Journal of Psychoeducational Assessment
1990, *8*, 290-302

JUST SAY NO TO SUBTEST ANALYSIS: A CRITIQUE ON WECHSLER THEORY AND PRACTICE

Paul A. McDermott and John W. Fantuzzo
University of Pennsylvania

Joseph J. Glutting
University of Delaware

Wechsler's beliefs about the nature of human intelligence and its measurement have profoundly influenced contemporary theory and practice. He encouraged interpretations not only of more global intellective indices, such as IQ, but encouraged as well the search for pathognomonic meaning in patterns of underlying, more specific, subtest scores. This article examines the evidence that concerns the interpretation of Wechsler and similar tests as measures of specific rather than global ability. Popular practices that involve use of subtests for both *intra*individual and *inter*individual assessment are viewed in the light of empirical research, and recommendations are presented.

Wechsler emphasized two major themes for the interpretation of his tests. What is remarkable about this is that by his own assertions, and from the available evidence, only one of those themes can be true. The themes hold that the tests are, at once, measures of the global capacity called intelligence (as the Verbal and Performance IQs) *and* measures of myriad specific abilities that have their own integral meanings and are manifested through the independent and conjoint variation of subtests.

It is said, on the one hand, and we will draw upon Zachary's (1990) summary of Wechsler's position, that the purpose of the subtests is *not* to discover whether an individual is high or low on one or more subtests, but to facilitate inferences about the global capacity that presumably underlies *all* subtests. The subtests do not exist to measure any specific ability in isolation . . . but to assess something that will emerge from the individual's performance as a whole.

On the other hand, and turning to one of many contraposed statements by Wechsler, we read in the opening sections of the manual for the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) an argument for the relevance and pathognomonic significance of the highs and lows of single subtest variation and covariation. We are told explicity to be vigilant for such phenomena as being signs of mental deficiency, anxiety, or even organic brain disease. Wechsler noted that subtest patterns "add much to an examiner's diagnostic armamentarium" and urged that "every clinical psychologist needs to be aware of their ascribed significance" (1974, p. 7).

All of this illustrates what we call the theme of the one vs. the theme of the many. It is a dialectic about two reciprocally exclusive theoretical notions, but it is not

Paper presented at the National Conference on Intelligence: Theory and Practice, Memphis, May 1990.

Correspondence that concerns this article should be addressed to Paul A. McDermott, Graduate School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104.

our dialectic. It is one seminated by Wechsler, with two themes fundamentally dissonant and at variance, but played simultaneously by many, many test users and also by developers of tests that would emulate Wechsler.

Doubtless were we to say everything we might about the Wechsler scales and their contributions to research and practice, by far our comments would be quite positive. Space is limited, and so having expressed our deep respect for most of the Wechsler heritage, we are compelled to focus more intensively on the problems wrought by the dissonant themes of the one and the many.

Recently, a heated debate was covered in the *Communiqué* of the National Association of School Psychologists (1988). It concerned the merits of the IQ and related intelligence constructs. Essentially, the diagnostic and treatment utility of global intelligence measures was under attack with the funereal chant that IQ forever rest in peace. The attack was met by several distinguished psychologists. But to our dismay, few respondents offered denials or even lamentations over the death. Instead, most respondents dismissed the global intelligence construct as having mere historical value and, thereafter, touted the merits of viewing intelligence as a multi-differentiated construct whose only value lies in the extent to which it enables us to discern individual profiles of abilities—in other words, the dialectic of the one vs. the many, whereby the one is relegated to prescientific archives, and the many is deemed victorious and advanced as the standard of good science and practice.

We are not here to defend the concept of IQ. The literature is replete with reasons why one might have misgivings about IQ based on its frequent abuse. Our concern is that many are finding it easier to dismiss the IQ's import than to deal with those who persist in abusing it. In a near-desperate struggle to disassociate from malfeasance, many have espoused too quickly the notion that the true value of intelligence tests comes with the interpretive richness of tell-tale subtest profiles and studies of individuals' specific strengths and weaknesses.

