagnostic procedure). Basic methodological standards stipulate that whenever a plausible, alternative explanation can be offered to account for phenomena, this *rival hypothesis* must be accepted until *it* is disproved (Campbell & Stanley, 1963; Krathwohl, 1993).

In light of this discussion and in direct contradistinction to the reversed assumptions of past inquiry, negative outcomes from the current study should simply be viewed as raising the rival hypothesis that subtest profiles from the DAS had zero or limited criterion validity for estimating children's propensity for placement in special education, concurrent achievement levels, or classroom deportment. This rival hypothesis is all the more compelling because our use of heterogeneous samples and criterion-related methodology controlled two of the most pervasive methodological confounds found in prior studies.

Practical Implications: Applicability of Results to Referred Children

A reviewer of a former version of this article expressed concern about whether the current results would hold for referral samples. After all, referral samples (i.e., the children typically seen by practicing psychologists) are quite different from unselected cohorts. The unselected cohort evaluated in this study actually offers a *superior* method for investigating the diagnostic criterion-related validity of IQ subtest profiles. As noted throughout this article, unselected cohorts avoid the dual problems of circular reasoning and inverse probabilities. Moreover, the current sample was representative of the child population between the ages of 6 through 17 years for the variables of age, gender, race, geographical region, grade level, and mothers' and fathers' education levels. In other contexts such as unselected cohort has been referred to as an epidemiological sample (Rutter, 1989).

Epidemiological studies are expensive and difficult to undertake. More importantly, epidemiological samples offer advantages that studies which examine populations of referred children do not. Referral samples are unrepresentative of the population as a whole and subject to referral bias (Rutter, 1989; Shaywitz, Bennett, Shaywitz, Fletcher, & Escobar, 1990), and they are subject to circularity problems and inverse probabilities (Glutting et al., 1997; McDermott et al., 1990).

Furthermore, referral samples tend to include a higher proportion of children with multiple diagnoses and often contain too many of the most severe cases (Shaywitz, Fletcher, Holahan, & Shaywitz, 1992). Such biases are avoided through the use of unselected cohorts (i.e., epidemiological samples) that provide an encompassing view of the *entire* population of children who meet operational criteria for unusual subtest profiles. In other words, the unselected cohort approach was particularly effective in allowing us to study the full range of children with disabilities, including those previously referrred by their schools as well as those who were not referred. Thus, the results of this study most definitely apply to referred children.

Suggestions for Future Research

The present investigation is not without limitations. Although multiple methods were used to identify unusual subtest profiles and comparisons were completed among several important criteria, it would have been better to use predictive, versus concurrent, criteria. Furthermore, the study did not examine the utility of factor deviation quotients. It clearly would be worthwhile to investigate the diagnostic and/or criterion-related validity of deviation quotients because, unlike subtest scores, their construct validity has been supported by factor analysis. Therefore, more research is needed on the criterion-related validity of factor scores from the DAS, Wechsler's various scales, and intriguing Gf-Gc cluster scores from the Woodcock-Johnson Tests of Cognitive Ability-Revised (Woodcock & Johnson, 1989).

Conclusion

Despite many early claims and continued widespread use, the weight of the evidence has begun to shift against advocates of subtest analysis. It is becoming increasingly clear, and scientifically necessary, that proponents present convincing empirical support for the interpretation of subtest profiles. This evidence must be free of the most common methodological pitfalls and deficiencies that have plagued past inquiry. Otherwise, psychologists will be compelled to relegate subtest analysis to the archives of past, rather than best, practice.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

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

Bannatyne, A. (1974). Diagnosis: A note on recategorization of WISC scaled scores. *Journal of Learning Disabilities, 7*, 272-274.

Bowers, T. G., Risser, M. G., Suchanec, J. F., Tinker, D. E., Ramer, J. C., & Domoto, M. (1992). A developmental index using the Wechsler Intelligence Scale for Children: Implications for the diagnosis and nature of ADHD. *Journal of Learning Disabilities, 25*, 179-185.

Campbell, D. T., & Stanley, J. C. (1963). Experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally.

Cattell, R. B. (1949). r_p and other coefficients of pattern similarity. *Psychometrika, 14*, 279-298.

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

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159).

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.

Delaney, E. A., & Hopkins, T. F. (1987). *Examiner's handbook: An expanded guide for fourth edition users*. Chicago: Riverside.

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

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

Flanagan, D. P., Andrews, T. J., & Genshaft, J. L. (1997). The functional utility of intelligence tests with special education populations. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 457-483). New York: Guilford.

Glasser, A. J., & Zimmerman, I. L. (1967). *Clinical interpretation of the Wechsler Intelligence Scale for Children (WISC)*. New York: Grune & Stratton.

Glutting, J. J., & Bear, G. G. (1989). Comparative efficacy of K-ABC subtests vs. WISC-R subtests in the differential classification of learning disabilities. *Learning Disability Quarterly, 12*, 291-298.

Glutting, J. J., Konold, T. R., McDermott, P. A., Watkins, M. W., & Kush, J. C. (in press). Structure and diagnostic benefits of a normative subtest taxonomy developed from the WISC-III standardization sample. *Journal of School Psychology*.

