Pergamon

Journal of School Psychology, Vol. 37, No. 1, pp. 29–48, 1999

Printed in the USA
0022-4405/99 $–see front matter

PII S0022-4405(98)00024-7

Structure and Diagnostic Benefits of a Normative Subtest Taxonomy Developed from the WISC-III Standardization Sample

Timothy R. Konold
University of Virginia

Joseph J. Glutting
University of Delaware

Paul A. McDermott
University of Pennsylvania

Joseph C. Kush
Duquesne University

Marley M. Watkins
Pennsylvania State University

The structure and composition of profile types most representative of the 2,200 children (6 years 0 months to 16 years 11 months) comprising the normative sample for the Wechsler Intelligence Scale for Children—Third Edition (WISC-III) are identified. Profiles from the 10 mandatory WISC-III subtests are sorted according to similar shape and level using multistage cluster analysis with independent replications. The final solution of eight most common (or core) profile types fulfills all formal heuristic and statistical criteria, including complete coverage, satisfactory within-type homogeneity, between-type dissimilarity, and replicability. Profile types are described according to population prevalence, ability level, subtest configuration; and each type is examined for membership trends by child demography, family characteristics, and unusual IQ discrepancies. Two methods are given for determining the relative uniqueness of WISC-III profile patterns in future research and clinical work. The article concludes with a case example using the method recommended for "everyday" decision making. © 1999 Society for the Study of School Psychology. Published by Elsevier Science Ltd

Keywords: Classification, Assessment, Subtest analysis.

Received September 11, 1995; accepted October 8, 1997.
Address correspondence and reprint requests to Timothy R. Konold, Ph.D., Department of Leadership, Foundations, and Policy, University of Virginia, 405 Emmet Street South, Charlottesville, VA 22903-2495. E-mail: Konold@virginia.edu

It is common practice to interpret the elevation and depression of subtest scores on tests of children's intelligence. The practice seems to stem from the fact that better intelligence tests, such as the Wechsler Intelligence Scale for Children—Third Edition (WISC-III) (Wechsler, 1991), require substantial administration times, and clinicians feel compelled to draw as much information as possible from the assessment process. Moreover, the subtests themselves retain some reliable and distinct variability that may help to support individual interpretations (Bracken, McCallum, & Crain, 1993; Kaufman, 1994).

Strategies for analyzing subtest scores can be divided into two major categories, although the two are not always distinct in their application. The first consists of examining statistical significance levels between one or more sets of subtest scores. The second documents variations in score base rates.

Psychologists are well versed in evaluating the *statistical significance* of WISC-III score differences (i.e., using p values such as $<.05$, $<.01$, and so forth). Characteristically, a child's performance on one or more subtests is compared to either the group average (i.e., a normative approach) or to the child's personal mean (an ipsative approach).

Establishing the statistical significance of a score difference is important because it greatly enhances the probability that the difference is not merely a chance occurrence (Silverstein, 1981). However, statistically significant differences can be quite common and ordinary. They simply reflect the distinct but natural variation of test scores and are not necessarily a reason for concern (Cahan, 1986; Silverstein, 1993).

Moreover, as a preliminary matter to the main study in this article, we examined the number of children from the WISC-III standardization sample (N = 2,200) who showed at least one statistically significant subtest deviation. Scores from the 10 Mandatory subtests were compared one at a time to children's personal means (optional WISC-III subtests were excluded). Statistically significant deviations were determined by tabled $p < .05$ critical values presented in the WISC-III manuals Table B.3 (p. 264). We restricted our analysis to the identification of subtest weaknesses (i.e., children showing subtest scores significantly below their own mean). We did not investigate the number of strengths. Our analysis showed that 42.7% of the children had at least one statistically significant weakness. Thus, when statistical significance is employed as a guideline, psychologists are willing to identify some sort of learning problem on the WISC-III, or generate an interpretive hypothesis, for over 40% of the children in the United States.

Partly as a consequence of the limitations of statistical significance testing, textbooks on intelligence testing have recently begun to encourage an examination of *univariate base rates* (Kamphaus, 1993; Kaufman, 1994; Sattler, 1992). The comparisons customarily start by subtracting a child's lowest WISC-III subtest score from their highest. The difference is compared to cumulative percentages reported for the test's standardization sample and a decision is made as to whether the obtained discrepancy is unusual. The procedure is univariate because only one difference is derived, even though two subtest scores are used.

Despite their benefits, both statistical significance testing and the univariate base-rate approach suffer from three drawbacks. First, the methods rarely account for correlations among subtest scores. As a result, some sets of subtests are prone to showing larger (or smaller) differences as a consequence of the magnitude of their interrelationships. Second, the methods are univariate. Only one subtest score at a time, or one composite formed from several sets of subtest scores, or one difference score is compared to the appropriate distributional statistics (i.e, standardization sample mean and standard deviation). The comparisons are then repeated as necessary. Third, *profiles* are quite unlike individual subtest scores or linear composites formed from groups of subtest scores.

Regarding correlations, Salvia and Good (1982) present a univariate method of comparison that accounts for correlations among subtest scores. Nonetheless, the more popular procedure is to use Davis's (1959) formula for examining statistical significance levels. The formula does not control intercorrelations and Davis's method is incorporated in such widely-used comparison tables as found in the WISC-III manual [(1991), p. 264], Sattler's [(1988), p. 815] textbook on intelligence testing, as well as, other sources.