Among the participants in the NASP debate, Reynolds (1988) seemed the only one to emphasize the vast body of empirical evidence to demonstrate the construct and prognostic utility of global intelligence. We cannot ignore a century of evidence that places global ability among the most dominant and enduring factors, both causal and corollary, associated with scholastic and occupational success; environmental adaptation; physical propensity and morbidity; and scientific, cultural and political acumen (Brody, 1985; Crano, Kenny, & Campbell, 1972; Eysenck & Barrett, 1985; Jencks, 1972; Jensen, 1982; Mackenzie, 1984; Terman & Oden, 1959). Moreover, for tests such as the WISC-R, research has demonstrated that global IQ measures provide construct continuity across gender and cultural majority and minority groups (Gutkin & Reynolds, 1981; Reynolds & Gutkin, 1980b) and provide equitable accuracy and no substantive bias in forecasting relevant performance criteria for females and larger minority groups (Hale, 1980; Reynolds & Gutkin, 1980a; Reynolds, Gutkin, Dappen, & Wright, 1979).

However, with recognition of the practical and political limitations of global IQ measures, popular practice has given way to belief in the superiority of specific over general ability constructs. We are encouraged to search for true individual differences and even treatment utility in speculation about the meaning of subtest variation. Indeed, the art of heaping inferences upon piles of other inferences about specific abilities has become a fashionable pastime.

Thus, we are compelled to examine the evidence that would favor interpretation of the Wechsler and similar tests as measures of differentiated rather than global ability. Within this context, we will examine first the popular practice of using Wechsler subtests for *intra*individual assessment, and, second, their use for *inter*individual assessment.

INTRAINDIVIDUAL ASSESSMENT

Clearly, the most popular approach to discovery of differential abilities within a given child or adult is the method of ipsative assessment. As would follow for a typical case on the WISC-R (illustrated in Table 1), the average scaled score across subtests is subtracted from each component subtest score. The resulting *ipsative* scores show both positive and negative deviations from the child's own average ability and reveal areas of relative strength (large positive deviations) and weakness (large negative deviations). Indeed, the ipsative perspective holds a certain intuitive appeal because, by removing the average ability element from a person's scores, the resulting profile seems to isolate and amplify aspects of differential ability.

Table 1
WISC-R Subtest Scaled Scores and Ipsative Scores for a Given Child

Score type	IN	SM	AR	VO	CM	DS	PC	PA	BD	OA	CD	Sum
Scaled	8	12	14	13	15	17	13	14	6	10	12	132[a]
Ipsative	−4	0	+2	+1	+1	+5	+1	+2	−6	−2	0	0

Note.—IN = Information; SM = Similarities; AR = Arithmetic; VO = Vocabulary; CM = Comprehension; DS = Digit Span; PC = Picture Completion; PA = Picture Arrangement; BD = Block Design; OA = Object Assembly; CD = Coding.
[a]Average scaled score across 11 subtests = 132/11 = 12.

Advanced as a rich alternative to the global ability concept, ipsative assessment is now popularized in leading assessment texts and extended to other intelligence tests (Delaney & Hopkins, 1987; Kaufman, 1979, 1990; Kaufman & Kaufman, 1983; Sattler, 1988) and is now performed through commercial microcomputer software (Ingram, 1985; Psychological Corporation, 1986a, 1986b). Notwithstanding appreciable popularity, ipsative assessment has not been well researched, and there is ample evidence to militate against its current applications.

Ipsative measures are evaluated against the same standards applied for other units of measurement; namely, their reliability, validity, and actual utility. Accordingly, McDermott, Fantuzzo, Glutting, Watkins, and Baggaley (1990) evaluated ipsatized subtest scores in terms of their construct validity, predictive criterion validity, temporal stability, and utility in psychological practice and measurement theory.

It has been assumed widely by proponents of ipsative assessment that the targeted ability constructs remain the same after as before ipsatization. Thus, if by some salient theory one is willing to surmise that a given subtest or group of subtests reflects a specific ability, one trusts that upon transformation to an ipsative form, the subtest or group of subtests continues to measure the same thing. However,

as predicted under the earlier theorem by Clemens (1965), McDermott et al. (1990) established with the WISC-R standardization sample that the underlying phenomena are vastly inconstant before and after ipsatization. Whereas with all major omnibus intelligence tests the subcomponents of the WISC-R correlate positively and substantially prior to ipsatization (grand average intercorrelation of subtests = .42, with all correlations positive), the relationships subsequent to ipsatization deteriorate markedly in strength and uniform directionality (unsigned grand average intercorrelation = .15 with 76% of the correlations negative). This follows from the fact that ipsatization automatically removes from a person's scores all common variance, as associated with Spearman's *g*. By necessity and design across all major intelligence tests, the common variance is the largest, most integral, and robust portion of score variability. Thus, for the WISC-R, ipsatization means the loss of nearly 60% of the test's reliable variance.

