

What Can the WISC-R Measure?

**Anthony J. Conger, Judith Cohen Conger,
Albert D. Farrell, and David Ward
Purdue University**

The WISC-R was investigated by using measures of profile (multivariate) reliability in order to determine its most reliable dimensions and the precision and similarity of the multivariate structure across age groups. Due to differences among the 11 age groups in both subscale reliabilities and true score covariance matrices, it was concluded that the precision of measurement differed across age groups. This finding was further supported by a comparison of canonical reliability coefficients and composites computed for each age group. However, exhaustive analyses of Varimax rotated profile dimensions indicated that the structure of the WISC-R subscales is rather stable across age groups, but the reliability of that structure differs systematically. A synthesis of the analyses indicated that (1) the WISC-R allows highly reliable comparisons of profile levels (Full-Scale IQ) at each age level; that (2) reasonably reliable comparisons of Verbal-Performance differences can be made at each age level; but that (3) for other comparisons, caution should be exercised because of age group differences and potentially high unreliability. Two strategies for the interpretation of WISC-R profiles, which take into account the above findings, are offered.

Since the advent of the original Wechsler scales (Wechsler, 1949), psychologists have attempted to derive diagnostic significance from various combinations and permutations of differences among subscale scores. The division of

the test into separate subscales has undoubtedly contributed to this endeavor. Despite the search for meaningful interpretations based on either individual subscale scores or comparisons among subscales, little of worth has been empirically demonstrated. Littell (1960) equated prediction from a single subscale with little more than a rationalized hunch; and Zimmerman and Woo-Sam (1972) in their review of the Wechsler Intelligence Scale for Children (WISC) from 1960-1970, concluded that while some results have yielded suggestive clues, specific subtest patterns have really not been established. Despite caveats about questionable interpretations of subscale differences, conservative alternatives have been ignored (Anastasi, 1968; McNemar, 1957; Tellegen & Briggs, 1967), and clinicians have continued to use the Wechsler scales in various and sundry diagnostic pursuits. Recent examples include the work of Robb, Bernardoni, and Johnson (1972) and Sattler (1974), who make liberal use of subscale comparisons in their quest for "psychodiagnosticity" based on the Wechsler scales.

With the advent of the Wechsler Intelligence Scale for Children—Revised (WISC-R; Wechsler, 1974), practitioners will undoubtedly continue to engage in interpretative procedures based on subscale comparisons. Indeed, there is an already growing literature suggesting new ap-

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 3, No. 4 Fall 1979 pp. 421-436
© Copyright 1979 West Publishing Co.

proaches to interpreting scatter on the WISC-R (Kaufman, 1976a) as well as suggestions for the interpretation of Verbal-Performance or other subscale discrepancies (Kaufman, 1976b; Piotrowski & Grubb, 1976).

A basic problem in interpreting any Wechsler composite or profile, raised by Tellegen and Briggs in 1967, is that the reliability of the profile is simply not taken into account. While techniques have been available for establishing the reliability of composites (cf. Tellegen & Briggs, 1967), it is only recently that multivariate techniques have been developed to determine profile reliability. Conger and his colleagues (Conger, 1974; Conger & Lipshitz, 1973; Conger & Stalard, 1976) have developed a general measure of multivariate profile reliability that can be used to establish the reliability of all possible subscale comparisons as well as to establish a set of independent maximally reliable profile composites. This technique is similar to the approach discussed by Bock (1966), and its application to the WISC and similar instruments has been recommended by Cronbach, Gleser, Nanda, and Rajaratnam (1972).

Conger and Conger (1975) calculated the multivariate profile reliability of the WISC subscales for three age groups. Results showed that while all possible differences among subscales were relatively unreliable ($r = .51$), a subset of five uncorrelated dimensions could be established with a more satisfactory average reliability of .87. However, they suggested that only two dimensions really have unambiguous and readily interpretable patterns (total IQ and Verbal-Performance difference) and cautioned clinicians to restrict themselves to these until adequate reliability and validity could be established for other dimensions.

Although factor analyses of the WISC-R have been done (Kaufman, 1975; Wallbrown, Blaha, Wallbrown, & Engin, 1975), the results of those analyses do not address the issue of what a battery of subscales can reliably measure. This follows from the mathematics of common factor analysis, which not only eliminates unreliable

variance but also eliminates reliable specific scale variance. In the extreme case of uncorrelated scales, there would be no common factors, yet differences among subscales would be maximally interpretable. Because of this inappropriateness of factor analysis and because of purported substantial changes and improvements in the WISC-R compared to the WISC (Kaufman, 1975; Wechsler, 1974), it seems prudent to investigate what the WISC-R subscales can reliably measure using the methods discussed by Conger and Conger (1975) in their analysis of the WISC. These methods are particularly appropriate for establishing (1) the reliability of all possible subscale comparisons; (2) a set of uncorrelated maximally reliable profile dimensions (or composites); and (3) a subset of profile dimensions rotated to provide an interpretable structure.