In actuality, univariate methods are inadequate to analyze groups of subtests because profile analysis requires multiple dependent comparisons. Profiles are integrated sets of test scores and so require appropriate hypotheses and statistical treatments (Cattell, 1949; Horst, 1941; Mosel & Roberts, 1954). Two classes of multivariate methods can be used to examine profiles. Cattell (1949) referred to the procedures as either *R* or *Q* analysis. Both account for correlations among subtest scores. Moreover, the procedures are capable of completing multiple comparisons *simultaneously*—the typical situation that occurs during psychodiagnostic appraisals. Thus, multivariate methods better honor multidifferentiated views of intelligence, as well as, the full network of relationships that exist among such abilities (Sternberg, 1984).

R analysis is based on the linear variation of test scores. Examples include factor, multiple regression, discriminant, and canonical analysis. However, by their nature, subtest 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). *R* analysis is based on linear modeling and is insensitive to differences in *both* profile level and shape. *Q* analysis respects both types of variation and is better able to address nonlinear, configural hypotheses (Cattell, Coulter, & Tsujioka, 1966; Tatsuoka, 1974).

A more appropriate approach to profile analysis would begin by applying *Q* methodology to group children simultaneously according to the level and shape of their subtest scores. A *normative* taxonomy of the most common subtest profiles would result when *Q* analysis is conducted with a test's standardization sample. The normative taxonomy would offer important diagnostic benefits because it could be used as a viable contrast or null condition for testing the uniqueness of profiles believed to be unusual or clinically important. Uniqueness would become plausible only when it could be demonstrated that a child's pattern of subtest scores is atypical of the most common (or core) patterns found in the population.

We undertook a series of studies to establish an empirical taxonomy of core subtest profiles existing within the standardization sample of the WISC-III. We also recognized it would be important to develop a taxonomy that would have the widest application during clinical assessments. This paper summarizes the culminating steps in that effort and relates the final taxonomy to known population prevalence for children's demographic and personal characteristics.

METHOD

Participants

The overall taxonomy was based on the entire sample of 2,200 children and adolescents used in the WISC-III's standardization study (Wechsler, 1991). Participants were selected according to a stratified quota system including 200 children at each of 11 age levels from 6 (6 years; 0 months) through 16 (16 years; 11 months), with 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 education vs. special education) were arranged to approximate distributions identified in the 1988 U.S. Census.

Profile Components

The WISC-III comprises 13 subtests. Ten are mandatory and contribute to two scale IQs: the Verbal Scale IQ (VIQ) which is composed of 5 subtests and the Performance Scale IQ (PIQ) which encompasses another 5 subtests. The VIQ and PIQ are combined to form the Full Scale IQ (FSIQ).

Each child's profile was based on scaled scores (Mean = 10, SD = 3) from the 10 mandatory subtests [(Wechsler, 1991), p. 5]. The supplementary subtests of Digit Span, Symbol Search, and Mazes were excluded because they do not influence formation of the FSIQ, VIQ, and PIQ. Moreover, as preparation for this study, we examined 6,424 WISC-III protocols obtained from psychodiagnostic evaluations conducted in Arizona, Delaware, New Jersey, Pennsylvania, Texas, and Virginia. All protocols had scores from the 10 mandatory subtests; less than 1% had Mazes and 31.9%

included Digit Span and Symbol Search. Thus, whereas a taxonomy encompassing supplementary subtests would be appropriate for some psychologists, a 10 subtest taxonomy would be useful for everyone.

Criterion Variables

Internal criteria. Children's obtained deviation IQs (Mean = 100, SD = 15) for the FSIQ, VIQ, and PIQ were used to help describe and interpret the final taxonomy. Also, prevalence of unusual VIQ/PIQ discrepancies within profile types was used to support interpretations regarding unusual profile configurations. Unusual discrepancies were defined in the clinical sense as those that occur in no more than 3% of the general population and are consistent with classifications used during development of a core profile taxonomy for the *Wechsler Intelligence Scale for Children-Revised* (Wechsler, 1974; McDermott, Glutting, Jones, Watkins, & Kush, 1989).

External criteria. Unlike deviation IQs that are actually transformed linear composites of the subtests themselves, certain variables were used to describe and lend validity to the taxonomy. These included the WISC-III stratification variables of children's age, sex, race, educational placements, region of the country, and the occupational status of the head of the household.

Procedure

Our primary goal was to identify and describe the most common profile types existing in the normal child population. This meant sorting the 2,200 profiles according to level and shape so that those within each group were maximally similar to one another (maximum homogeneity) and dissimilar to those in other groups (minimum overlap). However, the groups of similar profiles (called *profile types*) must be reasonably replicable across age levels rather than spurious mergers, as would occur by chance. Considering the overall solution (or *taxonomy*), it should account for all profile variation in the population (known as full coverage) and not discount profiles that happen to diverge from the popular trend. This is particularly important for a taxonomy intended to be fully representative of the general population of children.

Cluster analysis (*Q* methodology) was used to sort the 2,200 profiles. After evaluating numerous clustering algorithms, Ward's (1963) minimum-variance procedure was determined to best satisfy our research goals. Monte Carlo studies of competing clustering methods have consistently shown that when full coverage is required, Ward's method better recovers known taxonomic structure (Kuiper & Fisher, 1975; Mojena, 1977) and it outperforms other methods in reducing overlap (Bayne, Beauchamp, Begovich, & Kane, 1980). Ward's method also is the most accurate under mixture model testing where individuals must be classified to diverse known populations (Blashfield, 1976; Overall, Gibson, & Novy, 1993). In contrast, average-linkage clustering (the best alternative to Ward's approach) does comparatively poorly in reducing overlap (Bayne et al., 1980; Milligan, 1980), and in previous investigations of ability-profile taxonomies, average-linkage clustering performed poorly in comparison to Ward's method (Glutting, McGrath, Kamphaus, & McDermott, 1992; McDermott, Glutting, Jones, & Noonan, 1989).

