Psychological Assessment:
A Journal of Consulting and Clinical Psychology
1989, Vol. 1, No. 4, 292–299

1040-3590/89/$00.75

Core Profile Types in the WISC–R National Sample: Structure, Membership, and Applications

Paul A. McDermott
Graduate School of Education
University of Pennsylvania

Joseph J. Glutting
Department of Educational Studies
University of Delaware

Jeffrey N. Jones
Ball Systems Engineering
Columbia, Maryland

Marley W. Watkins
SouthWest EdPsych Services
Phoenix, Arizona

Joseph Kush
Deer Valley Unified Schools
Phoenix, Arizona

The structure and composition of subtest profile types most representative of the 2,200 6½- to 16½-year-old children comprising the normative sample for the Wechsler Intelligence Scale for Children–Revised (WISC–R; Wechsler, 1974) were explored. Profiles were sorted according to similar level and shape using multistage cluster analysis with independent replications. A final solution of 7 core profile types met all formal heuristic and statistical criteria, including satisfactory homogeneity, coverage, replicability, and stability over a 1-mo. period. Core types were described according to population prevalence, ability level, and configuration, and each type was examined for membership trends by child demography, family characteristics, and abnormal IQ discrepancies. Methods are given for determining the relative uniqueness of WISC–R profile patterns in future research and clinical work.

Forty years ago, Wechsler (1949) introduced his first tests of childhood intelligence. Originally intended as alternative measures of global ability that permitted distinction between aspects of verbal and nonverbal functioning, the tests soon were viewed as useful for even finer distinctions among children's cognitive styles, especially as reflected in patterns of score elevations and depressions across subtest areas. Based mainly on popular theory and inductive analysis about skills required for good performance on Wechsler subtests, interpretations have since been offered for more than 75 different patterns of subtest variation (Bannatyne, 1974; Glasser & Zimmerman, 1967; Guilford, 1967; Kaufman, 1979; Saccuzzo & Lewandowski, 1976; Selz & Reitan, 1979; Wechsler & Jaros, 1965; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962).

With publication of the Wechsler Intelligence Scale for Children–Revised (WISC–R; Wechsler, 1974), a new and somewhat more empirical interest in subtest analysis emerged. Clinical researchers recognized the need to validate inferences by establishing relationships between specific subtest profiles and meaningful external criteria. Thus, direct comparisons of groups of diversely diagnosed children have led some investigators to conclude that subtest profiles are helpful in differentiating among the emotionally, mentally, and learning impaired and among subtypes of underachieving and delinquent children (Dean, 1978, 1980; Hubble & Groff, 1980; Naglieri, 1980; Rourke & Strang, 1984; Vance, Fuller, & Ellis, 1983). In contrast, Hale (1979), Hale and Landino (1981), and Thompson (1980, 1981) found subtest patterns rather ineffective for such group discrimination—Hale and his associates (Hale & Raymond, 1981; Hale & Saxe, 1983) further demonstrated that the discriminatory and predictive efficiency of WISC–R profiles adds nothing to that already afforded by global IQ measures.

As preparation for the research reported in this article, we reviewed 53 empirical studies pertaining to WISC–R profile analysis, each having been drawn from the larger body of some 2,000 WISC–R works appearing during the 1974–1988 period.[1] Our review uncovered several pervasive methodological problems. First, researchers generally assume that intact groups of similarly diagnosed children accurately represent meaningful, if not homogeneous, categories—an assumption at variance with the evidence in child psychology and special education (Garfield, 1978; McDermott, 1988). Second, available studies fail to preclude circular use of WISC–R profiles for both initial formation of diagnostic groups and subsequent searches for profiles that might naturally define those groups. Third, it has been demonstrated that precision of measurement for WISC–R

Correspondence concerning this article, and requests for indicated supplementary materials, should be addressed to Paul A. McDermott, Graduate School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia, Pennsylvania 19104-6216.

[1] An earlier meta-analytic review by Kavale and Forness (1984) considered 94 investigations on Wechsler profile analysis. However, only 31 of the investigations pertained exclusively to the WISC–R; the remaining 63 concerned the original Wechsler Intelligence Scale for Children (Wechsler, 1949) or other Wechsler tests.

subtests tends to vary, sometimes dramatically so, from childhood through adolescence (Conger, Conger, Farrell, & Ward, 1979), which makes it difficult to generalize research conclusions from one age level to another.

Perhaps the most important consideration in profile research is the choice and evaluation of relevant hypotheses. In our review of research from the past 15 years we found few instances where claims for discovery of a unique WISC–R profile were assessed against a viable null hypothesis—namely, that such a profile was commonplace in the general population of children and thus unremarkable. Instead, claims for profile uniqueness tend to be grounded in surmises either that the average profile for a group of similarly diagnosed youngsters is inherently characteristic of the diagnostic category and uncharacteristic of alternative categories, or that evidence for differences in average profiles between diagnostic categories is tantamount to proof that such profiles are unlikely to emerge in the overall normal population. Frankly, without a normative typology of core profiles commonly existing among children, we simply cannot know whether subtest profiles elsewhere discovered are uncommon, distinctive, or clinically meaningful.