Glutting, J. J., McDermott, P. A., & Konold, T. R. (1997). Ontology, structure, and diagnostic benefits of the normative subtest taxonomy from the WISC-III standardization sample. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 349-372). New York: Guilford.

Glutting, J. J., McDermott, P. A., Watkins, M. M., Kush, J. J., & 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. *Journal of Special Education, 26*, 85-115.

Gough, H. (1971). Some reflections on the meaning of psychodiagnosis. *American Psychologist, 26*, 160-167.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Heller, K. A., Holtzman, W. H., & Messick, S. (Eds.). (1982). *Placing children in special education: A strategy for equity*. Washington, DC: National Academy Press.

Holland, A. M., & McDermott, P. A. (1996). Discovering core profile types in the school-age standardization sample of the Differential Ability Scales. *Journal of Psychoeducational Assessment, 14*, 131-146.

Horst, P. (1941). The prediction of personal adjustment. *Social Science Research Council Bulletin* (No. 48).

Humphries, T., & Bone, J. (1993). Use of IQ criteria for evaluating the uniqueness of the learning disability profile. *Journal of Learning Disabilities, 26*, 348-351.

Hunter, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257-266). Hillsdale, NJ: Erlbaum.

Jencks, C. (1972). *Inequality: A reassessment of the effect of family and schooling in America.* New York: Basic Books.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Kamin, L. J. (1984). *The science and politics of IQ.* New York: Halsted Press.

Kamphaus, R. W. (1993). *Clinical assessment of children's intelligence.* Boston: Allyn and Bacon.

Kamphaus, R. W., & Reynolds, C. R. (1987). *Clinical and research applications of the K-ABC.* Circle Pines, MN: American Guidance Service.

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.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman assessment Battery for Children: Administration and scoring manual.* Circle Pines, MN: American Guidance Service.

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.

Kazdin, A. E. (1995). *Conduct disorders in childhood and adolescence* (2nd ed.). Thousand Oaks, CA: Sage.

Kercher, A. C., & Sandoval, J. (1991). Reading disability and the Differential Ability Scales. *Journal of School Psychology, 29,* 293-307.

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., & Schnellenberg, R. P. (1987). The viability of scatter analysis on the WISC-R and SBIS: Examining a vestige. *Journal of Psychoeducational Assessment, 5,* 37-47.

Krathwohl, D. R. (1993). *Methods of educational and social science research: An integrated approach.* New York: Longman.

McDermott, P. A. (1981). Sources of error in the psychoeducational diagnosis of children. *Journal of School Psychology, 19,* 31-44.

McDermott, P. A. (1993). National standardization of uniform multisituational measures of child and adolescent behavior pathology. *Psychological Assessment, 5,* 413-424.

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. *Journal of Special Education, 25,* 504-526.

McDermott, P. A., Glutting, J. J., Jones, J. N., Watkins, M. W., & Kush, J. (1989). Core profile types in the WISC-R national sample: Structure, membership, and applications. *Psychological Assessment, 1,* 292-299.

McDermott, P. A., Marston, N. C., & Stott, D. H. (1993). *Adjustment Scales for Children and Adolescents.* Philadelphia: Edumetric and Clinical Science.

McGrew, K. (1986). *Clinical interpretation of the Woodcock-Johnson tests of cognitive ability.* Orlando, FL: Grune & Stratton.

McIntosh, D. E., & Gridley, B. E. (1993). Differential Ability Scales: Profiles of learning disabled subtypes. *Psychology in the Schools, 30,* 11-24.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York: Macmillan.

Moffitt, T. E., Gabrielli, W. F., Mednick, S. A., & Schulsinger, F. (1981). Socioeconomic status, IQ, and delinquency. *Journal of Abnormal Psychology, 90,* 152-156.

Mosel, J. N., & Roberts, J. B. (1954). The comparability of measures of profile similarity: An empirical study. *Journal of Consulting Psychology, 18,* 61-66.

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.

Plante, T. G., & Sykora, C. (1994). Are stress and coping associated with WISC-III performance among children? *Journal of Clinical Psychology, 50,* 759-762.

Prifitera, A., & Dersh, J. (1993). Base rates of WISC-III diagnostic subtest patterns among normal, learning disabled, and ADHD samples. In B. A. Bracken & R. S. McCallum (Eds.), *Journal of Psychoeducational Assessment Monograph Series. Advances in psychoeducational assessment: Wechsler Intelligence Scale for Children-Third Edition* (pp. 43-55). Germantown, TN: Psychoeducational Corporation.

Psychological Corporation. (1994). *WISC-III writer: The interpretive software system.* San Antonio, TX: Author.

Reynolds, C. R., & Kaiser, S. M. (1990). Test bias in psychological assessment. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (2nd ed., pp 487-525). New York: Wiley.

Rourke, B. P., & Strang, J. D. (1984). Subtypes of reading and arithmetical disabilities: A neuropsychological analysis. In M. Rutter (Ed.), *Developmental neuropsychiatry* (pp. 473-488). New York: Guilford.