Our clustering strategy comprised three stages and began with Ward's (1963) agglomerative algorithm. Specifically, the aggregate sample of 2,200 children was partitioned by age levels to form 11 blocks of 200 children, and profiles for children comprising each block were clustered independently. Clusters derived from the 11 independent analyses were pooled to form a proximity matrix of first-stage clusters that were themselves subject to second-stage clustering by Ward's method.

Second-stage clustering began with a proximity matrix whose diagonal elements held error sums of squares (ESS) statistic values for respective first-stage clusters, with off-diagonal elements corresponding to potential ESSs for merging each pair of first-stage clusters. Group centroids from the second-stage solution served as starting partitions for the third-stage, iterative-partitioning analysis conducted using K-means passes.

The choice of similarity measure was ESS statistic for the first- and second-stage agglomerative analyses and Euclidean distance for the third-stage, iterative-partitioning analysis. Correlation coefficients were rejected as similarity measures inasmuch as their sensitivity is constrained to differences in profile shape (Cattell et al., 1966; Cronbach & Gleser, 1953; Sneath & Sokal, 1973). Alternatively, the ESS statistic and Euclidean distance simultaneously account for differences in profile elevation and shape (Aldenderfer & Blashfield, 1984).

Several stopping rules were employed during the first- and second-stage analyses. Appropriate agglomerative solutions were required to: (a) correspond to a hierarchical step preceding an atypical inflection in the similarity measure; (b) fulfill Mojena's (1977) first stopping rule; and (c) satisfy Wishart's (1982) t-test. Stopping occurred during the third-stage, iterative-partitioning analysis when subject relocations ceased to improve within-cluster homogeneities.

Results from the final solution were compared along three internal criteria. That is, an ideal solution was required to: (a) show a replicability rate $\geq 60\%$ for the absorption of first-stage clusters into final taxonomy; (b) yield an average within-profile type homogeneity coefficient, $\overline{H}$ (Tryon & Bailey, 1970) $>.60$; and (c) provide an average between-profile-types similarity coefficient, $\overline{r_p}$ (Cattell, 1949), $<.40$. H and r_p are each sensitive to similar profile levels and shapes, where a value of 1.0 indicates profiles' identical in level and shape, 0.0 indicates chance similarity based on the full

WISC-III sample, and negative values indicate gross dissimilarity. The respective .60 and .40 a priori criteria were established through clustering and classification studies with normative samples from the WISC-R (McDermott, Glutting, Jones, Watkins, et al., 1989) and Wechsler Adult Intelligence Scale-Revised (WAIS-R) (Wechsler, 1981) (McDermott, Glutting, Jones, & Noonan, 1989).

The various internal and external criterion measures were used to describe or lend validity to each final profile type. Thus, considering the prevalence distribution of each pertinent demographic variable within a profile type, we conducted two-tailed tests of the standard error of proportional differences for all possible pairwise comparisons across levels of the criterion variables (Ferguson & Takane, 1988). Type I error rates were apportioned by the Bonferroni correction (Stevens, 1986). By this approach, the expected prevalence for a given characteristic within a profile type (e.g., Anglos versus Blacks) was based on its prevalence in the U.S. population, and unusual prevalence for a particular profile type was determined by statistically significant deviations from the general expectancy.

To achieve better balance in distributions of age intervals and in power for subsequent statistical tests, the original 11 age groups were reduced to 3; i.e., 6–8, 9–12, and 13–16 years. (This arrangement pertained exclusively to post hoc testing of age as an external criterion variable. As described above, development and replication of the taxonomy itself considered variation separately within each of the 11 original age intervals.)

RESULTS

Typal Structure

First-stage clustering produced 110 profile groups (an average of 10.0 per analysis). These were submitted to second-stage agglomerative clustering based on an 110 × 110 similarity matrix and the solution at all hierarchical steps was evaluated against the stated internal criteria. The second-stage, 8-cluster solution was the only one to satisfy all criteria. Therefore, it was submitted to a third-stage, iterative-partitioning analysis. Subject relocations during the third-stage analysis ceased after 34 iterations.

Table 1 displays, for each of the final 8 core profile types, its average coefficient for within-type homogeneity, between-types similarity, and replication rate. The average $\overline{H}$ value (.67) satisfies the a priori criterion of ≥.60 and the average $\overline{r_p}$ (.20) satisfies the <.40 criterion. Moreover, the average $\overline{H}$ is nearly identical to that found for the previous WISC-R taxonomy (i.e., .67 for the WISC-III vs. .63 for the WISC-R and the average $\overline{r_p}$ is noticeably superior (.20 for the WISC-III vs. .33 for the WISC-R). The types replicated 100% of the time across the 11 independent experiments and satisfied its a priori criterion of ≥60%.

Table 1
Prevalence and Psychometric Properties of the WISC-III Subtest Taxonomy

Cluster Number (N = 2,200)	% Population Prevalence	Internal Profile Cohesion[a] (H)	External Isolation[b] (r_p)	Independent Replications Across 11 Age Blocks[c] (%)
2	14.9	0.66	0.26	100
3	10.1	0.69	0.29	100
4	13.4	0.69	0.34	100
5	17.2	0.72	0.35	100
6	12.9	0.70	0.25	100
7	14.0	0.68	0.23	100
8	8.4	0.63	−0.10	100
Average	100.0	$\overline{H} = 0.67$	$\bar{r}_p = 0.20$	100

[a]Internal cohesion values indicate the "tightness of fit" of profiles of standard scores within each profile type. If the patterns of children's profiles within each profile type were identical, the homogeneity value would be 1.0. Conversely, as the variability of profile patterns within a type approaches that for the entire WISC-III standardization sample, homogeneity drops toward 0.0. A negative value would indicate variability greater than that for the standardization sample.