An additional issue addressed is whether the WISC-R is measuring the same things at different ages and, if it is, whether it is measuring with equivalent precision (i.e., as measured by reliability or standard error of measurement). Along these lines, Kaufman (1975) has concluded, on the basis of clinical comparisons of verbally interpreted factor structures, that the WISC-R is essentially constant across ages. Even if Kaufman is correct about the relative similarity of the structure of the common factors, questions will remain about the similarity of what can be reliably measured and how well similar dimensions are measured at different ages.

Data

Data were taken from the WISC-R manual (Wechsler, 1974). Wechsler provided correlation matrices among subscales for 200 children at each of 11 age levels (6½ through 16½). Split-half reliability estimates for these samples were provided for each age group for 10 of the 12 subscales. (Split-half estimates were not available for Digit Span and Coding.) Wechsler also provided test-retest reliability estimates for all sub-

Table 1
Analysis of Scale Reliabilities Across Age Groups

Scale	(10)	Signif- icance Level	Correlation With Age	Range of Reliabilities
Information	(I) 75.04	.001	.86	.67-.90
Similarities	(S) 22.56	.05	-.44	.74-.87
Arithmetic	(A) 16.88	n.s.	.14	.69-.81
Vocabulary	(V) 98.76	.001	.84	.70-.92
Comprehension	(C) 48.87	.001	.51	.69-.87
Picture Completion	(PC) 40.71	.001	-.70	.68-.84
Picture Arrangement	(PA) 7.46	n.s.	-.19	.69-.78
Block Design	(BD) 27.36	.01	.66	.80-.90
Object Assembly	(OA) 11.90	n.s.	.18	.63-.76
Mazes	(MZ) 53.84	.001	-.83	.57-.82

scales; however, these were provided for only three pooled age groups and were based on subsamples of 100 children for each pooled age group. Consequently, it was decided to use the split-half reliability estimates and to restrict the analyses to the 10 subscales for which these were provided.

Analyses and Results

Tests of Age Group Differences

Age-related differences in the WISC-R were investigated by comparing subscale reliability estimates, observed score correlation matrices, and true score correlation matrices across the 11 age groups. These analyses were done in order to test for age group differences in the structure and precision of measurement and also to determine whether age groups might be pooled to minimize the number of profile reliability analyses.

Subscale reliabilities were compared by χ^2 tests on Fisher log transformed values (cf. Hays, 1973, p. 664). The results (Table 1) indicated that reliabilities differed across ages for 7 of the 10 subscales. Correlations of reliabilities with

age indicated that reliabilities of 4 subscales (Information, Vocabulary, Comprehension, and Block Design) increased with age; and reliabilities of three subscales (Similarities, Picture Completion, and Mazes) decreased with age (Table 1).

Differences in the structure of measurement among age groups were investigated by comparing covariance¹ matrices of observed and true scores across age groups by Bartlett's χ^2 test of homogeneity of covariance matrices (cf. Timm, 1975, pp. 251-252). No significant difference (χ^2 (550) = 591.84, $p > .08$) was found among age groups for observed score covariances (comparing correlation matrices with unities in the diagonal); however, a highly significant difference (χ^2 (550) = 2237.18, $p < .001$) was found among age groups for true score co-

¹The term "covariance" is used to describe the observed score correlation matrix with unities and reliabilities in the diagonal in order to maintain continuity between them. In this regard, the observed score correlation matrix with unities in the diagonal represents covariances among standardized observed scores, while the same matrix with reliabilities in the diagonal represents covariances among true scores whose variances are equal to reliabilities.

variances (comparing correlation matrices with reliabilities in the diagonal).

In addition, tests of homogeneity of covariance matrices were done within each of three adjacent age groups (6½ to 8½, 10½ to 12½, and 14½ to 16½) to determine whether some pooling of these groups might be possible. The groups are comparable in age range to those studied by Conger and Conger (1975) in their analysis of the WISC. The results indicated, however, that the true score covariance matrices were not sufficiently similar (all *p*'s < .001) to warrant pooling. Based on the above results, it was concluded that the structure of reliable information in the WISC-R differed across ages. Furthermore, the differences among age groups were manifest in the subscale reliabilities and true score covariance matrices but not in the intercorrelations among observed scores.

Profile Reliability and Maximally Reliable Composites

Because of the differences among age groups in the subscale reliabilities and true score covariance matrices, canonical reliability coefficients and composites were computed for each age group using the procedures described by Conger and Conger (1975). An index of overall profile reliability was computed for each age group by the formula

$$\hat{\rho} = \frac{1}{K} \text{Trace} (R^{*tt} R_{xx}^{-1}) \quad [1]$$

where

- K* is the number of subscales,
- R_{xx}⁻¹* is the inverse of the correlation matrix with unities in the diagonal, and
- R^{*tt}* is the correlation matrix with reliabilities in the diagonal

This index takes the intercorrelations among subscales into account and can be interpreted as a measure of the reliability of a randomly selected composite or the average reliability of all possible composites (Conger & Lipshitz, 1973).