The research reported in this article was undertaken to define an empirical typology of core subtest profiles existing within the population of normal children. Such a normative typology would provide the necessary contrasts for testing hypotheses about unique profile variation. In addition, we recognized that, given the unavailability to date of a population typology to explain how subtests typically vary on a child-by-child basis, it would be important to determine how distinct ability patterns vary as a function of vital demographic and environmental factors.

Method

Samples

The overall typology was based on the entire sample of 2,200 children and adolescents used in the WISC–R national standardization study (Wechsler, 1974). Subjects were selected according to a stratified quota system including 200 children at each of 11 age levels from 6½ through 16½ years, with equal numbers of boys and girls at each level. Quotas for distribution of children's race, occupation group for head of the household, geographic region, and urban versus rural residence were arranged to approximate distributions identified in the U.S. Census. Children exhibiting severe emotional disturbance and those institutionalized with mental deficiency were excluded from the sample.

Typological stability analyses were conducted with the 303 children selected from the WISC–R standardization sample for test–retest study (as reported by Wechsler, 1974, pp. 29–31). The subsample included 51 children at age 6½, 46 at 7½, 50 at 10½, 52 at 11½, 51 at 14½, and 53 at 15½, with distributions for children's sex, race, and parental occupation proportionate to U.S. Census figures.

Profile Components

Each child's profile was based on scaled scores ($M = 10$, $SD = 3$) for 11 WISC–R subtests, including the "mandatory" 5 Verbal and 5 Performance subtests (Wechsler, 1974, p. 8) and the supplementary Digit Span subtest. Because the more popular Coding subtest was included as a mandatory part of the Performance scale, the alternate Mazes subtest was unused. Both Digit Span and Coding are regarded as primary components in most profile analysis schemes (Kaufman, 1979, pp. 149–152, 170–171), whereas Mazes often is excluded (e.g., see Bannatyne, 1974; Guilford, 1967; Witkin et al., 1962).

Criterion Variables

Internal criteria. Children's obtained deviation IQs ($M = 100$, $SD = 15$) for the Full Scale (FSIQ) and for the Verbal (VIQ) and Performance (PIQ) scales were used to help describe and interpret the final typology. As prescribed in the WISC–R manual (Wechsler, 1974), these values were based on mandatory subtests only, thus excluding Digit Span. Also, prevalence of abnormal VIQ/PIQ discrepancies within profile types was used to support interpretations regarding unusual profile configurations. Abnormal discrepancies were defined in the clinical sense (McDermott & Watkins, 1987) as those that occur in no more than 3% of the general population.

External criteria. Unlike deviation IQ measures that are actually transformed linear composites of the subtests themselves, certain variables were used both to describe and lend validity to the typology. These included the WISC–R stratification variables of child age, sex, and race and the occupational status of the head of the household. Also included were supplementary variables of interest collected at the time of standardization (although not reported in the WISC–R manual)—father's and mother's level of education, the child's birth order, and the number of children in the family.

Procedure

By their nature, subtest profiles are doubly defined according to *level* (position toward the upper, central, or lower range of the ability continuum) and *shape* (the pattern of peaks and valleys across subtest scores). The idea was to sort 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). Moreover, the groups of similar profiles (called *profile types*) must be reasonably replicable across age levels rather than spurious mergers that would occur by chance. The overall solution (or *typology*) should account for all profile variation in the population (known as full coverage) and not discount profiles that happen to diverge from the more popular trend. This is particularly important for a typology intended to be fully representative of the general population of children. Finally, the typology should yield reasonable stability for typal membership; that is, children's initial association with specific core profile types should remain constant on subsequent assessment—at least within the limits set by the temporal stability of underlying ability constructs.

To this end, we evaluated the appropriateness of numerous procedures for clustering profiles and determined that Ward's (1963) minimum-variance procedure best satisfied the research goals. Monte Carlo studies of competing clustering methods have shown consistently that when full coverage is required, Ward's method gives superior recovery of known typological structure (Kuiper & Fisher, 1975; Mojena, 1977) and outperforms other methods in reducing overlap (Bayne, Beauchamp, Begovich, & Kane, 1980). Ward's is also the most accurate under mixture model testing, where individuals must be classified to diverse known populations (Blashfield, 1976). In contrast, average-linkage clustering (the best alternative to the minimum-variance approach) does comparatively poorly in reducing overlap (Bayne et al., 1980; Milligan, 1980), and, in preliminary analyses with the WISC–R data, both average-linkage clustering and clustering based on ipsatized subtest scores produced typologies that were unreplicable across experiments, temporally unstable, and uninterpretable in the light of external criterion variables.[2]

[2] These results are consistent with the fact that, as based on the first unrotated principal factor extracted from the WISC–R normative data