The worth of the remaining variance that defines ipsative scores is assessed most directly through their ability to predict accurately to some meaningful criteria. For school-age children, academic success is among the most relevant such criteria. Consequently, McDermott et al. (1990) compared conventional norm-based with ipsatized subtest scores for predicting a variety of standardized achievement criteria across a wide range of samples. On the average, it was found that, whereas conventional subtest scores were able to account for 38.1% of academic performance, ipsative subtest scores accounted for only 5.8% to 9.0%. This represented a two-thirds to three-quarters drop in predictive efficiency for ipsative measures. To the extent that the ultimate purpose of any psychological measure rests on its ability to improve prediction, the observation that ipsative scores fail to exceed, match, or even approach conventional scores in predictive efficiency effectively vitiates any claim that ipsative assessment has relative merit.

In a further exploration of the reasons for the reduced validity of ipsative measures, the investigators postulated decrements in score stability associated with the removal of common variance. They showed for WISC-R scores across 1-month and 3-year intervals that ipsative scores were significantly less reliable than norm-based scores and that only for two subtests, which happened in their normative form to reflect less global intellective ability, did stability not falter.

Finally, McDermott and his colleagues examined the presumed diagnostic and treatment utility of ipsative approaches. Cattell (1944), who introduced the term "ipsative," warned that such measures had limited relevance. Thus, as McDermott et al. (1990) noted, once ability scores have been ipsatized, they can have no application for comparison across or among individuals. For instance, given the ipsative scores for two individuals on the same subtest (e.g., Arithmetic), whereby one individual has a rather high score (e.g., + 8) and the other a low score (– 2), there is no way to tell which individual has more Arithmetic ability. Moreover, because ipsativity alters every person's scores by a different value (their personal average), ipsative scores mark a true change of metric. Rather than remaining parametric across individuals, ipsative scores are nonparametric and polytonic. This condition effectively precludes application of all those statistical procedures that require parametric measures, and, unfortunately, this includes nearly all of the more sophisticated algorithms for testing group differences, defining latent dimensions, and multivariate classification.

Apparent also is what investigators called the *see-saw effect* in treatment applications. Because deviational ipsative scores must always sum to zero (refer to Table 1), one learns that as performance in one ipsative area is improved, in another it must tautologically deteriorate. Thus, before one devotes oneself to a remedial program that might enhance ablity in one area (as manifest through an ipsative score elevation), one must realize that the natural economy of ipsativity dictates that ability in other areas will diminish by an equal amount. It is a perfect zero-sum enterprise that can never be beneficial in the absolute sense.

Before we turn to the topic of interindividual assessment, we should touch upon one other aspect of the presumption that many measures of specific human abilities have more treatment validity than one measure of global ability. We are reminded of the landmark reviews by Cronbach and Snow (1977) of the many years of "aptitude-treatment" interaction research. They discovered that, across the literature and with few exceptions, more differentiated and *specific* views of intellectual abilities were *not* the most useful bases for individualizing intruction and curriculum. On the contrary, attention to global measures of ability reaped noticeably better results.

INTERINDIVIDUAL ASSESSMENT

Among users of Wechsler scales, there is perhaps no practice more popular than interpretation of subtest patterns. Based primarily on inductive theory and speculative task analysis about skills required for succcess on given subtests, interpretations have been offered for more than 75 different patterns of subtest variation (Bannatyne, 1974; Glasser & Zimmerman, 1967; Guilford, 1967; Kaufman, 1979; Kaufman, Harrison, & Ittenbach, 1990; Matarazzo, 1972; Saccuzzo & Lewandowski, 1976; Selz & Reitan, 1979; Wechsler & Jaros, 1965; Winne & Gittinger, 1973; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962; Zimmerman & Woo-Sam, 1973). With publication of the WISC-R and Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981), 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, for example, some writers reported discovery of subtest patterns unique to groups of emotionally disturbed subjects (Dean, 1977; Morris, Evans, & Pearson, 1978; Paget, 1982) or learning impaired subjects (Holcomb, Hardeste, Adams, & Ponder, 1987; McCue, Shelly, Goldstein, & Katz-Garris, 1984; Schiff, Kaufman, & Kaufman, 1981; Smith, 1978; Smith, Coleman, Dokecki, & Davis, 1977; Strichart & Love, 1979; Tabachnick, 1979; Vance, Gaynor, & Coleman, 1976; Zingale & Smith, 1978). Other investigators (e.g., Leton & Myamoto, 1987; Rourke & Strang, 1984; Wallbrown, Vance, & Blaha, 1979), who regard collections of disabled learners as rather heterogeneous, have applied objective profile sorting techniques to identify apparently unique subtypes of disability. Unfortunately, none of the aforementioned studies provided experimental control or contrast groups that might have supported claims for distinctive WISC-R or WAIS-R profiles.

