School Psychology Review
Volume 23, No. 4 1994, pp. 640–651

WECHSLER SUBTEST ANALYSIS: THE RIGHT WAY, THE WRONG WAY, OR NO WAY?

Marley W. Watkins
Arizona State University

Joseph C. Kush
Duquesne University

Abstract: Methods of Wechsler subtest analysis have been challenged on both statistical and theoretical grounds with normative, rather than ipsative, methods recommended. Cluster analysis has previously been applied to the WISC-R standardization sample and resulted in identification of seven core profile types distinguished primarily by levels of global ability. This study compared the WISC-R profiles of 1,222 special education students to those seven core types. More than 96% of the special education cases were found to be probabilistically similar to one of the core types identified in the standardization sample, and were therefore considered to be common variants of normal intellectual abilities. Only 3.6% of the clinical cases were unique in that they could not be categorized into one of the seven core profile types. No regularity was found in their deviations from normality and statistically homogeneous subgroups could not be formed. It was concluded that these unique profiles reflected essentially random and uninterpretable subtest variation. Based upon these data and other literature on subtest profile analysis, it was recommended that "no way" be the response to Wechsler subtest analysis.

The practice of analyzing Wechsler subtest scores has a long, controversial history in school psychology (Kehle, Clark, & Jenson, 1993; Sandoval, 1993; Zachary, 1990). Although Wechsler's original intent was to design a measure of intelligence which also would allow for distinctions between verbal and nonverbal functioning (Zachary, 1990), many clinicians have been unwilling to limit their use of the Wechsler scales to assessing global intellectual abilities and have instead attempted to find unique patterns of subtest scores that could differentially diagnose children as suffering from learning disabilities, emotional disabilities, and mental retardation. This tendency was encouraged by the passage of federal and state legislation which required school psychologists to render psychoeducational diagnoses in determining eligibility for special education services. Initial correlational research comparing subtest scatter to childhood disorders was promising (Dean, 1977; Waugh & Bush, 1977), but this manifestation of subtest analysis eventually failed due to lack of empirical support (Kavale & Forness, 1984; Kramer, Henning-Stout, Ullman, & Schellenberg, 1987; Macmann & Barnett, 1992; McDermott, Fantuzzo, & Glutting, 1990; Mueller, Dennis, & Short, 1986; Sattler, 1988; Zachary, 1990).

In current practice, subtest profile analysis is generally used to make inferences about an individual's cognitive strengths and weaknesses; that is, to generate hypotheses about a person's cognitive abilities which might assist the evaluator in arriving at recommendations, treatments, and programs (Kaufman, 1976; Kaufman & Kaufman, 1993; Sattler, 1988). Kaufman's (1979) influential book directly stated the case for this type of subtest analysis and presented a systematic method for implementing it with the WISC-R. More recently, Kaufman (Kaufman, Harrison, & Ittenbach, 1990) reiterated the importance of a "systematic, logical approach to the understanding of subtest fluctuations" (p. 294)

We gratefully acknowledge the assistance of Joseph Glutting, PhD who generously shared the computer algorithms and supplementary information that made these analyses possible.

Address all correspondence concerning this article to Marley W. Watkins, 511 West Wood Dr., Phoenix, AZ 85029.

and clearly differentiated between the right and wrong ways to use subtest analysis. Briefly, the wrong way was described as one which treats each subtest as an isolated set of skills to be interpreted one or two at a time without placing them in the context of verbal, performance, or full scale IQ scores. In contrast, the right way was delineated as: (a) comparing subtest scores against the mean of the broader cognitive factor of which each is a component, (b) determining statistically if each subtest significantly differs from the broader mean, and (c) interpreting a subtest score as representing a specific strength or weakness in ability only if it differs significantly from the mean of its factor score.

Kaufman's systematic method of subtest analysis has achieved immense popularity in school psychology training and practice. Examples of Kaufman's approach, or similar systems, abound in case reports and textbooks (Kamphaus, 1993; Knoff, 1986; Sattler, 1988, 1992) as well as in computer software applications (Psychological Corporation, 1986; Watkins & Kush, 1988).

Subtest analysis for hypothesis generation has, however, been challenged on both statistical and theoretical grounds. Reschly and Grimes (1990) noted that the differences identified in subtest analyses are likely to be unreliable. Cahan and Cohen (1988) detailed the low statistical power and high proportion of classification errors involved in testing for the statistical significance of subtest score differences and concluded that subtest significance testing should be abandoned. Similarly, difficulties with multiple statistical tests of subtest score differences have been noted by Krauskopf (1991) and Silverstein (1993). Kramer and colleagues (Kramer et al., 1987) considered that weak profile stability and treatment validity, among other factors, warranted against the usefulness of subtest analysis.