The maximally reliable profile composites are a set of mutually orthogonal composites with sequentially decreasing reliabilities. These were found (for each age group) by solving the eigenvalue-eigenvector equation

$$R^{*tt} \underline{V}_j = \gamma_j R_{xx} \underline{V}_j \quad [2]$$

where *V_j* is a vector of weights to be applied to the scales to obtain the *j*th maximally reliable profile composite and *γ_j* is the reliability of the *j*th composite. That is, the weight vector found from Equation 2 could be used to form a derived score *Y_{ij}* for each individual *i* as follows:

$$Y_{ij} = \sum_{k=1}^K X_{ik} V_{kj} \quad [3]$$

where, in the case of the WISC-R, *X_{ik}* would be the standard score for individual *i* on scale *k*. The composite score *Y_{ij}* would then have a reliability of *γ_j* and would be uncorrelated with composite score *Y_{ij'}* (*j* ≠ *j'*).

When the canonical profile dimensions are ordered according to the magnitude of their reliabilities, the reliability of the subspace spanned by the most reliable dimensions (cf. Bock, 1966; Conger, 1974) is

$$\hat{\rho}_s = \frac{1}{s} \sum_{j=1}^s \gamma_j \quad [4]$$

In case *s* equals the number of subscales *K*, this index equals the overall index of profile reliability found by Equation 1.

The canonical reliabilities, number of dimensions exceeding specific reliability values, and the reliability of selected subspaces are given in Table 2. The indices of overall profile reliability ranged from .53 to .61 with a median value of .58. Despite the similar magnitude of the overall profile reliabilities, the reliabilities of sequentially maximum profile dimensions showed differences across ages. At age 9½, for example, only 5 dimensions had a reliability of at least

Table 2
Multivariate and Canonical Dimension Reliabilities

Age	Mean Reliabilities ^a			Multivariate Reliabilities: Number of Dimensions			No. of Dimensions Exceeding a Reliability of			
	Total	Verbal	Perfor- mance	10 ^b	7	5	.95	.78	.64	.50
6½	.77	.75	.80	.60	.70	.77	1	2	4	7
7½	.77	.76	.78	.57	.69	.76	1	2	4	7
8½	.77	.78	.76	.60	.71	.77	1	2	4	7
9½	.78	.78	.75	.53	.65	.72	1	2	3	5
10½	.75	.79	.71	.55	.67	.73	1	2	3	6
11½	.81	.84	.78	.61	.72	.77	1	2	4	7
12½	.79	.85	.73	.59	.69	.75	1	2	4	7
13½	.79	.83	.74	.58	.67	.74	1	2	3	6
14½	.79	.83	.75	.56	.67	.74	1	2	4	6
15½	.76	.82	.72	.57	.68	.76	1	2	4	6 ^c
16½	.78	.83	.73	.60	.70	.76	1	2	4	6 ^c

^a These also correspond to a Cronbach - Gleser distance function reliability (cf Conger and Lipshitz, 1973).

^b This is the overall profile reliability as based on a Mahalanobis distance function (cf Conger and Lipshitz, 1973).

^c 16½ year olds had two additional dimensions at .50.

.50; at age 16½, 8 dimensions were as large as .50. The results also indicated some similarities across ages. For example, there was one extremely reliable dimension and a second highly reliable dimension at each age. In addition, the reliabilities of the 5- and 7-dimensional subspaces were generally comparable across ages.

Across age, similarity of unrotated solutions was further investigated in two ways. First, Tucker's ϕ coefficient of factor congruence (Harman, 1967, p. 270) was computed for each pair of weight vectors. The ϕ coefficients are like correlation coefficients in that they range from -1 to +1 with 0 indicating no agreement and -1 or +1 indicating perfect agreement. Second, weight vectors (Equation 2) derived for each age group were cross-applied to each of the other age groups to determine "cross-applied reliabilities"

using the general composite reliability formula (cf. Tellegen & Briggs, 1967):

$$\hat{\rho}_{jk'} = \frac{V'_{jk} R^{*tt} V_{jk}}{V'_{jk} R_{xx} V_{jk}} \quad [5]$$

where $\rho_{jk'}$ is the reliability for age group k' of a composite based on weight vector j from age group k . The true and observed score matrices R^{*tt} and R_{xx} were based on group k' .