Our global clustering strategy was a variation on the model suggested by Overall and Klett (1972, pp. 215–216) for determining the most representative profile patterns in a population. 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 through Ward's method. For each block, identification of the hierarchical step providing the most ideal clustering solution was based on several criteria: An ideal solution must (a) have a ratio of within-profile-type (cluster) variances to variance for the full standardization sample < 1.0, (b) correspond to a hierarchical step preceding atypical inflection in Ward's (1963) total error sums of squares statistic (E) with no reduction in increment rates occurring at subsequent steps, (c) yield an average within-profile-type homogeneity coefficient, $\bar{H}$ (Tryon & Bailey, 1970), >.60, and (d) yield an average between-profile-types similarity coefficient, $\bar{r}_p$ (Cattell, 1949), < .40. $\bar{H}$ and $\bar{r}_p$ are both sensitive to similar profile levels and shapes and are interpreted much like correlation coefficients, where the value 1.0 indicates profiles identical in level and shape, 0.0 indicates chance similarity based on the full WISC–R sample, and negative values indicate gross dissimilarity. The respective .60 and .40 a priori criteria were established through clustering and classification studies with random samples larger than the WISC–R sample (McDermott, 1980; McDermott & Watkins, 1987).

Clusters derived from the 11 independent analyses were pooled to form a set of first-stage clusters that were themselves subjected to second-stage clustering by Ward's method. Second-stage clustering began with a similarity matrix whose diagonal elements held E values for respective first-stage clusters, with off-diagonal elements corresponding to potential Es for merging each pair of first-stage clusters. The final clustering solution used the same criteria employed for first-stage clustering and, in addition, ensured that the solution satisfied the more stringent criteria afforded by Mojena's (1977) first stopping rule, Wishart's (1982) t test, and a replication rate for each final cluster > 50% across 11 independent experiments.

The two-stage clustering model served a dual purpose. On the one hand, it circumvented the practical infeasibility of a one-stage model requiring simultaneous consideration of nearly 4.8 million data points for the 2,200 subjects. On the other hand, it gave an opportunity for built-in replications of the final typology. Thus, the replicability rate for each final cluster was determined by the number of first-stage solutions (independent experiments) in which it also emerged, where emergence was verified by its subsequent second-stage absorption into the same cluster and corresponding $r_p \geq .90$ for that cluster.

The various internal and external criterion measures were used to describe or lend validity to each final profile type. Thus, considering the distribution (prevalence) of each pertinent criterion variable within a profile type, we conducted two-tailed tests of the standard error of proportional differences (Ferguson & Takane, 1988) for all possible pairwise comparisons across levels of the criterion variable, with Type I error apportioned across comparisons by the Bonferroni correction (Miller, 1966). By this approach, expected prevalence for a given characteristic (e.g., Whites vs. non-Whites) within a profile cluster was based on prevalence for the entire population, and unusual prevalence for the profile cluster was determined by statistically significant deviations from general expectancy.

(computed from figures given by Kaufman, 1979, Table 4.1, p. 110), most reliable subtest variance is associated with children's global ability levels. Ward's method tends to emphasize such level differences when sorting profiles (Blashfield & Aldenderfer, 1988). Alternatively, average-linkage clustering and clustering of ipsatized subtest scores substantially diminish or eliminate the influence of ability levels, yielding solutions unacceptably insensitive to the prevailing source of reliable WISC–R profile variation.

Profile stability was determined using the subsample of 303 children selected for WISC–R test–retest analysis. Nearly all children were retested between 3 and 5 weeks after first testing (the average interval being 1 month). Because all retest scores showed increments due to practice (see Wechsler, 1974, pp. 32–33, Table 11), the mean test–retest increment for each subtest by age level was subtracted from respective retest scores. Assignment of children's profiles to core profile types at initial testing corresponded to original membership in second-stage clusters, as described previously. Assignment of profiles produced at the time of retesting was based on an iterative classification procedure. Essentially, the 303 retest profiles were, first, classified into second-stage clusters using Cattell's formula for assessing similarity of a profile to the mean profiles of diverse groups (see Tatsuoka & Lohnes, 1988, pp. 377–378) as based on average profiles for second-stage clusters and, second, reclassified using Tatsuoka's (1974, pp. 21–28) maximum-probability method based on a priori probabilities gained from classifications at the first iterative step and within-cluster covariance matrices from second-stage clusters. Thereafter, we assessed agreement between test and retest profile classifications through Fleiss's (1971) adaptation of coefficient κ using computer program CONGRU (Watkins & McDermott, 1979), where stability for individual profile types corresponded to partial κ and stability for the entire typology corresponded to overall κ.

Results

Typal Structure

First-stage clustering produced 80 profile groups (an average of 7.3 per analysis). These were submitted to second-stage analysis based on an 80 × 80 similarity matrix, and the solution at all hierarchical steps was assessed against the stated criteria. The seven-cluster solution was the only one to satisfy all criteria and, therefore, was selected as the best overall typology of core WISC–R profile types. This solution showed a ratio of sum of within-profile-type variances to full sample variance of 0.8, whereas the ratio for the subsequent (six-cluster) solution was 1.3. Next we considered the pattern of increments in E: The seven-cluster solution occurred immediately prior to an increment 3.5 times greater than any prior increment, with no reductions or plateauing of increment rates occurring thereafter; this indicated that no solution containing fewer than seven types would better explain the underlying typology. Moreover, whereas the preceding eight-cluster solution satisfied Wishart's significance test, $t(76) = 4.30$, $p < .0005$, but failed Mojena's stopping criterion, the seven-cluster solution satisfied both criteria, $t(76) = 8.69$, $p < .0005$, thus indicating that no solution containing more than seven types offered better structure resolution.