Subtest analysis is paradigmatic of *ipsative* measurement. Ipsative scores, in the case of the WISC-R, are typically calculated by subtracting each subtest score from the average score of the battery, scale, or factor. This produces a profile of positive and negative deviations from the average performance commonly interpreted as a pattern of cognitive strengths and weaknesses singularly characteristic of the child being examined. When viewed from an ipsative measurement perspective, subtest analysis is considered to be a measure of intraindividual differences to which a large body of statistical knowledge can be applied (Cattell, 1944; Clemans, 1965; Hicks, 1970). Gorsuch (1974) theorized that ipsatization minimizes a general factor by eliminating the mean differences. After considering ipsative ability measures and numerical theory, Jensen (1992) 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, Fantuzzo, and Glutting (1990), who demonstrated that the ipsatization of WISC-R scores produced a loss of almost 60% of the test's reliable variance. McDermott, Fantuzzo, Glutting, Watkins, and Baggaley (1992) also analyzed an ipsative approach to the WISC-R and found that: (a) the correlation between WISC-R ipsative scores and general intelligence was significantly reduced, (b) short-term and long-term reliability of WISC-R ipsative scores were considerably lower than WISC-R normative scores, and (c) ipsative WISC-R scores exhibited a very weak relationship to academic achievement tests when contrasted to normative WISC-R scores. These authors concluded that comparisons among subtests violate the primary principles of valid test interpretation.

Psychometrically, it has been suggested that multivariate normative, rather than ipsative, procedures are required if individual cognitive differences are to be studied (Sternberg, 1984). McDermott, Glutting, Jones, Watkins, and Kush (1989) presented such a method for the WISC-R, based upon cluster analysis. Cluster analytic techniques also have been referred to as numerical taxonomy, pattern recognition, and Q factor analysis. As described by Ward (1963), these techniques are used when it is desirable to cluster large numbers of persons into smaller numbers of mutually exclusive groups, each optimally homogeneous with respect to the elevation and shape of the clustering variables. Conceptually, cluster analysis might be considered the inverse of factor analysis where similar individuals are grouped together. rather than similar variables.

Using WISC-R subtest scores as clustering variables, McDermott, Glutting, Jones, Watkins, and Kush (1989) identified seven normative profile types based on the WISC-R standardization sample. These core types, as presented in Figure 1, were primarily distinguished by differences in levels of global ability and to a lesser extent by subtest configurations based upon factor analytic dimensions (Kaufman, 1975).

The foregoing discussion has reviewed two major motivations for WISC-R subtest profile analysis: differential diagnosis and generation of hypotheses concerning an examinee's cognitive strengths and weaknesses. The former method has generally been abandoned due to a lack of empirical and referential support. The later approach has achieved widespread popularity in school psychology even though statistical and theoretical analyses suggest that it is fatally flawed and should be replaced with multivariate normative methods.

A comparison of WISC-R clinical profiles to the normative population would determine the rarity or uniqueness of subtest scatter in populations ordinarily seen by school psychologists. The purpose of this study, therefore, was to apply the WISC-R normative typology identified by McDermott, Gluttiing, Jones, Watkins, and Kush (1989) to WISC-R profiles obtained from a large special education population. If the special education clinical profiles are determined to be probabilistically similar to one of the seven core types, then they must be considered to be an undistinctive variant of normal intellectual functioning and not open to interpretation through subtest analysis. On the other hand, clinical WISC-R subtest profiles that deviate significantly from normative population types could be analyzed further to determine what attributes distinguish them from common or normal childhood intellectual variation. The use of subtest analysis to identify cognitive strengths and weaknesses would receive empirical support if stable discriminating attributes can be identified in these nonnormal clinical profiles. Conversely, essentially random and meaningless variation within these nonnormal subtest profiles would fail to substantiate the utility of subtest analysis.

METHOD

Subjects

Students enrolled in a southwestern, suburban school district special education program who received comprehensive psychological evaluations during a 6-year period served as subjects. They were selected from special education records based upon two criteria: (a) cognitive assessment included 11 subtests of the WISC-R and (b) diagnosis of learning disability (LD), emotional handicap (EH), or mental retardation (MR). State special education rules and regulations, which governed diagnostic decisions, were very similar to PL 94-142 rules. That is, (a) learning disability was defined as a significant ability-achievement discrepancy, (b) emotional handicap was conditioned upon one of five emotional characteristics affecting educational progress, and (c) mental retardation required deficits in intellectual functioning and adaptive behavior.

These selection criteria identified 1,222 subjects. Of this total, 90% were enrolled in Grades 1–8. Special education membership was 80% LD, 16% EH, and 4% MR. Gender distribution was 69% male and 31% female. Ethnic membership was 92% White, 1% Black, 6% Hispanic, and 2% other.