The 10 dimensions at each of 11 ages generated a 110×110 matrix of ϕ coefficients. To the extent that the weight vectors were consistent across ages, this matrix would resemble a multitrait-multimethod matrix (Campbell & Fiske, 1959) with high agreement among numerically similar weight vectors (traits) and

low agreement elsewhere. The only strong patterning of this type occurred for the first and second dimensions. For the first dimension (Table 3, below the diagonal) the ϕ coefficients for ages $8\frac{1}{2}$ to $16\frac{1}{2}$ ranged from .90 to .99, with a median value of .97. Among these ages there was apparently excellent agreement in the patterning of weights. By comparison, ages $6\frac{1}{2}$ and $7\frac{1}{2}$ had weight vectors which agreed with one another ($\phi = .97$) but which showed increasing disagreement with age for the remaining age groups. For the second dimension (Table 3, above the diagonal) the ϕ coefficients indicated good agreement among ages $10\frac{1}{2}$, $11\frac{1}{2}$, and $13\frac{1}{2}$ to $16\frac{1}{2}$, with ϕ 's ranging from .92 to .98. Ages $8\frac{1}{2}$, $9\frac{1}{2}$, and $12\frac{1}{2}$ also tended to belong to this grouping; but although ages $6\frac{1}{2}$ and $7\frac{1}{2}$ had good agreement with each other ($\phi = .86$), their weight vectors agreed only moderately with those of the remaining ages (ϕ 's from .45 to .78).

Numerically similar unrotated dimensions beyond the second showed little across-age agreement; however, when the numerical ordering was ignored, some agreement among weight vec-

tors was evident. In these cases the most striking finding was that the magnitude of discrepancies in the numerical ordering was rather large. For example, the third most reliable dimension for ages $6\frac{1}{2}$ to $8\frac{1}{2}$ showed good agreement with the ninth most reliable (or second least reliable) dimension at age $12\frac{1}{2}$, with ϕ 's ranging from .70 to .84.

The cross-applied reliabilities (not shown) strongly supported the patterning of ϕ coefficients. The first and second most reliable composites always produced the highest and second highest cross-applied reliability, respectively. Beyond these two dimensions, a general trend of cross-applied reliabilities decreasing with degree of age difference could be noted for the third and fourth most reliable dimensions and cross-applied reliabilities increasing with age discrepancy for the two least reliable dimensions. Inconsistent patterns were observed elsewhere.

Both the weight vector patterns and magnitudes of cross-applied reliabilities indicated that the first and second most reliable dimensions can be generalized across age groups. The simi-

Table 3
Agreement Among First and Second
Most Reliable Composites^a

Age	Age										
	$6\frac{1}{2}$	$7\frac{1}{2}$	$8\frac{1}{2}$	$9\frac{1}{2}$	$10\frac{1}{2}$	$11\frac{1}{2}$	$12\frac{1}{2}$	$13\frac{1}{2}$	$14\frac{1}{2}$	$15\frac{1}{2}$	$16\frac{1}{2}$
$6\frac{1}{2}$		86	62	54	58	53	58	59	50	45	48
$7\frac{1}{2}$	97		78	78	67	62	60	72	65	57	55
$8\frac{1}{2}$	91	92		91	84	90	78	94	92	86	85
$9\frac{1}{2}$	89	92	97		80	78	65	89	89	84	78
$10\frac{1}{2}$	87	93	96	97		96	81	94	92	93	96
$11\frac{1}{2}$	85	93	96	96	99		86	97	94	93	97
$12\frac{1}{2}$	83	90	95	98	97	98		84	76	74	74
$13\frac{1}{2}$	80	86	94	98	98	98	98		98	95	94
$14\frac{1}{2}$	77	83	93	96	95	94	97	98		95	96
$15\frac{1}{2}$	71	80	90	93	94	94	94	98	98		95
$16\frac{1}{2}$	75	82	94	94	97	95	95	98	98	97	

^aElements below the diagonal are Tucker ϕ coefficients for 1st unrotated weight vectors; elements above the diagonal are for 2nd unrotated weight vectors. Decimal points are omitted.

larity of weight vectors beyond the first two when order and magnitude of reliability were ignored indicated some degree of agreement across age groups in what was being measured but a great degree of disparity in how well these remaining composites were being measured. Consequently, interpretation was sought only for the first two dimensions.

Weights for the most reliable dimension (Table 4) were all positive and relatively uniform. In addition, correlations (not shown) between WISC-R subscales and composites formed from these weights were all positive and relatively equal. These characteristics indicate that the most reliable dimension was essentially "profile level" or Full-Scale IQ. Despite the relative uniformity of weights, large systematic changes with age were manifested for seven scales (Table 4). Weights for Information and Vocabulary were positively correlated with age (r 's $\geq .86$); and weights for Similarities, Picture Completion, Picture Arrangement, Object Assembly, and Mazes were negatively correlated

with age (r 's $\leq -.83$). These age-related differences accounted for little variation in reliability, however, as indicated by the cross-applied reliabilities.