Direct comparisons of groups of diversely diagnosed students have led some investigators to conclude that subtest profiles are helpful in differentiating between the emotionally and learning impaired (Dean, 1978, 1980; Fuller & Goh, 1981; Vance, Fuller, & Ellis, 1983), the learning impaired and mentally deficient (Naglieri,

1980), and classes of delinquents who are incarcerated rather than probationary (Hubble & Groff, 1980).

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 Wechsler profiles adds nothing to that afforded by global IQ measures.

In preparation for a series of related studies, McDermott and his colleagues identified among the 2,100 published works on the WISC-R and WAIS-R some 70 studies that pertain to profile analysis (McDermott, Glutting, Jones, Watkins, & Kush, 1989; McDermott, Glutting, Jones, & Noonan, 1989). They observed several serious and pervasive methodological problems throughout those studies, which essentially operated to negate or equivocate their import. We will enumerate those problems briefly.

First, in the existing literature it generally is assumed that groups of similarly diagnosed subjects represent meaningful, if not homogeneous categories. Yet, it is clear from the research on diagnostic congruence in clinical and school psychology and special education (Achenbach, 1988; Garfield, 1978; McDermott, 1981, 1988) that no such assumption is warranted.

Second, available research fails to preclude circular use of Wechsler subtest profiles for *both* initial formation of diagnostic groups *and* subsequent searches for profiles that might inherently (or independently) define those groups.

Third, much of the published research, by pooling convenient samples that span broad age ranges, assumes reasonably equitable measurement error for subtests across those age levels. The assumption generally is not justified. For example, with regard to the WISC-R, it has been shown that precision of measurement for the subtests tends to vary, sometimes dramatically, from childhood through adolescence (Conger, Conger, Farrell, & Ward, 1979).

Fourth, throughout the Wechsler profile literature, hypothesis testing (when actually carried out) is based on linear modeling such as univariate or multivariate tests of group mean differences, correlation, and multiple regression or discriminant 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). Thus, as measurement constructs, profiles are quite unlike simple test scores or score composites that can be represented adequately through their additivity or linear relationship. To illustrate, one will note that information about a profile's relative level is actually a nomothetic (norm-referenced) comparison of its aggregate value relative to other profiles in a given population, while a profile's shape constitutes a nomothetic comparison of the similarity of a pattern of greater and lesser abilities to prototypic patterns in that same population. Moreoever, it can be shown that the interpretive and mathematical weight of a deviant score (either elevation or depression) for a given subtest exceeds that for an invariant score on the same subtest. Inasmuch as score values for the same subtest can take on greater or lesser weight depending upon the configuration of other subtest scores, profiles retain nonlinear as well as linear elements. (See Cattell, 1949; Hore, 1941.) Consequently, questions that address profile uniqueness or dissimilarity (as in hypothesis testing) cannot be resolved

through conventional statistical procedures to assess group mean differences (as in ANOVA or MANOVA) or that rely on linear combinations of subtest scores (as in multiple regression, discriminant, or canonical analysis). As noted by Mosel and Roberts (1954), inquiries about profiles are matters of "configural" rather than linear variation and so require configural hypotheses with statistical treatments sensitive to trends in both level and shape.

Fifth, and most important, claims for discovery of unique Wechsler profiles are never made against a viable null hypothesis; namely, that the profile believed to be unique is really commonplace in the population of normal children or adults and, thus, unremarkable. Instead, claims for profile uniqueness are grounded in surmises that either (a) the *average* profile for a group of similarly diagnosed subjects is inherently characteristic of the diagnostic category *and* uncharacteristic of alternative categories; or (b) evidence for *differences* in average profiles *between* diagnostic categories is tantamount to proof that such profiles are unlikely to emerge in the overall general population. Frankly, without clear knowledge of the types and prevalence of subtest profiles that exist in the population of normal individuals, we simply cannot know whether profiles elsewhere discovered are uncommon, distinctive, or clinically meaningful.