[b]External isolation values are averages of the similarity coefficients between the mean standard score pattern for a given profile type and the mean pattern for every other profile type, where the similarity coefficient is calculated by Cattell's (1949) r_p formula. If the mean score pattern for a given type was identical to that of another type, r_p would equal 1.0. As the mean score patterns become more dissimilar and approach a degree of similarity that would approximate chance, r_p drops toward 0.0. A negative r_p results when mean score patterns betwen types are grossly dissimilar.

[c]A replication for a given profile type is defined as confirmation of the existence of that type within a first-stage clustering solution (i.e., within a solution for an age level block of 200 children) using the same profile type confirmation criteria applied to the final solution involving all 2,200 children in the WISC-III standardization sample. Thus, a profile type confirmed in 5 out of 8 first-stage solutions has a 62.5% replication rate.

Corresponding mean subtest scores and deviation IQs are presented in Table 2, along with a descriptive name for each type. The types are arranged in order of descending FSIQs and names are assigned on the basis of this variation plus outstanding VIQ/PIQ contrasts. Terminology such as High and Below Average are chosen to avoid confusion with standard WISC-III intelligence classifications such as "Very Superior," "Low Average," and "Borderline" (Wechsler, 1991, p. 32), the latter referring to normal curve IQ distributions only and not to discrete subtest profile types.

Figure 1 illustrates the relative level and shape of each profile type. The most noticeable distinction among types is general ability level. Also apparent, however, is that the profiles are *not* flat and nearly all tend to display score differences within general ability levels.

Table 2
Mean Subtest Score Patterns and Associated Deviation IQs for the WISC-III Subtest Taxonomy

Profile Type Number	Mean Subtest Score[a]										Mean Deviation Quotient[b]			
(N = 2,200)	PC	IN	CD	SM	PA	AR	BD	VO	OA	CM	FSIQ	VIQ	PIQ	Name and Description
1	13	14	13	14	13	14	15	14	14	14	126	124	124	High ability
2	13	13	10	12	12	12	12	12	12	12	114	113	112	Above average ability
3	10	12	13	12	10	12	10	12	10	13	109	112	104	Above average ability and VIQ > PIQ
4	10	9	13	10	12	10	11	9	11	10	103	97	108	Average ability and PIQ > VIQ
5	10	11	8	11	9	10	10	10	10	10	99	102	96	Average ability and VIQ > PIQ
6	9	7	9	7	9	8	10	7	10	7	89	86	96	Below average ability and PIQ > VIQ
7	7	8	9	8	8	8	6	9	7	9	88	92	85	Below average ability
8	6	5	7	5	6	6	5	5	6	6	73	75	76	Low ability

Abbreviations: WISC-III = Wechsler Intelligence Scale for Children—Third Edition; PC = Picture Completion; IN = information; CD = Coding; SM = Similarities; PA = Picture Arrangement; AR = Arithmetic; BD = Block Design; VO = Vocabulary; OA = Object Assembly; CM = Comprehension; FSIQ = Full Scale IQ; VIQ = Verbal Scale IQ; PIQ = Performance Scale IQ.

Tabled values are rounded to nearest whole number for convenient presentation. The data in this Table are copyright 1995 by The Psychological Corporation. For permission to reproduce, transform, or otherwise adapt these data, contact The Psychological Corporation.

[a]The population standard score Mean = 10 and SD = 3 for each age group.

[b]Deviation quotients are conventional IQ equivalents specific to each age group, where the population Mean = 100 and SD = 15.

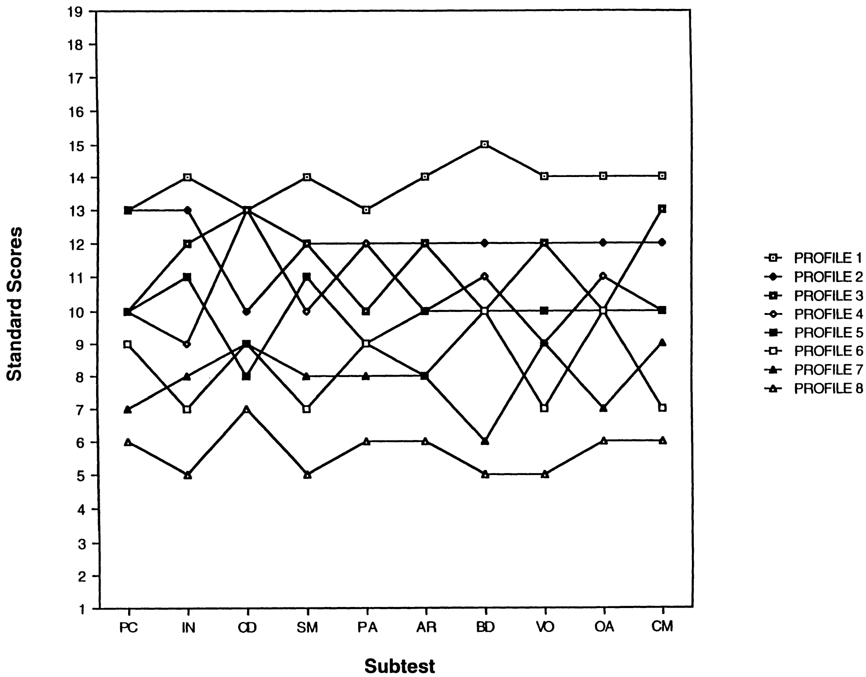

Figure 1. Taxonomy of subtest types in the WISC-III standardization sample. Abbreviations: PC = Picture Completion; IN = information; CD = Coding; SM = Similarities; PA = Picture Arrangement; AR = Arithmetic; BD = Block Design; VO = Vocabulary; OA = Object Assembly; CM = Comprehension.