The second most reliable dimension at each age was clearly a Verbal-Performance contrast. The weights for Verbal scales were (with one minor exception) always of the same sign and differed in sign from those of the Performance scales (Table 5). Correlations between the second most reliable composite and subscales at each age had the same pattern as the weights but were substantially more uniform. Correlations between weights and age (Table 5) revealed substantial ($r \geq .50$) positive correlations for Information, Vocabulary, Picture Completion, and Mazes and substantial ($r \leq -.50$) negative correlations for Similarities and Block Design. Although differences in weights were evident, the largest increments in reliability over the reliability of a simple average Verbal score minus average Performance score contrast were .06 units for 16½ year olds and .05 units for 6½ year

Table 4
Weights for
Most Reliable Dimension by Age Group

Age	Scale									
	I	S	A	V	C	PC	PA	BD	OA	MZ
6½	09	23	14	12	09	19	14	16	12	16
7½	17	22	12	11	10	14	10	20	12	14
8½	16	16	08	25	11	19	09	21	08	10
9½	16	15	13	24	14	11	10	15	09	06
10½	20	17	12	23	10	09	08	24	08	06
11½	22	14	12	19	14	10	06	22	07	06
12½	19	15	10	22	19	08	09	17	05	04
13½	22	13	13	27	14	08	07	18	06	03
14½	23	14	08	31	14	07	07	14	06	04
15½	31	11	12	32	09	07	07	16	05	04
16½	23	14	07	33	10	08	05	22	06	03
Mean	.20	.16	.11	.24	.12	.11	.08	.19	.08	.07
St.Dev.	.06	.04	.02	.07	.03	.04	.02	.03	.02	.04
Correl. with Age	.86	-.83	-.47	.87	.19	-.85	-.83	-.07	-.88	-.88

Note. Decimal points are omitted.

Table 5
Weights for Second Most
Reliable Dimension

Age	Scale									
	I	S	A	V	C	PC	PA	BD	OA	MZ
6½	06	92	08	13	06	-18	-12	-31	-29	-36
7½	23	52	19	19	09	-36	-03	-22	-16	-60
8½	28	21	01	52	18	-47	-01	-46	-19	-28
9½	34	10	29	48	07	-48	-21	-32	-27	-30
10½	43	18	18	32	16	-13	-13	-81	-16	-25
11½	38	13	01	43	29	-21	-03	-81	-23	-18
12½	30	28	-12	20	66	-26	-31	-63	-23	-17
13½	31	12	05	40	17	-25	-11	-57	-32	-24
14½	27	03	09	57	14	-23	-13	-62	-30	-27
15½	45	02	01	43	01	-16	-20	-61	-24	-16
16½	30	09	10	53	09	-14	-09	-84	-22	-11
Mean	.30	.24	.08	.38	.17	-.26	-.12	-.56	-.24	-.26
St.Dev.	.11	.27	.11	.15	.18	.12	.09	.21	.05	.13
Correl. with Age	.53	-.74	-.32	.56	.07	.50	-.29	-.70	-.26	.74

Note. Decimal points are omitted.

olds. Larger increments were evident when the age-derived Verbal-Performance composites were compared to Verbal IQ-Performance IQ reliabilities computed from Wechsler's data (increments ranged from .01 to .08, with a median value of .07). The differences between the canonical dimensions and Wechsler's approach indicate that a more reliable contrast could be established; however, the simplicity of comparing average Verbal and average Performance scores or Verbal and Performance IQs may well warrant the loss in reliability.

Rotated Dimensions

The preceding results provided a clear picture only for the two most reliable dimensions. Because of the lack of commonality across ages for the remaining dimensions, further clarity was sought among Varimax-rotated subspaces; however, the differing numbers of "reliable" dimensions across ages indicated that no single cut-off for the number of dimensions rotated would be

entirely appropriate for all ages. Thus, 4, 5, 6, and 7 dimensions were rotated for each age group. The rotated subspaces were then compared both within age groups and across age groups. Comparisons of rotated subspaces were based on ϕ coefficients of factor congruence and cross-applied reliabilities.

Within-age comparison of different rotations indicated that some of the weight vectors were invariant across rotations but others changed drastically. The invariant dimensions were generally "specific" contrasts, i.e., one subscale had a very high weight (.90 or above) and all other subscales had low weights (.30 to $-.30$), or "doublets" in which two subscales had comparable weights and all other subscales had low weights. Every subscale except Object Assembly emerged as a specific canonical contrast for some age group: Picture Arrangement emerged for every age group; Block Design and Arithmetic for all but one age group; and Information, Picture Completion, Mazes, and a Vocabulary-Comprehension sum for more than

Table 6
 Canonical Specifics and Doublets By Age,
 Scale and Rotation Canonical Dimension

AGE	Specifics										Doublets		
	I	S	A	V	C	PC	PA	BD	OA	MZ	I+A	I+V	V+C
6½		(4-7)	7			(4-7)	7	7		(4-7)			
7½		(5-7)				(4-7)	7	7		(5-7)	(4-7)		6,7
8½			7	(4-7)		(4-7)	7	(4-7)		(4-7)	5,6		
9½	7		(4-7)			6,7	(5-7)	6,7		7		6	
10½	6,7		6,7			7	(5-7)	4,6,7		6,7	4,5		7
11½	7		7			5,7	7	(4-7)		5,7		4	5,6
12½	6,7	7	6,7	6,7	(4-7)		6,7	6,7				4,5	
13½			4,6			6	6,7					7	(4-6)
14½	4		7	7			(5-7)			7	5		4,5
15½	(4-7)		(4-7)	6			(5-7)	(4-7)		(4-7)	5		5,7
16½	(4-7)		(4-7)	6,7			6,7	4,5,7		6,7			

Note. Numerical entries denote presence of specific or doublet when that number of dimensions were rotated.

half of the age groups. Since the rotated 7-dimensional solution contained all of the specifics and doublets (listed in Table 6), across-age comparisons were based on this solution.