For this reason, normative typologies of *core profiles* recently have been established and validated using standarization samples for the WISC-R (McDermott, Glutting, Jones, Watkins, & Kush, 1989), WAIS-R (McDermott, Glutting, Jones, & Noonan, 1989), and Wechsler Preschool and Primary Scale of Intellingence (Glutting & McDermott, in press a), as well as for the Kaufman Assessment Battery for Children

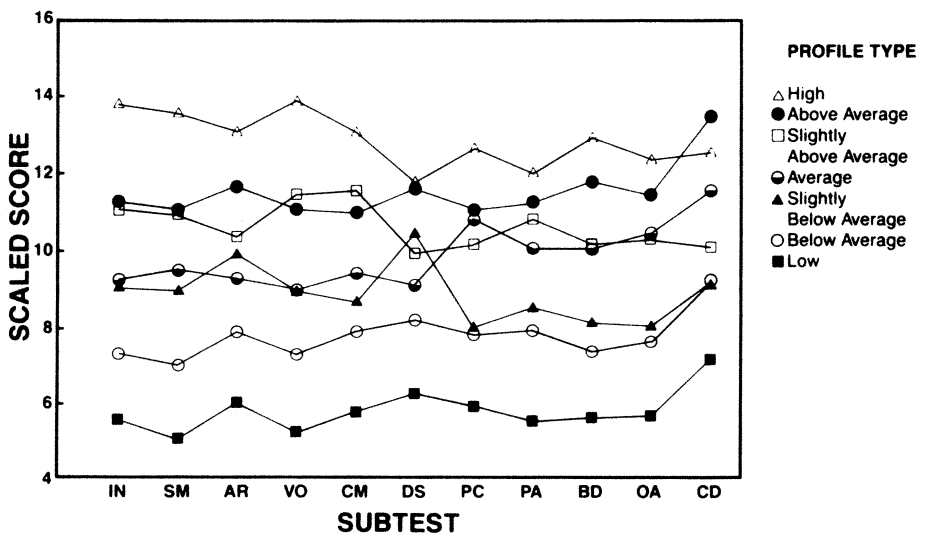

FIGURE 1. Core profile types in the WISC-R standardization sample. IN = Information; SM = Similarities; AR = Arithmetic; VO = Vocabulary; CM = Comprehension; DS = Digit Span; PC = Picture Completion; PA = Picture Arrangement; BD = Block Design; OA = Object Assembly; CD = Coding.

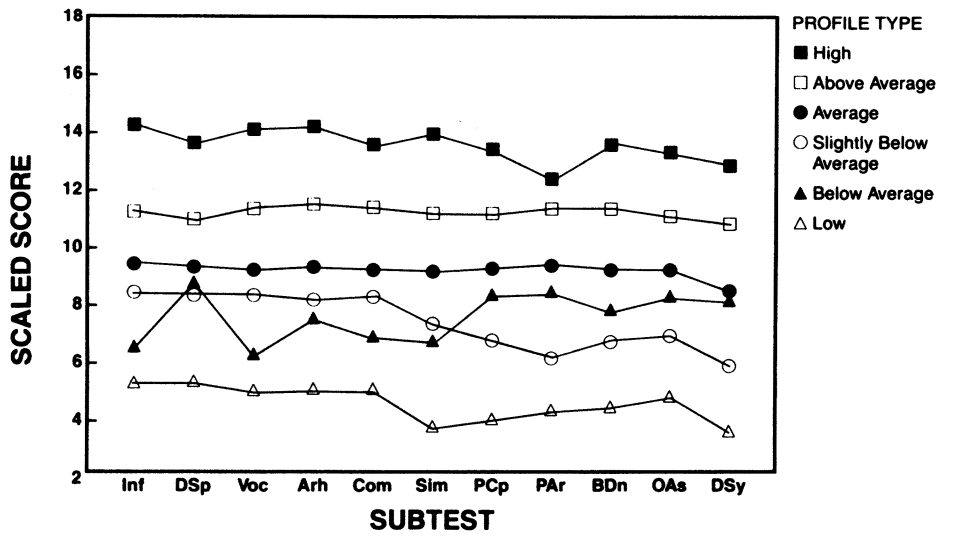

FIGURE 2. Core profile types in the WAIS-R standardization sample defined primarily by level of general intellectual functioning. Inf = Information; DSp = Digit Span; Voc = Vocabulary; Arh = Arithmetic; Com = Comprehension; Sim = Similarities; PCp = Picture Completion; PA = Picture Arrangement; BDn = Block Design; OAs = Object Assembly; DSy = Digit Symbol.