Academic achievement levels in reading, math, and written language were obtained for the most recently evaluated 280 students with the Woodcock Johnson-Revised achievement battery. Table 1 presents mean Verbal IQ (VIQ), Performance IQ (PIQ), Full Scale IQ (FSIQ), and achievement standard scores for subjects by special education classification. An estimate of the magnitude of ability-achievement discrepancies was computed by subtracting the lowest academic score (reading, math, or written language) from the highest intellectual score (VIQ, PIQ, or FSIQ). Scrutiny of the mean difference scores in Table 1 indicates that all students achieved at lower than expected levels, with students diagnosed as LD being most discrepant from expectancy.

TABLE 1
Students' Mean Intellectual, Achievement, and Discrepancy Scores by Special Education Category

	Special Education Category		
	LD	EH	MR
WISC-R IQs			
N	974	195	53
VIQ	91.7	92.1	64.9
PIQ	98.8	97.5	64.0
FSIQ	94.6	94.1	62.0
Woodcock-Johnson Revised Achievement Standard Score			
N	231	33	16
Reading	80.1	92.7	72.0
Math	85.3	90.8	57.6
Written Language	75.7	81.8	64.9
IQ-Achievement Discrepancy			
Highest IQ	101.2	100.3	68.5
Lowest Achievement	72.2	79.0	54.8
Discrepancy	29.3	20.6	15.2

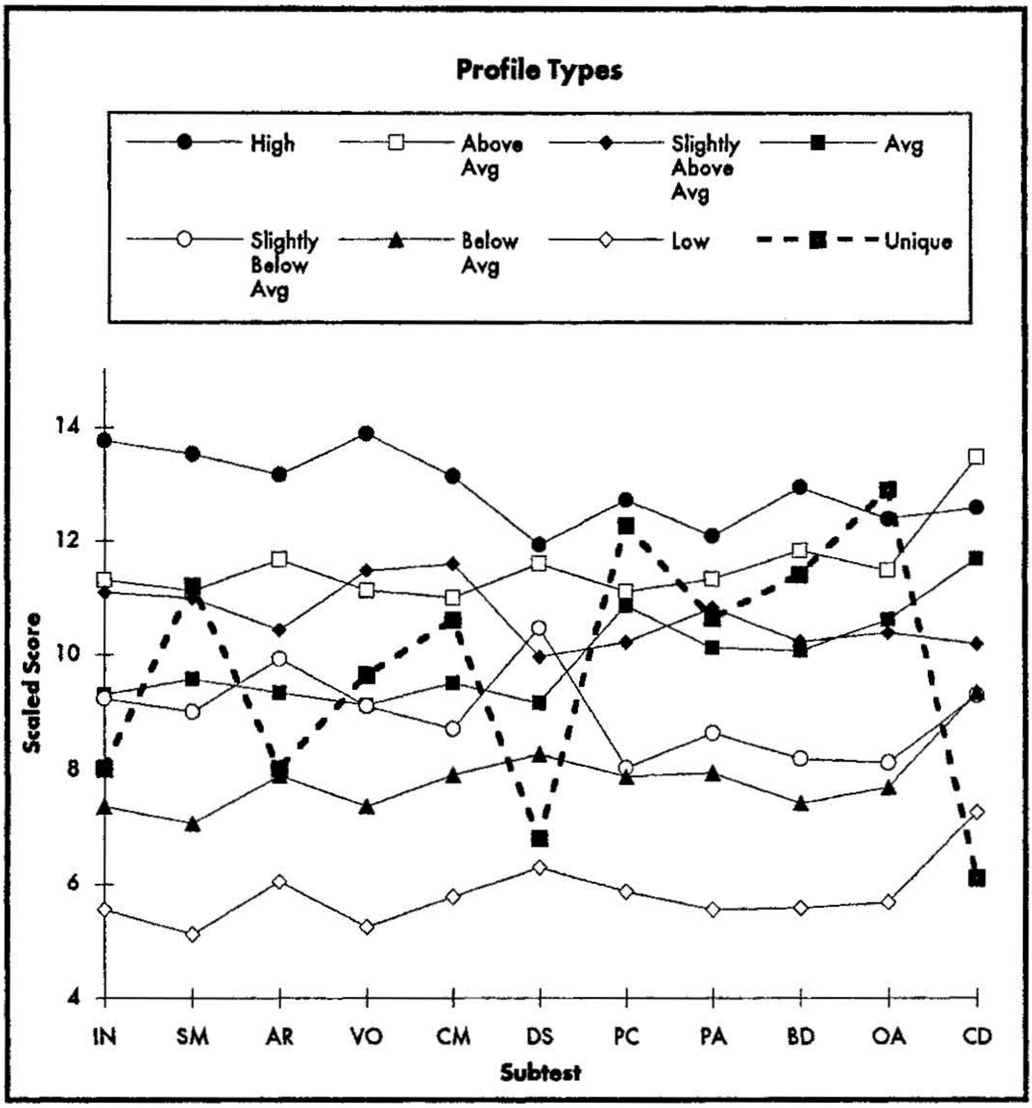
Procedure

The seven core profile types identified by McDermott, Glutting, Jones, Watkins, and Kush (1989) within the WISC-R standardization population served as the normative standard. That is, the empirical typology of core subtest profiles existing within the population of normal children served as the standard to which clinical WISC-R subtest profiles were compared to determine their relative rarity or uniqueness and consequent clinical meaningfulness.