The 7 rotated dimensions for each of the 11 age groups generated a 77×77 matrix of ϕ coefficients. A visual examination of this matrix indicated strong agreement among many weight vectors for all or most ages. Clear clusters of specific canonical contrasts were shown² for Information (ages 9½, 10½, 12½, 14½ to 16½); Similarities (6½ to 8½, 12½, 16½); Arithmetic (all ages); Picture Completion (6½ to 11½, 13½); Picture Arrangement (all ages); Block Design (all ages); and Mazes (6½ to 11½, 13½ to 15½). The only scales not showing an obvious clustering were Vocabulary, Comprehension, and Object Assembly.

The Vocabulary and Comprehension subscales did emerge as specifics at some ages, but at other ages they were present as a doublet. A closer analysis of the ϕ coefficients for these composites indicated a strong simplex patterning in which ages 12½ and 16½ showed clear canonical specifics for both Vocabulary and Comprehension, but the remaining age groups had a single doublet dimension with varying contributions of Vocabulary (V) and Comprehension (C). Ordering on the basis of the ratio of the Vocabulary weight to the Comprehension weight (from large V/C to small V/C) gave the following ordering of ages: 12½, 16½, 14½, 8½, 15½, 10½, 9½, 7½, 6½, 11½, 13½, 12½, and 16½.

While the analysis of ϕ coefficients indicated some definite commonality across ages, the mere absence of a composite as an axis of a rotated solution does not indicate that the composite does not exist somewhere in the subspace of reliable dimensions. The within-age comparisons (across rotated solutions) showed that some specifics were present when 5 or 7 dimensions were rotated, but not when 6 were. Consequently, weight vectors falling within a cluster

²Discrepancies between these results and those shown in Table 6 are due to the method of inferring a "specific." In Table 6, specifics were found within ages by the pattern of weights. For these results, ϕ 's among weight vectors were clustered on the basis of magnitude, and then a representative vector was given an interpretation. The discrepancies are minimal.

were cross-applied to all ages to verify the clustering and to determine the possible presence of dimensions that were not one of the axes of the rotated solution. The cross-applied reliabilities confirmed the clustering of weight vectors based on ϕ coefficients and also indicated that at some ages such a dimension, although not manifest as an axis, was as reliable as when it emerged as an axis.

The above results suggested the possibility that a clearer understanding might be obtained by selecting a subset of "marker" variables that most clearly showed specific contrast or doublet characteristics and calculating reliabilities for these markers across all ages. Table 7 shows the reliabilities of marker dimensions at each age along with reliabilities of comparable dimensions derived within that age group. It is important to note that of the 54 comparisons of marker-to-age-derived dimensions (excluding

marker = age derived), 30 markers were as good as the age-derived dimensions. This indicated that age-derived dimensions were not capitalizing on an optimization phenomenon and that the results were robust. Secondly, while the marker dimensions yielded reasonable reliabilities even if a dimension was not apparent in the rotated solution, some of the reliabilities were rather low and some general trends were apparent. For example, reliabilities of the Similarities, Picture Completion, and Mazes canonical specifics appeared to be negatively correlated with age; and reliabilities of the Information, Vocabulary, and Block Design specifics appeared to be positively correlated with age. The initial analysis of univariate subscale reliabilities showed similar results. In fact, the relationship between the specific canonical contrasts and univariate reliabilities were even more dramatic than the word "similar" conveys. Figure 1 shows

Table 7
Reliabilities of Marker Specifics
and Doublets by Age

AGE	Verbal								Performance			
	S	A	A+I	I	I+V	V	V+C	C	PC	PA	BD	MZ
6½	(80)	64	68	41	68	47	65	52	77	66	74	73
7½	76	60	(73)	58	75	40	(65)	50	(73)	63	74	(71)
8½	67	59	67	60	83	68	71	53	79	58	79	69
9½	61	67	73	56	(82)	55	71	59	65	66	70	60
10½	62	(66)	74	(65)	83	57	66	52	54	65	79	56
11½	65	70	77	68	87	61	77	66	71	65	83	68
12½	71	68	73	69	86	64	80	(74)	59	68	78	48
13½	61	70	75	67	87	66	78	62	63	62	80	57
14½	62	59	75	68	87	70	78	64	56	64	78	64
15½	55	72	78	77	89	70	71	51	51	65	78	53
16½	70	64	77	74	88	(72)	75	59	63	(63)	(85)	44

Source of Marker:

Age	6½	10½	7½	10½	9½	16½	7½	12½	7½	16½	16½	7½
Rotation	7	7	7	7	6	7	7	7	7	7	7	7

Note. Decimal points omitted. Numbers in parentheses indicate reliability of marker within age group from which it was taken.

graphs of univariate and specific canonical reliabilities across ages for each scale. A close look at these figures reveals that the canonical specific contrast reliabilities and univariate reliabilities have extraordinarily similar shapes.