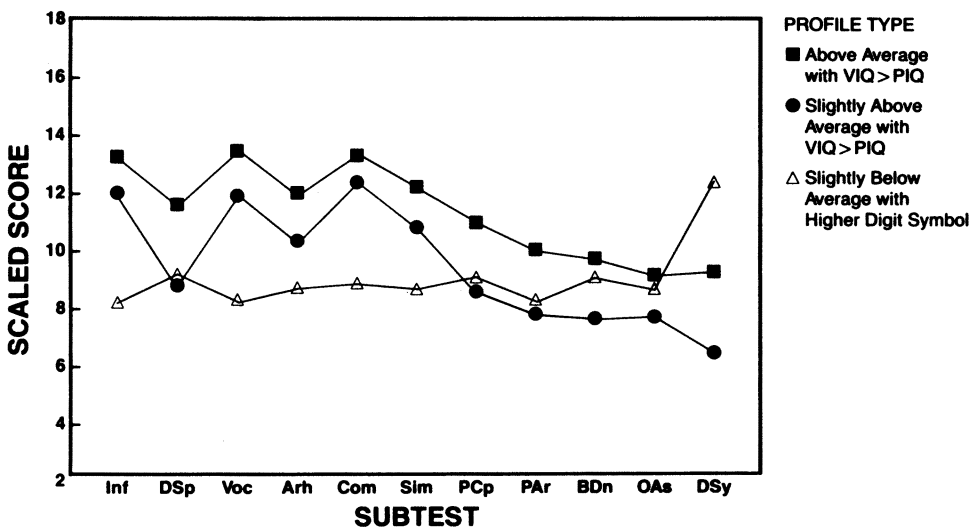

FIGURE 3. Core profile types in the WAIS-R standardization sample defined primarily by Verbal/Performance IQ discrepancies. Inf = Information; DSp = Digit Span; Voc = Vocabulary; Arh = Arithmetic; Com = Comprehension; Sim = Similarities; PCp = Picture Completion; PA = Picture Arrangement; BDn = Block Design; OAs = Object Assembly; DSy = Digit Symbol.

(Glutting, McGrath, & Kamphaus, 1990) and McCarthy Scales of Children's Abilities (Glutting & McDermott, in press b). Each study of core profiles is based on hierarchical, minimum-variance clustering, with multiple replications across age levels or random segments of the standardization samples. Furthermore, core profile types are validated against internal and external statistical and heuristic criteria and are described in terms of significant trends in prevalence by age, gender, race, educational and occupational level, marital status, birthplace, residence, birth order, number of siblings, handedness and abnormal Verbal/Performance IQ discrepancies. Figure 1 shows core profile types that exist in the WISC-R national sample, and Figures 2 and 3 show core profiles from the WAIS-R sample.

Also provided for each normative typology are straightforward methods for determining the relative uniqueness of any subtest profile believed to carry special meaning. One method enables researchers to define exactly through the group similarity coefficient for profiles with correlated components (Tatsuoka, 1974; Tatsuoka & Lohns, 1988) the level and shape similarity of any profile to the core profile types in the general population. Another procedure, based on generalized distance theory, takes only moments and is suggested for psychologists in everyday practice. Whenever the hypothesis testing finds a profile appreciably similar to one of the core types, the individual's profile must be considered commonplace and undistinctive.

CONCLUSION

We have said that for the Wechsler scales the one-many controversy was spawned by Wechsler himself. This was evident as early as 1958 in the fourth edition of *The measurement and appraisal of adult intelligence.* Wechsler advanced the idea that the subtests had special and pathognomonic meaning, while at the same time he argued that the subtests had no particular role other than as a means to the goal of contributing to the composite called intelligence. In fact, Zachary (1990) reminds us that the very selection of subtests was guided not by a comprehensive theory that would attempt to measure all of the vital ingredients that comprise intelligent behavior, but, rather, by the need to devise subtests that were reliable, brief, and easy to administer. Within this framework, the one and the many are thus made antithetical. Of course, apart from this observation, we do not mean to assert that intelligence tests cannot conceivably measure both global *and* specific abiilities. We do wish to note, however, that efforts thus far to identify specific abilities that approach the construct and especially predictive validity of global intelligence have been largely unsuccessful (e.g., Brody & Brody, 1976; Cronbach & Snow, 1977; Eysenck, 1979). As Brody (1985) has emphasized:

> Any theory of the structure of the intellect that proposes to replace [global ability] with a more differentiated concept ought to have predictive validity that exceeds that obtained from a general factor (p. 357).