Table 1 displays for each of the seven core profile types its prevalence in the child population, average coefficients for within-type homogeneity and between-types similarity, and replication rate. Note that the $\bar{H}$ value (.63) satisfied the a priori criterion of .60 and that the $\bar{r}_p$ value (.33) satisfied the .40 criterion. On average, across the 11 age-level experiments, the types replicated 84.4% of the time, the lowest replication rate (63.6%) being found for the two rarest profile types (as indicated by the 8.3% prevalence figure for Profile Type 5 and the 4.0% figure for Profile Type 7). The rarer types also were the only ones to fail replication at contiguous age levels, with Type 5 not emerging among 10- or 11-year-olds nor Type 7 among 11- or 12-year-olds.

Table 1
Prevalence, Psychometric Properties, and Names of WISC-R Core Profile Types

Profile type	% population prevalence	Within-type homogeneity (H)	Between-type similarity (r_P)	% replicability across 11 solutions	Descriptive name (and symbol)
1.	13.9	.58	.12	100.0	High (H)
2.	18.4	.62	.39	90.9	Above average (AAv)
3.	15.1	.64	.48	81.8	Slightly above average (SAAv)
4.	22.6	.65	.50	100.0	Average (Av)
5.	8.3	.69	.47	63.6	Slightly below average (SBAv)
6.	17.7	.64	.36	90.9	Below average (BAv)
7.	4.0	.62	−.03	63.6	Low (L)
Average		.63	.33	84.4	

Note. N = 2,200. WISC-R = Wechsler Intelligence Scale for Children-Revised. The data in this table are adapted from *Research Report No. 89-3* by P. A. McDermott, 1989, San Antonio, TX: The Psychological Corporation.

Also shown in Table 1 is a descriptive name for each profile type. Corresponding mean subtest patterns and deviation IQs are presented in Table 2. Types are arranged in order of descending FSIQs and bear corresponding names. Terminology such as High and Slightly Below Average was chosen to avoid confusion with standard WISC-R intelligence classifications such as "Very Superior," "Low Average (Dull)," and "Borderline" (Wechsler, 1974, p. 26), 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. Clearly, the predominant distinction among types is general ability level. Also apparent, however, is that among normal children, not only are prototypic profiles not flat, but they tend to follow various configural parallels across ability level. For example, the Below Average and Low types show nearly identical shapes across all subtests, and the Above Average type differs by only one subtest (Block Design). Similar parallelism is noted between profiles for Verbal subtests comprising the Above Average and Slightly Below Average types (Digit Span being the exception) and between Performance subtests for Slightly Above Average and Below Average types (Coding being the exception). It is interesting also that deviations for Digit Span and Coding often coincide directionally with deviations for Arithmetic as, for example, when all three covary to indicate relatively greater ability (refer to Above Average, Slightly Below Average, Below Average, and Low types) or lesser ability (Slightly Above Average type).

Typal Membership

Prevailing composition of each type was explained in terms of children's age, sex, race, birth order, and abnormal VIQ/PIQ discrepancies, the number of children in the family, family occupational status, and fathers' and mothers' education levels. In each case, prevalence percentages within a profile type were

Table 2
Mean Subtest Score Patterns and Deviation IQs for WISC-R Core Profile Types

	Mean subtest scaled score[a]											Mean deviation quotient[b]		
Profile type	IN	SM	AR	VO	CM	DS	PC	PA	BD	OA	CD	VIQ	PIQ	FSIQ
High	14	14	13	14	13	12	13	12	13	12	13	121	116	121
Above Average	11	11	12	11	11	12	11	11	12	11	13	107	109	109
Slightly Above Average	11	11	10	11	12	10	10	11	10	10	10	106	102	105
Average	9	10	9	9	10	9	11	10	10	11	12	96	101	98
Slightly Below Average	9	9	10	9	9	10	8	9	8	8	9	95	90	91
Below Average	7	7	8	7	8	8	8	8	7	8	9	84	85	84
Low	6	5	6	5	6	6	6	6	6	6	7	72	73	70

Note. N = 2,200. Tabled values are rounded to the nearest whole number for convenient presentation. WISC-R = Wechsler Intelligence Scale for Children-Revised; 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; V = Verbal; P = Performance; FS = Full Scale. The data in this table are adapted from *Research Report No. 89-3* by P. A. McDermott, 1989, San Antonio, TX: The Psychological Corporation.
[a] The population scaled score M = 10 and SD = 3 for each age group.
[b] Deviation quotients are conventional IQ equivalents specific to age group, where the population M = 100 and SD = 15.