Rational Alternatives

The rotated structure of the WISC-R may be described as a set of canonical specifics contrasting each of nine subscales with the remaining subscales. Although this overall structure did not emerge for every age group, the canonical specifics were common to many age groups, and their weight patterns were obviously robust. Despite these desirable characteristics, an alternative approach using rational weights may be proposed. That is, since the canonical specifics contrast one subscale with the remaining subscales, would it not be simpler to compare each subscale with the simple average of all scales or with the simple average of respective Verbal or Performance subscales?

To investigate these alternative formulations, two rational contrast composites were established for each subscale. These consisted of (1) contrasts with total, which had the form $X_j - M_a$, where X_j is the targeted subscale score and M_a is the average of all subscale scores; and (2) contrasts with Verbal/Performance, which had the form $X_j - M_{v/p}$, where X_j is the targeted subscale and $M_{v/p}$ is the average of Verbal or Performance scales depending on whether X_j is classified as a Verbal or Performance scale. Reliabilities for these composites were computed at each age (using Equation 5) and are shown in Figure 1. With few exceptions, reliabilities of the two rational contrasts have the same shape across ages as the univariate and canonical specific reliabilities. However, the canonical reliabilities were almost always the highest; and the within-Verbal/Performance reliabilities, the lowest.³ Apparently, the differential weighting of

the canonicals was primarily advantageous for Similarities and Block Design; however, the Block Design canonical specific deviated from the patterning of all other canonical specifics in that another scale (Object Assembly) had a relatively high same-signed weight. The ratio of the Block Design weight to Object Assembly weight was about 4:1; for every other canonical specific, the ratio of the weight of the targeted scale to the next largest same-sized weight exceeded 8:1.

Discussion

Based on the preceding analyses, which appear to be mutually confirmatory, a clear picture can be offered for what the WISC-R can reliably measure. First of all, the WISC-R allows a highly reliable comparison of profile levels (or total IQ) at every age. Secondly, reasonably reliable comparisons of Verbal and Performance profile levels (or Verbal and Performance IQs) can be made at every age. The results also indicated that beyond these two dimensions the WISC-R subscales at different ages possess a similar structure in terms of what can be measured, but the reliability or precision of measurement varies across ages, sometimes dramatically so. This structure, which emerged in the rotated subspaces, can best be conceptualized as a set of composites contrasting a single subscale with the average of all remaining subscales.

A strategy for profile interpretation that takes into account the above findings would be as follows:

1. Compare profiles on the basis of total IQ.
2. Compare Verbal IQ to Performance IQ.
3. Compare individual scales to the average of all scales.