Science always presents a certain tension between the global and specific—the global being the more generalizable and the specific the more accurate. However, the tension is resolved easily in that science advances only by its ability to offer what is most generalizable yet reasonably accurate. Only when the accuracy of the very specific exceeds that for the global is there mitigation for preferring specific

over global constructs. Based on our current knowledge about the realm of human intelligence, there is little to support the belief that many intelligence constructs are better than one. Until preponderant and convincing evidence shows otherwise, we are compelled to advise that psychologists just say "no" to subtest analysis.

REFERENCES

Achenbach, T. M. (1988). Integrating assessment and taxonomy. In M. Rutter, A. H. Tuma, & I. S. Lamn (Eds.), *Assessment and diagnosis in child psychopathology* (pp. 300-343). New York: Guilford Press.

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

Brody, N. (1985). The validity of tests of intelligence. In B. Wolman (Ed.), *Handbook of intelligence* (pp. 353-389). New York: John Wiley.

Brody, E. B., & Brody, N. (1976). *Nature, determinants, and consequences.* New York: Academic Press.

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

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

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

Crano, W. D., Kenny, J., & Cambpell, D. T. (1972). Does intelligence cause achievement?: A cross-lagged panel analysis. *Journal of Educational Psychology, 63,* 258-275.

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

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods.* New York: Irvington.

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

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.

Delaney, E. A., & Hopkins, T. F. (1987). *The Stanford-Binet Intelligence Scale: Fourth edition examiner's handbook.* Chicago: Riverside.

Eysenck, H. J. (1979). *The structure and measurement of intelligence.* Berlin: Springer Verlag.

Eysenck, H. J., & Barrett, P. (1985). Psychophysiology and the measurement of intelligence. In C. R. Reynolds & V. L. Wilson (Eds.), *Methodological and statistical advances in the study of individual differences* (pp. 1-49). New York: Plenum Press.

Fuller, G. B., & Goh, D. S. (1981). Intelligence, achievement, and visual-motor performance among learning disabled and emotionally impaired children. *Psychology in the Schools, 18,* 261-268.

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.

Glutting, J. J., & McDermott, P. A. (in press a). Patterns and prevalence of core profile types in the WPPSI standardization sample. *School Psychology Review.*

Glutting, J. J., & McDermott, P. A. (in press b). Score structures and applications of core profile types in the McCarthy scales standardization sample. *Journal of Special Education.*

Glutting J. J., McGrath, E. A., & Kamphaus, R. W. (1990). *Taxonomy and validity of core profile types in the K-ABC standardization sample.* Unpublished manuscript,

University of Delaware, Department of Educational Studies, Newark, DE.

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

Gutkin, T. B., & Reynolds, C. R. (1981). Factorial similarity of the WISC-R for white and black children from the standardization sample. *Journal of Educational Psychology, 72,* 227-231.

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. (1980). Sexual bias in the WISC-R. *Journal of Consulting and Clinical Psychology, 48,* 776.

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.

Holcomb, R. A., Hardeste, R. A., Adams, N. A., & Ponder, H. M. (1987). WISC-R types of learning disabilities—a profile analysis with cross validation. *Journal of Learning Disabilities, 20,* 369-373.

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

Hubble, L. M., & Groff, M. (1980). WISC-R profiles of adjudicated delinquents later incarcerated or released on probation. *Psychological Reports, 47,* 481-482.

Ingram, R. (1985). *The Kaufman method of WISC-R hypothesis generation* [Computer program]. New York: John Wiley.

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

Jensen, A. R. (1982). The chronometry of intelligence. In R. J. Sternberg (Ed.), *Recent advances in research on intelligence.* Hillsdale, NJ: Erlbaum.

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

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence.* Boston: Allyn & Bacon.

Kaufman, A. S., Harrison, P. L., & Ittenbach, R. F. (1990). Intelligence testing in the schools. In T. B. Gutkin & C. R. Reynolds (Eds.), *Handbook of school psychology* (2nd ed., pp. 289-327). New York: John Wiley.