Inferences of variation corresponding to the VIQ/PIQ dyad are supported by an examination of unusual prevalence distinctions (see Table 3). Unusual IQ discrepancies are defined in the clinical sense (Glutting et al., 1992; McDermott, Glutting, Jones, Watkins, et al., 1989) as those which occur in no more than 3% of the general population. The comparisons make apparent that profile types 3 and 5 are defined not only by general ability, but also by the presence of more VIQ > PIQ discrepancies than would typically be projected. Similarly, profile types 4 and 6 show more PIQ > VIQ discrepancies and profile type 7 is characterized by fewer PIQ > VIQ discrepancies. Interesting also is that deviations for Arithmetic and Coding often coincide directionally as, for example, when the two subtests covary to indicate relatively greater ability (profile types 7 and 8) or lesser ability (profile types 1 and 2).

Typal Membership

Associated characteristics of each type are explained in terms of children's age, sex, ethnicity, education placements, region of the country, and parent

Table 3
Distribution and Prevalence of Verbal/Performance IQ Differences in WISC-III Subtest Taxonomy

Profile Type Number $N = 2{,}300$	Percentage[a] Severe VIQ > PIQ Discrepancy	No Severe Difference	Severe PIQ > VIQ Discrepancy	Prevalence[b]
1	0.4	99.2	0.4	NS
2	3.4	96.0	0.6	NS
3	6.3	93.7	0.0	More VIQ > PIQ* and fewer PIQ > VIQ*
4	0.0	90.1	9.9	More PIQ > VIQ**** and fewer VIQ > PIQ*
5	5.6	94.4	0.0	More VIQ > PIQ* and fewer PIQ > VIQ*
6	0.0	91.6	8.4	More PIQ > VIQ*** and fewer VIQ > PIQ*
7	4.5	95.5	0.0	Fewer PIQ > VIQ*
8	2.2	96.7	1.1	NS

Abbreviations: VIQ = Verbal Scale IQ; NS = not significant; PIQ = Performance Scale IQ.
The sum of percentages across each row is 100%.
[a]Determinations of "severe" IQ differences in profile types is based on cut scores across the WISC-III normative sample, where VIQ-PIQ differences >22 points comprise 3% of VIQ > PIQ differences and PIQ-VIQ differences >24 points comprise x% of PIQ > VIQ differences. The 3% criterion approximates differences nearly two standard deviations above and below the population mean respectively, and is consistent with the standard established by McDermott et al. (1989).
[b]Identification of significant prevalence trends is based on tests of the standard error of proportional differences corrected for the number of simultaneous statistical contrasts by the Bonferroni method.
$*p < .05.$, $**p < .01$, $***p < .001$, $****p < .0001$.

education levels. In each case, prevalence percentages within a profile type are contrasted with expected prevalence as found for the overall WISC-III standardization sample. For reader convenience, we summarize below distinguishing trends for each type. Unless indicated otherwise, only trends found to be statistically significant (i.e., $p < .05$, or less) are reported.

Profile 1. High Ability

(Prevalence = 9.1%; FSIQ Mean = 126.2, SD = 5.5). More Anglos are present than would normally be projected on the basis of race distributions in the population. No Blacks are present and there are fewer Hispanics than normally presumed; the rate for Hispanics is approximately one-half of the population expectancy (i.e., 5.5% for the profile type vs. 10.8% for the WISC-III standardization sample). Over twice the expected number of children have parents who graduated college; more parents graduated from high school and fewer parents failed to complete high school. The proportion of children from the Northeast region exceeds its expectancy and the proportion of children from the South is below expectancy.

Profile 2. Above Average Ability

(Prevalence = 14.9%; FSIQ Mean = 113.9, SD = 4.4). More than 65% of the children are girls and a greater percentage of Anglos is present. Fewer Blacks and Hispanics are evident; Blacks are less than a quarter and Hispanics are less than a half of their anticipated rates. Comparatively more parents graduated college; more parents graduated from high school, and fewer failed to complete high school.

Profile 3. Above Average Ability and VIQ > PIQ

(Prevalence = 10.1%; FSIQ Mean = 108.5, SD = 5.2). The occurrence of unusual VIQ > PIQ discrepancies is higher and unusual PIQ > VIQ is less than that found in the general child population (Mean discrepancy = 8.5 points in favor of the VIQ). More boys are present (65.8%). This type is the only one to show an age effect; more older children (13–16 year olds) and fewer younger children (6–8 year olds) are apparent. There are more Anglos than anticipated and fewer Blacks and Hispanics. Slightly more parents than usual graduated from college, but no differences are present for percentages of parents who graduated high school, or failed to graduate high school.

Profile 4. Average Ability and PIQ > VIQ

(Prevalence = 13.4%; FSIQ Mean = 102.6, SD = 5.2). Profile type 4 shows the greatest disproportion in the number of atypical VIQ/PIQ differences. There are more PIQ > VIQ discrepancies and fewer VIQ > PIQ discrepancies than normally seen in the child population (Mean difference = 11.2 points in favor of the PIQ). Interestingly, with the exception of more boys being present (65.6%), this type shows the fewest demographic and environment disparities. Distributions for age, race, educational placements, parent education levels, and geographic region all align with expectations.