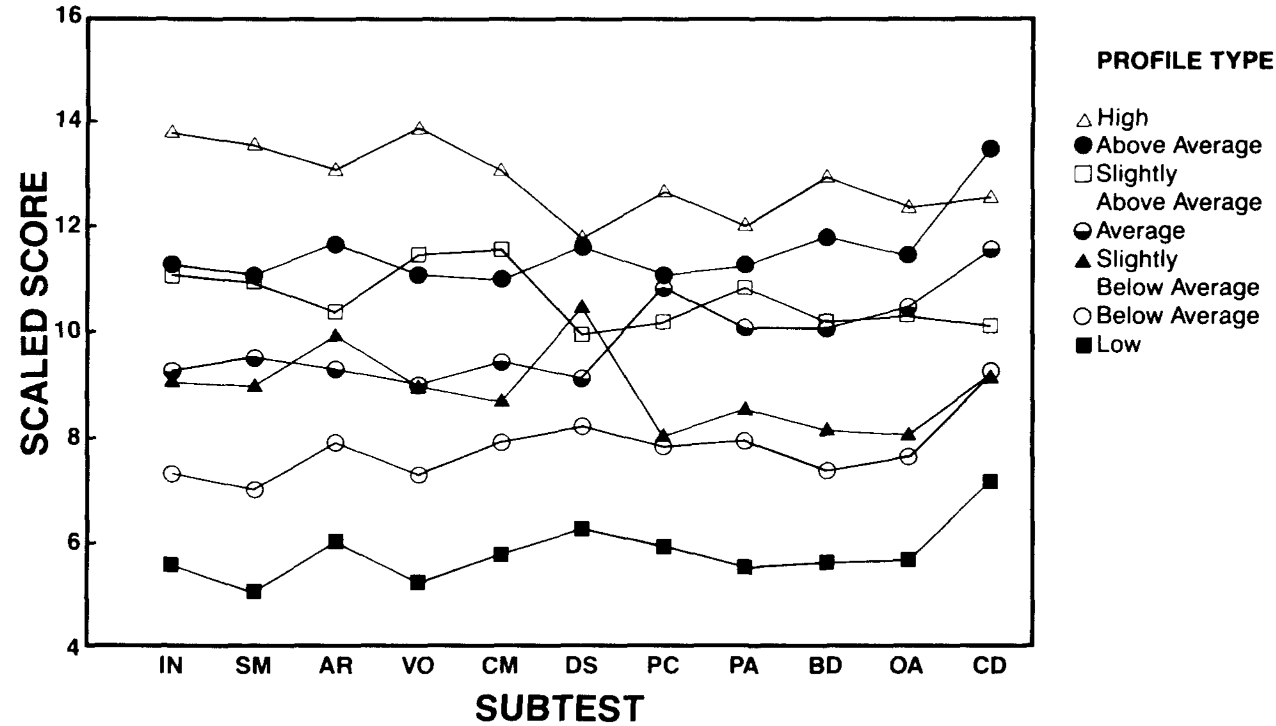

Figure 1. Mean subtest patterns for Wechsler Intelligence Scale for Children–Revised core profile types (IN = Information; SM = Similarities; AR = Arithmetic; VO = Vocabulary; CM = Comprehension; DS = Digit Span; PC = Picture Completion; PA = Picture Arrangement; BD = Block Design; OA = Object Assembly; CD = Coding).

contrasted with expected prevalence as found for the overall normative sample. For the reader's convenience, we next summarize distinguishing prevalence trends for each profile type. Only those trends found statistically significant are reported.[3]

High type. Prevalence = 13.9%; FSIQ, M = 121.4, SD = 8.5. The occurrence of abnormal VIQ > PIQ discrepancies is higher, and abnormal PIQ > VIQ discrepancies lower, than found in the general population (mean discrepancy = 5.1 IQ points in favor of the VIQ). More than 60% of this type are boys, with somewhat more preadolescents and fewer younger children than expected. The proportion of non-White children is less than one third that anticipated from general population racial distributions. Nearly one third are from families with professional occupational status, and two thirds are from families with at least white-collar occupational status. More than half of the fathers and a third of the mothers have some postsecondary education. The number of these children who are second born is double those who are fourth or later born, with significantly more coming from two- or three-child families than from families having five or more children.

Above average type. Prevalence = 18.4%; FSIQ, M = 108.7, SD = 7.8. The number of non-White children is less than half that anticipated from overall population expectancy. About half of these children have at least one parent with a white-collar job. Comparatively more fathers have at least finished high school (80%) than have attended elementary school only (7.4%), the general population figures showing only 68% of fathers finishing high school and about 12% attending elementary school only. Also, significantly more mothers have some postsecondary education (28.7% vs. 21.6% in the general population) than have an elementary education only (5.2% for this type vs. 8.3% for the overall population).

Slightly above average type. Prevalence = 15.1%; FSIQ, M = 104.7, SD = 6.9. There are more abnormal VIQ > PIQ discrepancies and fewer PIQ > VIQ discrepancies than found in the general population (mean discrepancy = 4.0 points in favor of the VIQ). These children are slightly more often pubescent or adolescent, with fewer being non-White than expected by population trends. More children's fathers than expected have 4 or more years of postsecondary education or at least some high school education than have elementary schooling only, and significantly more mothers have completed high school or additional schooling than have elementary schooling only. The proportion of firstborn children is relatively higher than the proportions of fourth- or later-born children.