Similarity of subjects' WISC-R profiles to the seven core types found in the population was assessed with the $r_p(k)$ group similarity coefficient for correlated variables (Tatsuoka, 1974; Tatsuoka & Lohnes, 1988). It has been empirically demonstrated that $r_p(k)$ is equal or superior to all other profile similarity measures in classification accuracy (Carroll & Field, 1974). Each $r_p(k)$ value reflects the level and shape similarity of the individual subtest profile to each core type in the population. The values of $r_p(k)$ range from –1.0 to +1.0 and are interpreted much like correlation coefficients where a value of +1.0 indicates identical profile level and shape, a zero value indicates chance similarity, and a negative value indicates dissimilarity.

As noted by Glutting, McGrath, Kamphaus, and McDermott (1992), no absolute criteria exists for the identification of unusual profiles. Selection of a criteria in the present study was guided by several considerations. First, reverse application of the $r_p(k)$ classification procedure to the WISC-R standardization sample suggested an $r_p(k)$ value of .30. This critical value translated into a WISC-R standardization sample prevalence of 1.5%. When applied to the special education sample, an $r_p(k)$ of .30 identified 3.6% of the subjects. Most school psychologists would agree that a prevalence rate of 1.5% to 3.6% is sufficiently infrequent to be considered uncommon. On the other hand, use of a larger $r_p(k)$ critical value led to loss of uniqueness (i.e., an $r_p(k)$ of .40 identified 11% of the standardization sample and 20% of the special education

FIGURE 1. Mean WISC-R subtest patterns for seven core profile types plus the unique special education profile.



Note: IN = Information, SM = Similarities, AR = Arithmetic, VO = Vocabulary, CM = Comprehension, DS = Digit Span, PC = Picture Completion, PA = Picture Arrangement, BD = Block Design, OA = Object Assembly, CD = Coding.

sample). Second, this .30 criterion is congruent with critical values reported in similar research with the K-ABC (Glutting et al., 1992), WISC-R (McDermott, Glutting, Jones, Watkins, & Kush, 1989), WPPSI (Glutting & McDermott, 1990), and WAIS-R (McDermott, Glutting, Jones, & Noonan, 1989). Third, analogous to correlation coefficients, the probability of obtaining a $r_p(k)$ value of .30 by chance is less than one in ten (Cattell, Coulter, & Tsujioka, 1966).

Based upon these considerations, an $r_p(k)$ similarity index of .30 was adopted as the criterion for identifying unique WISC-R profiles. Each WISC-R clinical case was then statistically compared to each of the seven core profiles, which generated seven $r_p(k)$ values for each case. These $r_p(k)$ val-

TABLE 2
WISC-R Subtest Profiles Classified into Seven Core Types and One Unique Group via $r_p(k)$ Across Special Education Categories with Mean VIQ, PIQ, and FSIQ for Each Type

Type	Percent of Total Sample	LD *N*	EH *N*	MR *N*	VIQ Mean	PIQ Mean	FSIQ Mean
High	2.1	23	3	0	117.5	119.4	120.9
Above average	3.3	32	8	0	101.9	118.6	112.0
Slightly above average	15.3	145	42	0	105.7	103.4	105.2
Average	31.0	327	52	0	92.1	105.8	98.1
Slightly below average	5.5	48	9	0	95.5	88.2	91.1
Below average	27.7	293	45	0	83.4	90.8	85.9
Low	11.5	68	20	53	71.4	72.0	69.6
Unique	3.6	38	6	0	96.7	104.3	99.8

ues expressed the similarity of that clinical case to each of the seven normative core profile types. If all seven $r_p(k)$ values for an individual profile were less than .30 then the individual subtest profile was considered unique in that it was probably not a member of (not similar to) a core type in the normal population. This was accomplished with a computer program which represented each core type with its mean subtest profile and subtest intercorrelations with the variance–covariance matrix specific to each core type (J. J. Glutting, personal communication, October 22, 1992).