Profile 5. Average Ability and VIQ > PIQ

(Prevalence = 17.2%; FSIQ Mean = 99.1, SD = 4.3). The occurrence of unusual VIQ > PIQ discrepancies is higher and unusual VIQ > PIQ discrepancies is lower than expected (Mean difference = 6.2 points in favor of the VIQ). There are more girls (59.8%) and slightly more Anglos than anticipated.

Profile 6. Below Average Ability and PIQ > VIQ

(Prevalence = 12.9%; FSIQ Mean = 89.3, SD = 4.7). Profile type 6 shows the second greatest disproportion in the number of atypical VIQ/PIQ differences. There are more PIQ > VIQ discrepancies and fewer VIQ > PIQ discrepancies than normal (Mean difference = 10.7 points in favor of the PIQ). Slightly fewer Anglos are included in this type. No difference in proportions is evident for Blacks. However, more Hispanics are present; the rate is over twice the population trend. Fewer parents graduated college or received technical training beyond high school, but proportions for parents who attended, or graduated, high school align with their national rates.

Profile 7. Below Average Ability

(Prevalence = 14.0%; FSIQ Mean = 87.6, SD = 4.8). The frequency of anomalous PIQ > VIQ discrepancies is *lower* than expected, but there is no appreciable difference in the number of VIQ > PIQ discrepancies (Mean difference = 7.5 points in favor of the PIQ). More boys (57.3%) than girls are associated with this type. Over twice the predicted number of Blacks are present; the proportion of Anglos is less than expected and there is no significant difference for Hispanics. Fewer parents graduated college or received technical training after high school than predicted from population trends. A tendency is present for parents to fail high school more often than anticipated, but the finding is significant only at $p < .10$. Similarly, a trend for children to be overrepresented in special education programs. However, here too, the trend is marginally significant (i.e., $p < .10$).

Profile 8. Low Ability

(Prevalence = 8.4%; FSIQ Mean = 73.1, SD = 6.2). Less than half the expected number of Anglos are present. There are more Blacks; the number is over two and one-half times above the population rate. Likewise, slightly more Hispanics are found in this type by than would normally be assumed. Less than a third of the typical number of parents graduated from high school. The proportion of children from the South is above expectancy and the proportion of children from the Northeast is below expectancy. Like profile type 7, there is a marginal trend for children to be overrepresented in special education (i.e., $p < .10$).

DISCUSSION

For over half a century, psychologists have recognized that questions about profile variation are best addressed through nonlinear methods of statistical analysis (Cattell, 1949; Horst, 1941; Mosel & Roberts, 1954; Osgood &

Suzi, 1952; Tatsuoka, 1974; Tatsuoka & Lohnes, 1988). Nevertheless, the predominant research strategy has been to investigate ability profiles using either linear-univariate or linear-multivariate methodologies. The WISC-III taxonomy developed in this paper offers an alternative to linear methods of analysis.

Three advantages can be identified for comparing subtest scores to the WISC-III taxonomy. First, the comparisons take into account the magnitude of intercorrelations among subtest scores. Second, the comparisons are multivariate rather than univariate. Third, and perhaps most importantly, the comparisons employ *nonlinear* multivariate methods, which unlike linear multivariate methods, simultaneously account for differences in profile level and shape. These three factors are not controlled when psychologists test whether statistically significant differences are present among children's subtest scores, nor are the factors controlled when psychologists evaluate whether an unusual univariate base-rate difference is present. Thus, the WISC-III taxonomy provides a mathematically superior method for identifying whether a given subtest profile is clinically relevant and atypical of the most common patterns of intellectual abilities.

The normative taxonomy also makes it possible to conduct at least two kinds of scientific inquiry. First, given the set of most representative profiles in the child population, we can assess and extend our perspective to external phenomena such as demography, environment, and personal factors. This we have attempted to accomplish through comparisons of background characteristics of children comprising the WISC-III standardization sample. Second, and more important diagnostically, comparing subtest scores to the normative taxonomy makes it possible to test the validity of profiles believed to be unusual or nomothetically exceptional.

Methods for Identifying Unusual Profiles

Several strategies can be used to assess claims for whether a given profile is uncommon. However, we recommend two methods that have been useful in our own work. A profile is deemed uncommon in both methods when it is shown that a child's score pattern probably is *not* a member of a core type.

The first method is mathematically more precise, but it is also more complex to calculate. Therefore, it may be more applicable to research than clinical practice. Each of the eight core types is represented by its average (prototypic) score profile within the WISC-III standardization sample. Likelihood of core typal membership is determined by a quadratic multiple discriminant classification (QMDC) which optimizes information concerning differential subtest correlations within profile clusters. Unique clusters are identified by their low probability for membership in any of the core types.

A somewhat less precise but more practical method is based on generalize distance theory (D^2) (Osgood & Suzi, 1952).[1] This second method is the one recommended for everyday decision making. It begins by comparing a child's profile to the three core types closest to his or her general ability level.[2] If the sum of the squared differences for a child's profile is ≥98 for each comparison, the profile may be interpreted as being uncommon. By contrast, if any of the sums is <98, the profile cannot be considered uncommon.

Cutting Scores

Critical values for the QMDC and D^2 procedures are prevalence based. For instance, the critical value just presented for D^2 is associated with a prevalence of 5.4%. This value is relative and can be altered to select a greater, or lesser, percentage of children showing unusual subtest configurations. It was chosen because most psychologists would agree that a prevalence of 5% is sufficiently rare to be considered uncommon. In addition, we developed a computer program to calculate D^2. The program was used to return children from the WISC-III standardization sample to the core profile types. The new placements were compared to children's original placements within the core types. Matches were appreciable when the highest of each child's D^2 values was <98, but decreased noticeably when the highest D^2 was raised successfully to value ≥98 (e.g., 98, 99, 100, etc.). This second line of support is data analytic. It reveals children whose D^2 values are ≥98 bear little resemblance to the core types. Thus, empirical as well as heuristic considerations point to applying critical values associated with a prevalence of 5.4%.