Average type. Prevalence = 22.6%; FSIQ, M = 97.9, SD = 7.0. The frequency of abnormal VIQ > PIQ discrepancies is lower than expected (mean discrepancy = 5.8 points in favor of the PIQ). Fewer of these children than expected are non-White, and significantly more are young or preadolescent children than adolescents of any age. Relatively more come from families having skilled or semiskilled occupational status, and fewer come

[3] Supplementary tables showing exact population and core profile prevalence rates and statistical tests for all internal and external criterion variables may be obtained by writing to Paul A. McDermott.

from families having professional status. Comparatively more children's mothers have attended or completed high school than have completed 4 or more years of postsecondary education.

Slightly below average type. Prevalence = 8.3%; FSIQ, M = 91.3, SD = 5.9. The proportion of girls and of non-White children is significantly higher than expected from population estimates. Two out of five of these children are young adolescents. Significantly more are third- than second-born children.

Below average type. Prevalence = 17.7%; FSIQ, M = 83.5, SD = 6.9. More girls than boys are associated with this profile type, as are substantially more non-Whites than predicted from population racial distribution. The proportion of younger children exceeds that for either preadolescents or early adolescents. More than three fourths are from households with blue-collar occupational status, compared to one fourth from households with white-collar status. Approximately 80% of the children's fathers have attained no formal education beyond high school, and a general trend is noted for mothers to have comparatively less formal education than found across the overall population. Significantly more fourth- or later-born children than firstborn children are evident, with significantly more coming from families having five or more rather than three children.

Low type. Prevalence = 4.0%; FSIQ, M = 70.3, SD = 8.5. Over half of these children are non-White, the frequency being threefold that expected. More than 65% of the children have semi- or unskilled working families compared to a 35% distribution of such families throughout the greater population. The number of fathers having no more than grammar school education is more than double the population expectancy, with the number of fathers having completed at least high school (39%) substantially lower than found in general (68%). Also, the number of mothers not having completed high school (59%) substantially exceeds the general population trend (24%). About 40% of the children come from families having five or more children, with the percentage of children from three-child families being less than half that seen in the general population.

Typal Stability

The simple stability rate for the overall typology was 64.7% (which is 57.5% beyond chance, $p < .0001$), thus indicating appreciable stability. Partial κ values for most profile types achieved significance beyond the .01 level, whereas stability for the Slightly Below Average type was significant only at the .06 level. Recall that the latter type was among the rarest and was not replicated among 10- or 11-year-old children. Inspection of the test–retest data revealed that because one third of the sample involved 10- and 11-year-olds, general stability for the Slightly Below Average type was underestimated (its significance reaching the .001 level upon exclusion of such age groups). Therefore, typal stability was supported for the full typology and for individual profile types.

Discussion

The most striking features of the typological structure are the distinctions between levels of global ability. Indeed, as estimated from alternate variance components and the first unrotated principal factor for the WISC–R standardization sample (using information drawn from Kaufman, 1979, Tables 4.1–4.2), it is evident that nearly 60% of the scale's reliable variance is associated with Spearman's *g*. This is consistent with the conclusion by Conger et al. (1979) that the most reliable comparisons for WISC–R profiles are those across general intelligence levels (as commensurate with the FSIQ).

Applying various factor-analytic procedures to assess subtest variation in the WISC–R normative sample, Kaufman (1975) and Kroonenberg and ten Berge (1987) extracted three salient group dimensions: (a) Verbal Comprehension, as defined primarily by Verbal scale subtests, (b) Perceptual Organization, based mainly on Performance scale subtests, and (c) Freedom From Distractibility, which, although its constituent subtests differ somewhat at certain ages, is usually defined by the Arithmetic, Digit Span, and Coding subtests. It is rather likely that the observed configural parallels for Verbal subtests across certain profile types and for Performance subtests across other types stem from the factorial bifurcation of the Verbal Comprehension and Perceptual Organization dimensions. This is further consistent with the evidence showing that, after the more robust and reliable variation of children's global ability has been considered, the most valuable WISC–R comparisons are those rooted in Verbal versus Performance scale variation (Conger et al., 1979). Moreover, it seems fair to conclude that the systematic directional covariation of the Arithmetic, Digit Span, and Coding subtests comports rather convincingly with the independent variation of the Freedom From Distractibility factor.

Prior research has suggested some gender differences in the WISC–R normative sample. Specifically, Kaufman and Doppelt (1976) reported boys' average VIQ to be 2.4 points and FSIQ 1.8 points above corresponding values for girls. Although significant statistically, the differences were interpreted as relatively inconsequential—a conclusion that makes sense if such differences are dispersed somewhat randomly throughout ability levels. However, those comparisons were based on aggregates of children irrespective of ability level. In contrast, the profile typology gives perspective on gender differences as they occur for various levels of intellectual ability. Given the prevalence for boys in High type profiles and for girls in both Slightly Below Average and Below Average profiles, it seems apparent that gender differences are not evenly distributed and do tend to produce some overinclusion of girls toward the lower portion of the ability continuum.