RESULTS

Figure 1 indicates the present sample's mean subtest patterns for the seven standardization sample WISC-R core profile types (based upon data presented by McDermott, Glutting, Jones, Watkins, and Kush, 1989) as well as the mean subtest profile of 44 atypical special education cases that could not be grouped into one of the seven core profiles.

Overall, the proportion of children classified into the seven core profiles from the special education sample was significantly different from the proportions found in the standardization sample ($x = 585.9$, $df = 6$, $p < .0001$). Typal comparisons of proportions were conducted to determine which profile types were more or less prevalent in the special education population when compared to the total populations. Statistical comparisons of the significance of the difference between two independent proportions (Ferguson, 1966) revealed that the special education population contained a greater proportion of students in the low, below average, and average profile types and a smaller proportion of students in the high and above average profile types.

Classification results from the special education population clearly indicate that 3.6% of the clinical cases were unusual when contrasted to the core types found in the normative populations (see Table 2). Analysis of the demographic structure of the 44 atypical students found that they were predominantly White ($N = 43$), male ($N = 42$), and learning disabled ($N = 38$). The distribution of cases across grade level was relatively even and reflected the distribution of the entire special education population.

Two hypotheses were explored to further interpret these results: (a) these cases are unique and reflect meaningful subtest fluctuation and (b) these cases are unique but reflect essentially random and meaningless subtest variation. To help explicate these two hypotheses, a hierarchical clustering analysis was conducted to detect the presence of meaningful groups embedded within the 44 unique WISC-R subtest profiles. The SPSS implementation of Ward's method of using squared Euclidian distance was employed (SPSS, 1990) because it allows for superior population recovery (Aldenderfer & Blashfield, 1984). A priori stopping criteria consisted of: (a) an initial inspection of the dendogram followed by scrutiny of the total error sums of squares statistic at each step for atypical inflections and (b) an average within-profile homogeneity coefficient (*H*; Tryon & Bailey, 1970) > .60. Conceptually, the first criterion is analogous to the more familiar scree test used to determine the number of factors to rotate in factor analysis (Aldenderfer & Blashfield, 1984). The second criterion, (*H*), is a function of within-cluster variance as a proportion of total variance and reflects the internal coherence or "tightness" of the clusters (Tryon & Bailey, 1970). Neither criterion was achieved. Thus, no recoverable homogeneous clusters existed in this group of unique WISC-R subtest profiles.

The effort to discriminate between meaningful and random subtest variation in these 44 atypical cases was expanded by applying Kaufman's clinical interpretation system (Kaufman, 1979). As noted by Kaufman, Harrison, and Ittenbach (1990), "diagnostic decisions should not be based, even partially, on a significant V–P discrepancy or on the existence of significant strengths and weaknesses in the subtest profile unless the fluctuations occur infrequently in the normal population" (p. 300). Of the 44 unique profiles, 26 exhibited PIQ >VIQ differences (mean difference = 19.1) and 17 displayed VIQ > PIQ differences (mean difference = 10.5). Based upon prevalence data provided by Kaufman, Harrison, and Ittenbach (1990), eight of these cases had VIQ–PIQ discrepancies that could be considered unusual in the general population (less than 5% prevalence). Thus, VIQ–PIQ differences were more common in these atypical special education cases than would be expected (18% versus 5%). Six of these eight uncommon discrepancies were of the PIQ > VIQ pattern. Descriptive statistics for the WISC-R subtests of the 44 unique special education profiles are presented in Table 3. Review of this descriptive information should be guided by results of previous data analyses. Caution must be exercised when making intuitive interpretations of randomness (Bar-Hillel & Wagenaar, 1993), when submitting sampling variability to causal explanation (Tversky & Kahneman, 1993), and applying group statistics to an individual (Dawes, Faust, & Meehl, 1993).

Factor deviation quotients were calculated from formulae developed by Gutkin (1979). Each subtest was compared to the mean of the factor of which it was a member utilizing Bonferroni error correction for multiple comparisons via formulae presented by Sattler (1988) based upon the specific age of each student. All deviations from factor mean scores of $p < .01$ were deemed significant. Approximately 20% of the cases exhibited no significant deviations from factor means. Another 24% had one deviation, 34% had two deviations, 20% had three deviations, and 2% had four deviations. The subtests which deviated most frequently ($n = 9$) were Information, Coding, Picture Completion, and Picture Arrangement while Vocabulary and Comprehension deviated least frequently ($n = 2$). Based upon prevalence data provided by Kaufman, Harrison, and Ittenbach (1990), 8 of these cases had a verbal scaled score range and 16 had a performance scaled score range that could be considered unusual in the general population (less than 5% prevalence). When using Sattler's (1988, p. 166) terminology, 40 of the 44 students in the current study showed moderate subtest variability and only 4 students showed extreme subtest variability. Thus, subtest score scatter was more common in these atypical special education cases than would be expected in a normal population (18–36% versus 5%).