Case Example

The paper concludes with a case example. The case is presented for two reasons: (a) to demonstrate how score comparisons to the core profile taxonomy can be applied in actual practice; and (b) to highlight the benefits of this methodology over an examination of univariate base rates.

The example uses generalize distance theory (D^2) because it is the method recommended for "everyday" decision making. A child named

[1] Copies of computer programs used to calculate quadratic multiple discriminant classifications (QMDC) and D^2 may be obtained from the authors. The QMDC program operates within SAS and the D^2 program works within SPSS. Both can be applied to any sample. The programs read subtest standard scores from a data file, match children to the core types, and print the best classification for each child (one for each of the 8 core types). The programs also can be altered to identify subjects who fail to fit the core types.

[2] Occasionally, it is necessary for clinicians to make 4, or even 5, comparisons to the core profile taxonomy. Such a situation is likely to occur when: (a) a child's subtest pattern is near types 4 and 5, or 6 and 7, and (b) when one of these types is the last of the 3 contrasts.

Mandy obtained a WISC-III FSIQ of 92. A worksheet was used to compare her WISC-III scores to the core types (Table 4). Scanning down the far right-hand column, the psychologist located the FSIQ closest to the one obtained by Mandy. The best match occurred with profile type 6 (FSIQ = 89). Scores for the 10 corresponding WISC-III subtests were entered onto the worksheet at Step 1. Next, Mandy's WISC-III subtest standard scores were entered at Step 2. The psychologist then subtracted 10 times, once for each subtest comparison (Step 3). The difference scores were squared (Step 4) and added (Step 5) to yield a total difference score of 133. The psychologist observed that the comparison did not fit a core type (i.e., a total difference ≥98). Consequently, a second set of comparisons was completed (Step 6). Here too, the total difference score (115) showed that Mandy's profile did not fit a core type and a third comparison was undertaken. Once again, the total difference (173) exceeded the critical value of ≥98.

The psychologist concluded that Mandy had an unusual subtest profile because *each* of the three comparisons yielded a total difference ≥98. Interestingly, Mandy's profile shows a univariate base-rate difference that is quite common and ordinary (i.e., highest subtest score of 13 − lowest subtest of 3 = 10, which then converts to a 19.5% base rate according to Table B.6 on page 266 of the WISC-III manual). Thus, the case example is informative because it demonstrates how accounting for nonlinear, *multivariate* aspects of a profile better identify unusual subtest variation than analyses founded on univariate base rates. Upon identifying an unusual profile, psychologists disposed to interpreting subtest score variation could begin to investigate patterns of dips and rises that may form the basis for hypothesis generation. Thus, profile analysis represents the first step in evaluating children's subtest scores by allowing clinicians to determine whether the profile is unique in a multivariate sense.

At present, however, we cannot advocate the practice of subtest analysis in the generation or formation of hypotheses surrounding children's problems. Our position stems from a lack of empirical evidence that demonstrates subtest analysis is useful for identifying children who demonstrate problems in learning, or psychopathologies in general. Although some research suggests that subtest comparisons provide an effective means for detecting groups with known disorders (Bowers, Risser, Suchanec, Tinker, Ramer, & Domoto, 1992; Plante & Sykora, 1994; Prifitera & Dersh, 1993), other evidence indicates that subtest comparisons are ineffective for this purpose (Glutting & Bear, 1989; Humphries & Bone, 1993; Kavale & Forness, 1984; Watkins & Kush, 1994). Our stance with the latter group emanates, in part, from the methodological pitfalls of circular evidence and inverse probabilities inherent to much of the research advocating the utility of subtest analysis (Glutting, McDermott, Konold, Snelbaker & Watkins, in press).

Table 4
Case Study for Everyday Method to WISC-III Profile Analysis

Profile Type Number	Mean Subtest Score										
	PC	IN	CD	SM	PA	AR	BD	VO	OA	CM	FISQ
1	13	14	13	14	13	14	15	14	14	14	126
2	13	13	10	12	12	12	12	12	12	12	114
3	10	12	13	12	10	12	10	12	10	13	109
4	10	9	13	10	12	10	11	9	11	10	103
5	10	11	8	11	9	10	10	10	10	10	99
6	9	7	9	7	9	8	10	7	10	7	89
7	7	8	9	8	8	8	6	9	7	9	88
8	6	5	7	5	6	6	5	5	6	6	73

Step 1. Enter standard scores for profile type whose FSIQ is nearest to child's FSIQ.

9 7 9 7 9 8 10 7 10 7

Step 2. Enter child's standard scores.

6 13 12 3 7 5 5 11 13 7

Step 3. Subtract to get 10 difference scores.

3 −6 −3 4 2 3 5 −4 −3 0

Step 4. Square each of the 10 difference scores.

9 36 9 16 4 9 25 16 9 0

Step 5. Sum the squared difference scores to get total = 133.

Step 6. Repeat Steps 1 through 5 for the profile type whose FSIQ is second nearest to the child's FSIQ.

Enter standard scores for profile type whose FSIQ is second nearest to child's FSIQ.

7 8 9 8 8 8 6 9 7 9

Enter child's standard scores.

6 13 12 3 7 5 5 11 13 7

Subtract to get 10 difference scores.

1 −5 −3 5 1 3 1 −2 −6 2

Square each of the 10 difference scores.

1 25 9 25 1 9 1 4 36 4

Sum the squared difference scores to get total = 115.

Step 7. Repeat Steps 1 through 5 for the profile type whose FSIQ is third nearest to the child's FSIQ.