The most popular view on gender differences in intelligence holds that females manifest relative superiority for verbal ability. Based on long-standing patterns of differential performance on verbally loaded academic tasks and intellective measures, researchers have noted a female advantage that begins to emerge during preschool years and becomes more obvious and stable with pubescence (Anastasi, 1958; Denno, 1982; Halpern, 1986; Maccoby, 1966). But a more recent review and meta-analysis by Hyde and Linn (1988) concluded that available research presents no grounds for the superiority contention. This conclusion is consonant with the typological coprevalence of boys and more VIQ > PIQ discrepancies in the High profile type and of girls and fewer VIQ > PIQ discrepancies within the Slightly Below Average type.

We have indicated that the absence of a normative typology of WISC–R subtest profiles has impeded research and clinical practice. A typological approach is more appropriate for studying individual differences when all areas of ability are to be considered simultaneously, as is most often the case in psychological assessment. However, one must exercise caution in accepting any particular cluster solution as ideal. It is possible that some alternative solution might better explain profile variation. Moreover, cluster solutions are difficult to validate in the absolute sense; that is, although the observed distributions of children's demography and other external criteria comport with grounded theory, such distributions could also emerge with other typological solutions. Nevertheless, we hold much confidence in the current solution because, as previously noted, we tested alternative solutions empirically in preliminary analyses. Invariably they produced results unreplicable across experiments, temporally unstable, and uninterpretable in the light of the available external criteria.

The WISC–R normative typology makes possible at least two kinds of scientific inquiry. First, given the set of most representative profile types in the child population, we can reassess and extend our perspective on how natural variation in human ability relates to external phenomena such as demography and environment. This we have attempted to accomplish with theoretically interesting characteristics of the children comprising the WISC–R national sample. Second, a normative typology makes it possible to test the validity of profiles believed to be descriptively or clinically unique.

A profile is deemed unique only when it can be shown that it is probably not a member of a core type in the population. Each core type is represented by its mean subtest profile, and subtest intercorrelations are represented by the variance–covariance matrix specific to each core type.[4] Likelihood of core typal membership is determined using the $r_p(k)$ group similarity coefficient for correlated variables (Tatsuoka, 1974), where a separate $r_p(k)$ value reflects level and shape similarity of the hypothetically unique profile to each core type. A coefficient $\geq .40$ suggests reasonable similarity to a core type. If all seven $r_p(k)$ values for a profile are $< .40$, the null hypothesis may be rejected and the profile regarded as appreciably distinct. Alternatively, the null hypothesis cannot be rejected and it must be concluded that the profile thought unique actually represents a common or natural variant of normal childhood intellectual abilities.

A less precise but more convenient method (based on generalized distance theory) is recommended for everyday clinical judgments about profile uniqueness. Here the subtest profile produced by a tested child is compared only to those core profile types within the child's general ability range (using the scaled scores reported in Table 2). Scanning across the subtests for a given core type and the child's profile, calculate the difference in scaled score points between corresponding subtests of each profile. Square each such difference and sum the squared difference across the 11 subtests. If the sum of squared differences between the child's profile and any core profile in the child's ability range is ≥ 80, the child's profile may be interpreted as uncommon in the general population.[5] Otherwise, the child's profile should be considered commonplace and indistinctive.

References

Anastasi, A. (1958). *Differential psychology* (3rd ed.). New York: Macmillan.

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

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

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

Blashfield, R. K., & Aldenderfer, M. S. (1988). The methods and problems of cluster analysis. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 447–473). New York: Plenum.

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

Conger, A. J., Conger, J. C., Farrell, A. D., & Ward, D. (1979). What can the WISC–R measure? *Applied Psychological Measurement, 3,* 421–436.

Dean, R. S. (1978). Distinguishing learning-disabled and emotionally disturbed children on the WISC–R. *Journal of Consulting and Clinical Psychology, 46,* 381–382.

Dean, R. S. (1980). They do when compared with emotionally disturbed children: Reply to Sattler. *Journal of Consulting and Clinical Psychology, 48,* 256–257.

Denno, D. (1982). Sex differences in cognition: A review and critique of longitudinal evidence. *Adolescence, 17,* 779–788.

Ferguson, G. A., & Takane, Y. (1988). *Statistical analysis in psychology and education* (6th ed.). New York: McGraw-Hill.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76,* 378–382.

Garfield, S. L. (1978). Research problems in childhood diagnosis. *Journal of Consulting and Clinical Psychology, 46,* 596–601.

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

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

Hale, R. L. (1979). The utility of WISC–R subtest scores in discriminating among adequate and underachieving children. *Multivariate Behavioral Research, 14,* 245–253.

Hale, R. L., & Landino, S. A. (1981). Utility of WISC–R subtest analysis in discriminating among groups of conduct problem, withdrawn, mixed, and non-problem boys. *Journal of Consulting and Clinical Psychology, 41,* 91–95.

Hale, R. L., & Raymond, M. R. (1981). Wechsler Intelligence Scale for Children–Revised (WISC–R) patterns of strengths and weaknesses as predictors of the intelligence–achievement relationship. *Diagnostique, 7,* 35–42.

Hale, R. L., & Saxe, J. E. (1983). Profile analysis of the Wechsler Intelligence Scale for Children–Revised. *Journal of Psychoeducational Assessment, 1,* 155–162.