DISCUSSION

McDermott, Glutting, Jones, Watkins, and Kush (1989) identified seven core profile types in the WISC-R normative population primarily distinguished by differences in levels of global ability. The present investigation applied multivariate normative methods to the study of individual WISC-R differences by comparing the WISC-R subtest profiles of 1,222 special education cases to the normative population to determine the rarity or ordinariness of those clinical profiles. It was found that 96.4% of the

TABLE 3
Standard Score Means, Standard Deviations, and Ranges for WISC-R Subtests, VIQ, PIQ, FSIQ, VC, PO, and FD for Unique Special Education Students

Variable	Mean	SD	Minimum	Maximum
Information	8.0	2.3	3	12
Similarities	11.2	2.9	2	18
Arithmetic	8.0	2.4	3	13
Vocabulary	9.6	2.4	2	14
Comprehension	10.6	2.2	6	15
Digit Span	6.8	2.7	2	17
Picture Completion	12.3	2.7	7	18
Picture Arrangement	10.6	4.0	1	18
Block Design	11.4	3.8	1	18
Object Assembly	12.9	3.8	4	19
Coding	6.1	2.7	1	14
Verbal IQ	96.7	8.4	70	114
Performance IQ	104.3	15.0	74	131
Full Scale IQ	99.8	8.6	89	123
Verbal Comprehension	99.2	10.2	72	122
Perceptual Organization	111.6	15.8	76	135
Freedom from Distractibility	80.1	10.6	63	113

special education population displayed subtest profiles that were probabilistically similar to the WISC-R standardization sample. That is, more than 96% of the special education clinical profiles were statistically similar to one of the seven core WISC-R types and must be considered common and undistinctive variants of normal intellectual abilities not open to interpretation via subtest profile analysis.

Only approximately 4% of the special education WISC-R subtest profiles could not be classified as members of one of the seven core population types and, therefore, were deemed unique and eligible for further analysis to identify attributes that could reliably distinguish them from common subtest profiles. These 44 subtest profiles, not parsimoniously similar to core population types, were submitted to a subsequent clustering algorithm to detect any meaningful grouping, but no homogeneous clusters were identified. Descriptively, the nonclassified sample, as a group, exhibited more VIQ-PIQ differences and more subtest scatter than is generally found in the normal population. There were no other discernible clinical patterns and it is unclear what the obtained VIQ-PIQ differences and subtest scatter can mean beyond reflecting the factorial structure of the WISC-R. Although these unique special education profiles deviate from normality in some ways, it appears that they depict essentially random subtest variation. Thus, the present research has failed to empirically support the utility of Wechsler subtest profile analysis.

The practical implications of these results for school psychology practice are clear. WISC-R profiles that reflect clinically unique scatter are quite rare in both normal and special education groups. Our analysis of the WISC-R profiles of more than 1,200 special education students indicated that these children were best grouped according to their general ability level and, to a lesser degree, by the verbal, perceptual organization, and freedom from distractibility factors. Based on these data, the practicing school psychologist can expect to encounter an unusual WISC-R profile in only approximately 4% of the special education population, and even less frequently (1.5%) in the normal population. Given their own referral rates, school psychologists can estimate the frequency with which they should expect to encounter nonnormal profiles: 25 evaluations in a year might produce one unique profile; 50 psychological evaluations in a year would result in one or two unique profiles; and a busy year of 100 psychological evaluations would yield only two to four unique profiles.

Beyond this rarity, there is no way for the school psychologist to reliably identify those few unique Wechsler profiles without applying the computationality complex $r_p(k)$ group similarity coefficient for correlated variables (Tatsuoka, 1974; Tatsuoka & Lohnes, 1988) to each case. Less onerous computational methods based upon generalized distance theory, as detailed by McDermott and colleagues (1989), are available but considerably less valid since they identified two false positives for each true positive when applied to all 1,222 special education cases.

Even after solving the problems of rarity and unreliable identification, the school psychologist has no valid way to interpret the variability of these cases. No regularity was found in these unique special education profiles' deviations from normality and statistically homogeneous subgroups could not be formed. Thus, it appears that they reflect essentially random and uninterpretable subtest variation.

Our findings strongly suggest that practitioners more closely attend to the cautions expressed in the literature surrounding the practice of subtest profile analysis (Cahan & Cohen, 1988; Glutting & McDermott, 1990; Glutting et al., 1992; Jensen, 1992; Kavale & Forness, 1984; Kramer et al., 1987; Krauskopf, 1991; Macmann & Barnett, 1992; Matarazzo, 1990; McDermott, Fantuzzo, & Glutting, 1990; McDermott et al., 1992; McDermott, Glutting, Jones, & Noonan, 1989; McDermott et al., 1989; Reschly & Grimes, 1990; Silverstein, 1993). The present study, when considered within the context of this extensive body of critical commentaries, can only question the probity of Wechsler subtest analysis. The interpretation of WISC-R subtest scatter, either for differential diagnosis or for hypothesis generation, has not been empirically supported. It is necessary, therefore, to conclude that discussions of the "right" and "wrong" ways to conduct subtest analysis be abandoned in favor of the conclusion that "no way" exists to reliably identify unique Wechsler subtest profiles which reflect anything other than essentially random, meaningless, and uninterpretable subtest variation.