Enter standard scores for profile type whose FSIQ is third nearest to child's FSIQ.

10 11 8 11 9 10 10 10 10 10

Enter child's standard scores.

6 13 12 3 7 5 5 11 13 7

Subtract to get 10 difference scores.

4 −2 −4 8 2 5 5 −1 −3 3

Square each of the 10 difference scores.

16 4 16 64 4 25 25 1 9 9

Sum the squared difference scores to get total = 173.

Step 8. Determine whether profile is unusual. If each of the three Total Sums of the Squared Differences scores in ≥98, the profile unusual.

Abbreviations: PC = Picture Completion; IN = information; CD = Coding; SM = Similarities; PA = Picture Arrangement; AR = Arithmetic; BD = Block Design; VO = Vocabulary; OA = Object Assembly; CM = Comprehension; FSIQ = Full Scale IQ.

The method of identifying unusual profiles presented here is empirically superior to approaches that employ clinical intuition, univariate statistical significance, and/or univariate base-rates. We suggest that clinicians who are inclined to interpret subtest variation, do so only after a child's profile is deemed uncommon in a multivariate sense (the method advanced in this article). As illustrated in the case example, had Mandy's profile been evaluated with the traditional univariate base-rate approach, it would have been deemed common. Previous typological testing research indicates similar inconsistencies between univariate and multivariate methods of evaluating subtest profiles (Glutting, McDermott, & Konold, 1997). Thus, clinicians who choose to interpret subtest scores in the absence of a normative taxonomy, are likely to arrive at erroneous conclusions.

This study advances our understanding of subtest analysis by providing a method for determining when a child's profile deviates from what we would commonly expect to observe in the population. However, future research is needed to determine whether children who exhibit uncommon profiles are more likely to experience learning problems than those who do not. This line of inquiry would involve drawing an unselected cohort of children, identifying those individual's whose profiles are distinct from the core types, and later determining whether these individuals are more likely to experience learning problems.

REFERENCES

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

Bayne, R., Beauchamp, J., Begovich, C., & Kane, V. E. (1980). Monte Carlo comparisons of selected clustering procedures. *Pattern Recognition, 12,* 51–62.

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.

Blashfield, R. K. (1976). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin, 83,* 377–388.

Bracken, B. A., McCallum, R. S., & Crain, M. (1993). WISC-III subtest composite reliabilities and specificities: Interpretive Aids. 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. 22–34). Germantown, TN: Psychoeducational Corporation.

Cahan, S. (1986). Significance testing of subtest score differences: The rules of the game. *Journal of Psychoeducational Assessment, 4,* 273–280

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

Cattell, R. B., Coulter, M. A., & Tsujioka, B. (1966). The taxonomic recognition of types and functional emergents. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 288–329). Chicago: Rand McNally.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing profile similarity. *Psychological Bulletin, 50,* 456–473.

Davis, F. B. (1959). Interpretation of differences among averages and individual test scores. *Journal of Educational Psychology, 50,* 162–170.

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., McDermott, P. A., & Konold, T. R. (1997). Ontology, structure, and diagnostic benefits of a 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., Konold, T. R., Snelbaker, A. J., & Watkins, M. W. (In press). More ups and downs of subtest analysis: Criterion validity of the DAS with an unselected cohort. *School Psychology Review.*

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.

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.

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

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

Kavale, K. A., & Forness, S. R. (1984). A meta-analysis of the validity of Wechsler scale profiles and recategorizations: Patterns or parodies? *Learning Disabilities Quarterly, 7,* 136–156.

Kuiper, F., & Fisher, L. (1975). A Monte Carlo comparison of six clustering procedures. *Biometrics, 31,* 777–783.

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, 1,* 118–125.

McDermott, P. A., Glutting, J. J., Jones, J. N., Watkins, M. W., & Kush, J. (1989). Identification and membership of core profile types in the WISC-R national standardization sample. *Psychological Assessment, 1,* 292–299.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika, 45,* 325–342.

Mojena, R. (1977). Hierarchical grouping methods and stopping rules: An evaluation. *Computer Journal, 20,* 359–363.

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

Osgood, C. E., & Suzi, G. J. (1952). A measure of relation determined by both mean differences and profile information. *Psychological Bulletin, 49,* 251–262.

Overall, J. E., Gibson, J. M., & Novy, D. M. (1993). Population recovery capabilities of 35 cluster analysis methods. *Journal of Clinical Psychology, 49,* 459–470.

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 Psychoeducationl Assessment monograph series, advances in psychoeducational assessment: Wechsler Intelligence Scale for Children-Third Edition* (pp. 43–55). Germantown, TN: Psychoeducational Corporation.

Salvia, J., & Good, R. (1982). Significant discrepancies in the classification of pupils: Differentiating the concept. In J. T. Neisworth (Ed.), *Assessment in special education* (pp. 77–82). Rockville, MD: Aspen Systems.

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

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

Silverstein, A. B. (1981). Reliability and abnormality of test score differences. *Journal of Clinical Psychology, 37,* 392–394.

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

Sneath, P., & Sokal, R. (1973). *Numerical Taxonomy.* San Francisco: Freeman.

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

Stevens, J. (1986). *Applied multivariate statistics for the social sciences.* Hillsdale, NJ: Erlbaum.

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* (2nd ed.). New York: Macmillan.

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

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

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. (1974). *Wechsler Intelligence Scale for Children-Revised.* New York: Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised.* New York: Psychological Corporation.

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

Wishart, D. (1982). *CLUSTAN users manual: Supplement* (3rd ed). Edinburgh, Scotland: Program Library unit, Edinburgh University.