Halpern, D. F. (1986). *Sex differences in cognitive abilities.* Hillsdale, NJ: Erlbaum.

Hubble, L. M., & Groff, M. (1980). WISC–R profiles of adjudicated de-

[4] Exact mean scaled score values and covariance matrices for subtests within WISC–R core profile types may be obtained by writing to Paul A. McDermott.

[5] If 10 subtests are compared, the sum of squared differences must be ≥ 73 to support the uniqueness hypothesis.

linquents later incarcerated or released on probation. *Psychological Reports, 47,* 481–482.
Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104,* 53–69.
Kaufman, A. S. (1975). Factor analysis of the WISC–R at 11 age levels between 6½ and 16½ years. *Journal of Consulting and Clinical Psychology, 43,* 135–147.
Kaufman, A. S. (1979). *Intelligent testing with the WISC–R.* New York: Wiley.
Kaufman, A. S., & Doppelt, J. E. (1976). Analysis of WISC–R standardization data in terms of the stratification variables. *Child Development, 47,* 168–171.
Kavale, K. A., & Forness, S. A. (1984). A meta-analysis of the validity of Wechsler scale profiles and recategorizations: Patterns or parodies. *Learning Disability Quarterly, 7,* 136–156.
Kroonenberg, P. M., & ten Berge, J. M. F. (1987). Cross-validation of the WISC–R factorial structure using three-mode principal components analysis and perfect congruence analysis. *Applied Psychological Measurement, 11,* 195–210.
Kuiper, F., & Fisher, L. (1975). A Monte Carlo comparison of six clustering procedures. *Biometrics, 31,* 777–783.
Maccoby, E. E. (Ed.). (1966). *The development of differences.* Stanford, CA: Stanford University Press.
McDermott, P. A. (1980). Prevalence and constituency of behavioral disturbance taxonomies in the regular school population. *Journal of Abnormal Child Psychology, 8,* 323–336.
McDermott, P. A. (1988). Agreement among diagnosticians or observers: Its importance and determination. *Professional School Psychology, 3,* 225–240.
McDermott, P. A., & Watkins, M. W. (1987). *McDermott multidimensional assessment of children.* New York: Psychological Corporation.
Miller, R. G., Jr. (1966). *Simultaneous statistical inference.* New York: McGraw-Hill.
Milligan, G. W. (1980). An examination of the effect of six types of error perturbation of fifteen clustering algorithms. *Psychometrika, 45,* 325–342.
Mojena, R. (1977). Hierarchical grouping methods and stopping rules: An evaluation. *Computer Journal, 20,* 359–363.
Naglieri, J. A. (1980). WISC–R subtest patterns for learning disabled and mentally retarded children. *Perceptual and Motor Skills, 51,* 605–606.
Overall, J. E., & Klett, C. J. (1972). *Applied multivariate analysis.* New York: McGraw-Hill.
Rourke, B. P., & Strang, J. D. (1984). Subtypes of reading and arithmetical disabilities: A neuropsychological analysis. In M. Rutter (Ed.), *Developmental neuropsychiatry* (pp. 473–488). New York: Guilford Press.
Saccuzzo, D. P., & Lewandowski, D. G. (1976). The WISC as a diagnostic tool. *Journal of Clinical Psychology, 32,* 118–124.
Selz, M., & Reitan, R. M. (1979). Rules for neuropsychological diagnosis: Classification of brain function in older children. *Journal of Consulting and Clinical Psychology, 47,* 258-264.
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.
Thompson, R. J. (1980). The diagnostic utility of WISC–R measures with children referred to a developmental deviation center. *Journal of Consulting and Clinical Psychology, 48,* 440–447.
Thompson, R. J. (1981). The diagnostic utility of Bannatyne's recategorized WISC–R scores with children referred to a developmental evaluation center. *Psychology in the Schools, 18,* 43–47.
Tryon, R. C., & Bailey, D. E. (1970). *Cluster analysis.* New York: McGraw-Hill.
Vance, H. B., Fuller, G. N., & Ellis, R. (1983). Discriminant function analysis of LD/BD children scores on the WISC–R. *Journal of Clinical Psychology, 39,* 397–388.
Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *American Statistical Association Journal, 58,* 236–244.
Watkins, M. W., & McDermott, P. A. (1979). A computer program for measuring levels of overall and partial congruence among multiple observers on nominal scales. *Educational and Psychological Measurement, 39,* 235–239.
Wechsler, D. (1949). *Wechsler Intelligence Scale for Children.* New York: Psychological Corporation.
Wechsler, D. (1974). *Wechsler Intelligence Scale for Children–Revised.* New York: Psychological Corporation.
Wechsler, D., & Jaros, B. (1965). Schizophrenic patterns on the WISC. *Journal of Clinical Psychology, 21,* 288–291.
Wishart, D. (1982). *CLUSTAN users manual: Supplement* (3rd ed.). Edinburgh, Scotland: Program Library Unit, Edinburgh University.
Witkin, H. A., Dyk, R. B., Faterson, H. G., Goodenough, D. R., & Karp, S. A. (1962). *Psychological differentiation.* New York: Wiley.

Received January 9, 1989
Revision received May 16, 1989
Accepted May 30, 1989 ■