Several limitations of this research should be understood when examining these results and conclusions. First, outcomes from our WISC-R analysis cannot be uncritically applied to the third edition of the Wechsler Intelligence Scale for Children. Although the WISC-R and WISC-III are quite similar (Little, 1992), WISC-R results should not be extended to the WISC-III without empirical verification. It seems unlikely, however, that data will be obtained from the WISC-III which contradict results from the K-ABC (Glutting et al., 1992), WISC-R (McDermott, Glutting, Jones, Watkins, & Kush, 1989), WPPSI (Glutting & McDermott, 1990), and WAIS-R (McDermott, Glutting, Jones, & Noonan, 1989).

Second, the $r_p(k)$ criterion applied in this research was directly related to the proportion of unique cases identified and, consequently, to the conclusions generated by that case data. The $r_p(k)$ level selected for this study was consonant with previous research in cognitive functioning (Glutting et al., 1992; Glutting & McDermott, 1990; McDermott et al., 1989; McDermott, Glutting, Jones, & Noonan, 1989) and should be a reasonably accurate standard. Nevertheless, the subjectivity involved in selecting an $r_p(k)$ criterion level might constitute a threat to the generalizability of these results.

Third, the special education population investigated in this article might be somehow unrepresentative of other special education students or clinical cases. The large sample size ($N = 1{,}222$) mitigates against this risk, but subtle selection biases also might represent a threat to generalization of our results.

Fourth, the negative results obtained in this research on subtest profile analysis should not be extrapolated to the use of intelligence tests in general. Despite the social and psychometric limitations of intellectual testing, the concept of global intelligence continues to offer much in terms of descriptive and predictive clinical utility. In addition, the verbal and performance factors of the WISC-R may provide important information concerning individual student strengths and weaknesses. It is becoming increasingly clear that cognitive ability tests predict job performance in a wide variety of occupations (Barrett & Depinet, 1991) and considerable evidence exists that IQ tests are a good guide to future academic achievement (Weinberg, 1989). Thus, use of the WISC-R as a measure of vocational or scholastic ability has not been addressed by our data.

Given the identified limitations of our research and the significance of this topic, it is important that further empirical investigations of Wechsler subtest profile analysis be conducted. One strongly recommended approach is replication of this research, substituting the WISC-III for the WISC-R within diverse special education and normal populations while applying complementary statistical grouping and classification methods.

REFERENCES

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis.* Beverly Hills, CA: Sage.

Bar-Hillel, M., & Wagenaar, W. A. (1993). The perception of randomness. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 369–393). Hillsdale, NJ: Erlbaum.

Barrett, G. V., & Depinet, R. L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist, 46,* 1012–1024.

Cahan, S., & Cohen, N. (1988). Significance testing of subtest score differences: The case of nonsignificant results. *Journal of Psychoeducational Assessment, 6,* 107–117.

Carroll, R. M., & Field, J. (1974). A comparison of the classification accuracy of profile similarity measures. *Multivariate Behavioral Research, 9,* 373–380.

Cattell, R. B. (1944). Psychological measurement: Normative, ipsative, interactive. *Psychological Review, 51,* 292–303.

Cattell, R. B., Coulter, M. A., & Tsujioka, B. (1966). In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 288–329). Chicago: Rand McNally & Company.

Clemans, W. V. (1965). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs* (No. 14).

Dawes, R. M., Faust, D., & Meehl, P. E. (1993). Statistical prediction versus clinical prediction: Improving what works. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 351–367). Hillsdale, NJ: Erlbaum.

Dean, R. S. (1977). Patterns of emotional disturbance on the WISC-R. *Journal of Clinical Psychology, 33,* 486–490.

Ferguson, G. A. (1966). *Statistical analysis in psychology and education.* New York: McGraw-Hill.

Glutting, J. J., & McDermott, P. A. (1990). Patterns and prevalence of core profile types in the WPPSI standardization sample. *School Psychology Review, 19,* 471–491.

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.

Gorsuch, R. L. (1974). *Factor analysis.* Philadelphia: Saunders.

Gutkin, T. B. (1979). The WISC-R verbal comprehension, perceptual organization, and freedom from distractibility deviation quotients: Data for practitioners. *Psychology in the Schools, 16,* 359–360.