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

Sattler, J. M. (1992). *Assessment of children: WISC-III and WPPSI-R supplement.* San Diego, CA: Author.

Schwean, V. L., Saklofske, D. H., Yackulic, R. A., & Quinn, D. (1993). WISC-III performance of ADHD children. In B. A. Bracken & R. S. McCallum (Eds.), *Journal of Psychoeducational Assessment Monograph Series. Advances in psychoeducational assessment: Wechsler Intelligence Scale for Children-Third Edition* (pp. 56-70). Germantown, TN: Psychoeducational Corporation.

Selz, M., & Reitan, R. M. (1979). Rules for neuropsychological diagnosis: Classification of brain function in older children. *Journal of Consulting and Clinical Psychology, 47,* 258-264.

Sines, J. O. (1966). Actuarial methods in personality assessment. In B. A. Maher (Ed.), *Progress in experimental personality research* (pp. 133-193). New York: Academic Press.

Stinnett, T. A., Harvey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologist: A national survey. *Journal of Psychoeducational Assessment, 12,* 331-350.

Teeter, P. A., & Smith, P. L. (1993). WISC-III and WJ-R:

Predictive and discriminant validity for students with severe emotional disturbance. In B. A. Bracken & R. S. McCallum (Eds.), *Journal of Psychoeducational Assessment Monograph Series. Advances in psychoeducational assessment: Wechsler Intelligence Scale for Children-Third Edition* (pp. 114-123). Germantown, TN: Psychoeducational Corporation.

U.S. Bureau of the Census. (1986). *Current population survey, October 1986* [machine-readable data file]. Washington, DC: Author.

U.S. Department of Commerce. (1990). *School enrollment-Social and economic characteristics of students* (Current population reports, Series P-20, No. 443). Washington, DC: U.S. Bureau of the Census.

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

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D., & Jaros, B. (1965). Schizophrenic patterns on the WISC. *Journal of Clinical Psychology, 21*, 288-291.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*, 461-481.

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

Wielkiewicz, K. R., & Daood, C. J. (1993). Correlations between WISC-R subtests and scales of the Personality Inventory for Children. *Psychological Reports, 73*, 1343-1346.

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

Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review, 25*, 9-23.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Tests of Cognitive Ability-Revised.* Chicago: Riverside.

Ysseldyke, J. E., & Christenson, S. L. (1988). Linking assessment to intervention. In J. L. Graden, J. E. Zins, & M. J. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for students* (pp. 91-109). Washington, DC: National Association of School Psychologists.

Footnotes

[1]Although the difference in the chi-square analysis was not significant, 11.6% of the children with unusual profiles were in special education versus 8.3% of the children with common profiles (see Table 1); that is, 40% more children from the unusual profile group were in special education compared to the group with common profiles (i.e., 11.6 - 8.3 = 3.3, which in turn, is 40% of 8.3). However, the obtained (i.e., a posteriori) magnitude of effect for the chi-square analysis was estimated using Cohen's (1988) *W* statistic. The obtained *W* of .015 is extremely small according to Cohen (1988, p. 224) who defined any $W < .10$ as "small."

The *W* statistic is not readily interpretable. Nonetheless, Cohen (1988) shows how *W* can be converted to a percentage of between-group variance using Cramer's r_p^2. The obtained r_p^2 (.000059), when multiplied by 100, reveals that there is only a .0059 percentage difference in the proportion of children with and without unusual subtest profiles who received special education. Thus, results from the chi-square analysis not only show *no* statistically significant difference, the obtained 40% difference is illusory. The difference actually represents such a small proportion of the between-group variability (.59 of 1%) that it must be considered either minuscule or as depicting no true difference in classification rates between groups with and without unusual subtest profiles.

[2]McDermott et al. (1992) present two methods for ipsatizing an ability profile. One is the method recommended by Kaufman (1994) and the other is the method recommended by Sattler (1988). Readers are encouraged to consult the Kaufman and Sattler articles for a more complete understanding of the two methods.

Joseph Glutting, PhD, is Professor of Educational Studies with specialization in school psychology, at the University of Delaware. His research includes intellectual assessment, multivariate strategies for typology development, and the interpretation of ability profiles.

Paul A. McDermott is Professor of Policy Research, Evaluation, and Measurement and Chair of the Division of Psychology in Education at the Graduate School of Education of the University of Pennsylvania. He is Director of the PhD Program in School, Community, and Clinical Child Psychology. His research includes assessment of youth psychopathology and learning styles, adult substance abuse, and multivariate strategies for typology development.

Timothy R. Konold is Assistant Professor in the Department of Leadership, Foundations, and Policy, Curry School of Education, University of Virginia. His research interests span applied psychometrics, test development, and classification systems.

Alisa J. Snelbaker is a graduate student in the MS+30 program in School Psychology at the University of Delaware. Her research interest include applied assessment issues, test development, and school consultation models.

Marley W. Watkins, PhD, is Associate Professor of School Psychology at Pennsylvania State University. He is a Diplomate in School Psychology, American Board of Professional Psychology, and his research interests included diagnostic assessment, the development of microcomputer interpretation programs, and computer-assisted instruction.