Kaufman, A. S., & Kaufman, N. L. (1983). *Interpretive manual for the Kaufman Assessment Battery for Children.* Circle Pines, MN: American Guidance Service.

Leton, D. A., & Myamoto, L. K. (1987). Psychometric classifications of learning disabled students. *Psychology in the Schools, 24,* 201-209.

MacKenzie, B. (1984). Explaining race differences in IQ: The logic, the methodology, and the evidence. *American Psychologist, 37,* 1214-1233.

Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore: Williams & Wilkins.

McCue, M., Shelly, C., Goldstein, G., & Katz-Garris, L. (1984). Neuropsychological aspects of learning disability in young adults. *International Journal of Neuropsychology, 6,* 229-233.

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

McDermott, P. A. (1988). Agreement among diagnosticians or observers: Its importance and determination. *Professional School Psychology, 3,* 225-240.

McDermott, P. A., Fantuzzo, J. W., Glutting, J. J., Watkins, M. W., & Baggaley, A. R. (1990). *Illusions of meaning in the ipsative assessment of human ability.* Unpublished manuscript, University of Pennsylvania.

McDermott, P. A., Glutting, J. J., Jones, J. N., & Noonan, J. V. (1989). Typology and prevailing composition of core profiles

in the WAIS-R standardization sample. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 118-125.

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

Morris, J. D., Evans, J. G., & Pearson, D. R. (1978). The WISC-R subtest profile of a sample of severely emotionally disturbed children. *Psychological Reports, 42,* 319-325.

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

Naglieri, J. A. (1980). WISC-R subtest patterns for learning disabled and mentally retarded children. *Perceptual and Motor Skills, 51,* 605-606.

National Association of School Psychologists. (1988). Debate over usefulness of IQ. *Communique, 17,* 4-7.

Paget, K. D. (1982). Intellectual patterns of conduct problem children on the WISC-R. *Psychology in the Schools, 19,* 439-435.

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

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

Reynolds, C. R. (1988). Of straw men and the practice of the ancients: Comment. *Communique, 17,* 6.

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.

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

Schiff, M. M., Kaufman, A. S., & Kaufman, N. L. (1981). Scatter analysis of WISC-R profiles for learning disabled children with superior intelligence. *Journal of Learning Disabilities, 14,* 400-404.

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.

Smith, M. D. (1978). Stability of WISC-R subtest profiles for learning disabled children. *Psychology in the Schools, 15,* 4-7.

Smith, M. D., Coleman, J. M., Dokecki, P. R., & Davis, E. E. (1977). Intellectual characteristics of school labeled learning disabled children. *Exceptional Children, 43,* 352-357.

Strichart, S. S., & Love, E. (1979). WISC-R performance of children referred to a university center for learning disabilities. *Psychology in the Schools, 16,* 183-188.

Tabachnick, B. G. (1979). Test scatter on the WISC-R. *Journal of Learning Disabilities, 12,* 60-62.

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.

Terman, L. M., & Oden, M. H. (1959). *The gifted group as adults.* Stanford, CA: Stanford University Press.

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.

Vance, H. B., Fuller, G. N., & Ellis, R. (1983). Discriminant function analysis of LD/BD children's scores on the WISC-R. *Journal of Clinical Psychology, 39,* 397-398.

Vance, H. B., Gaynor, P., & Coleman, M. (1976). Analysis of cognitive abilities for learning disabled children. *Psychology in the Schools, 13,* 477-483.

Wallbrown, F. H., Vance, H., & Blaha, J. (1979). Developing remedial hypotheses from ability profiles. *Journal of Learning Disabilities, 12,* 59-63.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams & Wilkins.

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

Winne, J., & Gittinger, J. (1973). Introduction to the Personality Assessment System. *Journal of Clinical Psychology: Monograph Supplement* (No. 38).

Witkin, H. A., Dyk, R. B., Faterson, H. G., Goodenough, D. R., & Karp, S. A. (1962). *Psychological differentiation.* New York: John Wiley.

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

Zimmerman, I. L., & Woo-Sam, J. M. (1973). *Clinical interpretation of the Wechsler Adult Intelligence Scale.* New York: Grune & Stratton.

Zingale, S. A., & Smith, M. D. (1978). WISC-R patterns for learning disabled children at three SES levels. *Psychology in the Schools, 15,* 199-204.