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced normative measures. *Psychological Bulletin, 74,* 167–184.

Jensen, A. R. (1992). Commentary: Vehicles of g. *Psychological Science, 3,* 275–278.

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

Kaufman, A. S. (1975). Factor analysis of the WISC-R at 11 age levels between $6^1/_2$ and $16^1/_2$ years. *Journal of Consulting and Clinical Psychology, 43,* 135–147.

Kaufman, A. S. (1976). A new approach to the interpretation of test scatter on the WISC-R. *Journal of Learning Disabilities, 9,* 160–168.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R.* New York: Wiley.

Kaufman, A. S., Harrison, P. L., & Ittenbach, R. F. (1990). Intelligence testing in the schools. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (pp. 289–327). New York: Wiley.

Kaufman, A. S., & Kaufman, N. L. (1993). *Manual: Kaufman Adolescent & Adult Intelligence Test.* 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.

Kehle, T. J., Clark, E., & Jenson, W. R. (1993). The development of testing as applied to school psychology. *Journal of School Psychology, 31,* 143–161.

Knoff, H. M. (1986). Personality assessment report. In H. M. Knoff (Ed.), *The assessment of child and adolescent personality* (pp. 570–596). New York: Guilford.

Kramer, J. J., Henning-Stout, M., Ullman, D. P., & Schellenberg, 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.

Krauskopf, C. J. (1991). Pattern analysis and statistical power. *Psychological Assessment: A Journal of Clinical and Consulting Psychology, 3,* 261–264.

Macmann, G. M., & Barnett, D. W. (1992). Redefining the WISC-R: Implications for professional practice and public policy. *Journal of Special Education, 26,* 139-161.

Little, S. G. (1992). The WISC-III: Everything old is new again. *School Psychology Quarterly, 7,* 136–142.

Matarazzo, J. D. (1990). Psychological assessment versus psychological testing: Validation from Binet to the school, clinic, and courtroom. *American Psychologist, 45,* 999–1017.

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

McDermott, P. A., Glutting, J. J., Jones, J. N., & Noonan, J. V. (1989). Typology and prevailing composition of core profiles in the WAIS-R standardization sample. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 118–125.

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: A Journal of Consulting and Clinical Psychology, 1,* 292–299.

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.

Psychological Corporation. (1986). *WISC-R microcomputer-assisted interpretive report* [Computer program]. New York: Author.

Reschly, D. J., & Grimes, J. P. (1990). Best practices in intellectual assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology–II* (pp. 425–440). Washington, DC: National Association of School Psychologists.

Sandoval, J. (1993). The history of interventions in school psychology. *Journal of School Psychology, 31,* 195–217.

Sattler, J. M. (1988). *Assessment of children* (Third Edition). San Diego: Author.

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

Silverstein, A. B. (1993). Type I, type II, and other types of errors in pattern analysis. *Psychological Assessment, 5,* 72–74.

SPSS. (1990). *SPSS base system user's guide.* Chicago: Author.

Sternberg, R. J. (1984). The Kaufman Assessment Battery for Children: A information-processing analysis and critique. *Journal of Special Education, 18,* 269–278.

Tatsuoka, M. M. (1974). *Classification procedures: Profile similarity.* Champaign, IL: Institute for Personality and Ability Testing.

Tatsuoka, M. M., & Lohnes, P. R. (1988). *Multivariate analysis techniques for educational and psychological research* (Second Edition). New York: Macmillan.

Tryon, R. C., & Bailey, D. E. (1970). *Cluster analysis.* New York: McGraw-Hill.

Tversky, A., & Kaheman, D. (1993). Belief in the law of small numbers. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 341–349). Hillsdale, NJ: Erlbaum.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *American Statistical Association Journal, 58,* 236–244.

Watkins, M. W., & Kush, J. C. (1988). *Rational WISC-R analysis* [Computer program].Pphoenix, AZ: Southwest EdPsych Services.

Waugh, W. J., & Bush, K. W. (1977). *Diagnosing learning disorders.* Columbus, OH: Charles E. Merrill.

Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debates. *American Psychologist, 44,* 98–104.

Zachary, R. A. (1990). Wechsler's intelligence scales: Theoretical and practical considerations. *Journal of Psychoeducational Assessment, 8,* 276-289.

Marley W. Watkins, PhD, received his doctorate in School Psychology from the University of Nebraska-Lincoln. He is a Diplomate in School Psychology, American Board of Professional Psychology, and is now with the Deer Valley U.S.D., Phoenix, AZ.

Joseph C. Kush, PhD, received his doctorate in School Psychology from Arizona State University. He is currently Assistant Professor of School Psychology at Duquesne University. His research interests include cognitive and intellectual theory and assessment.