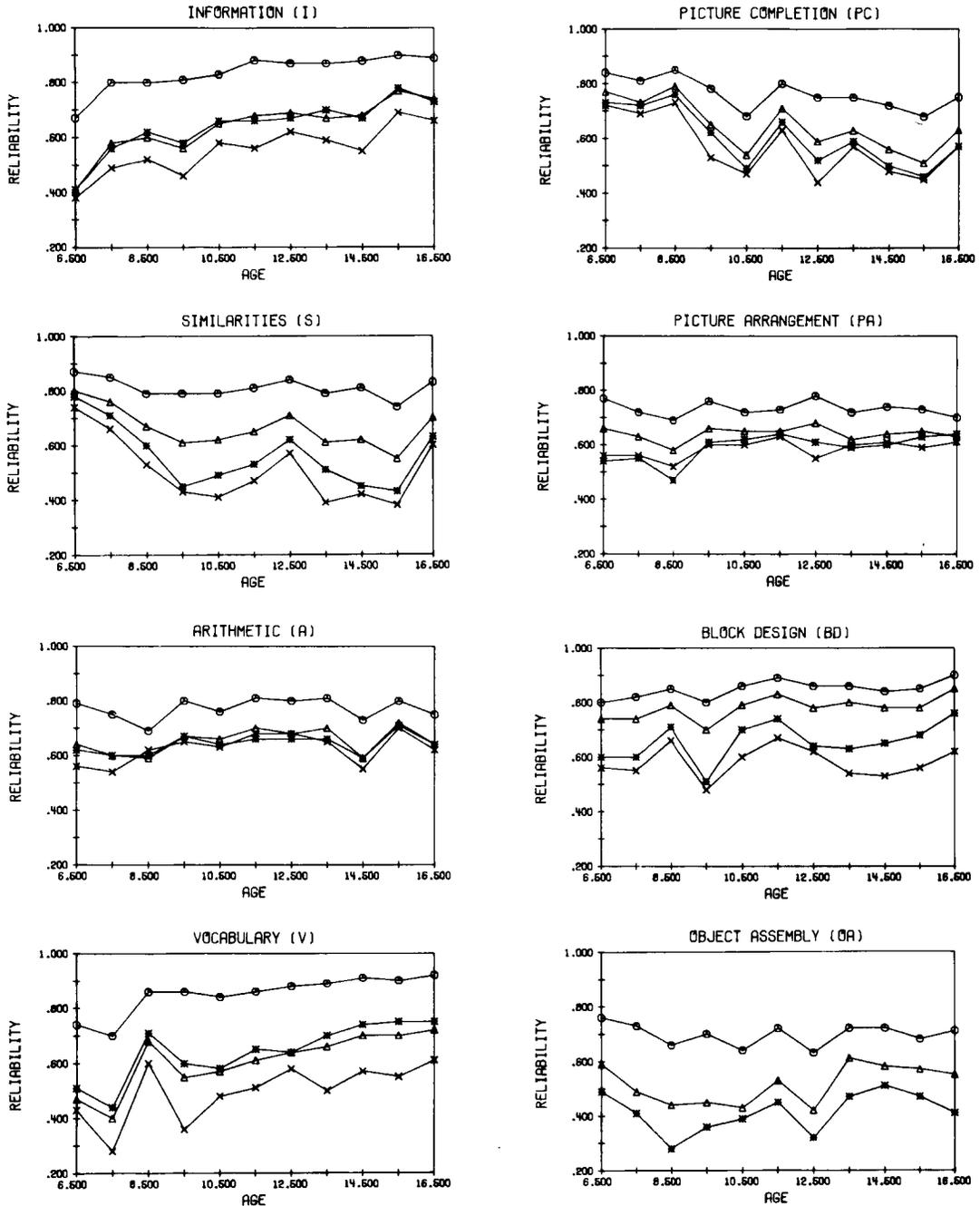
An alternative strategy involving less reliable comparisons would be as follows:

1. Compare profiles on Verbal and Performance IQ separately;
2. Compare Verbal IQ to Performance IQ; and

³Deviations of rational specifics from the canonical/univariate pattern can be accounted for by unusually high or low correlations of that scale at that age with Total score, Verbal or Performance score, or with some other scale.

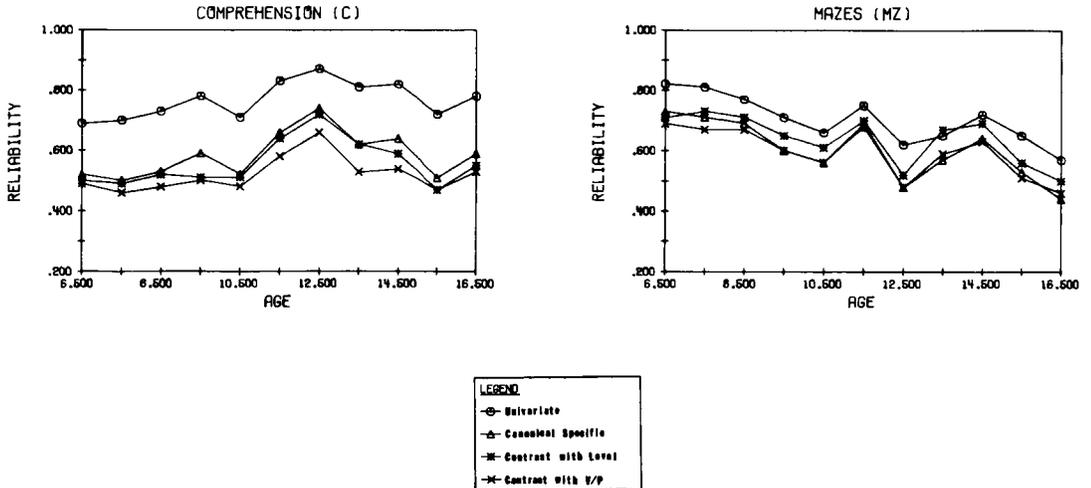
Figure 1

Univariate, Canonical Specific, and Rational Contrast Reliabilities
as a Function of Age



(continued on next page)

Figure 1 (continued)



3. Compare each subscale to respective Verbal or Performance scale averages.

More complex comparisons are, of course, possible, but their reliability should be computed at each age (using Equation 5). It is possible, however, to generalize along the following lines: Doublets, triplets, and so forth, which compare a sum of Verbal scales to a sum of Performance scales will have reasonable reliabilities; however, contrasts among two or more Verbal or among two or more Performance scales will have low reliabilities.

The strategy offered above may be conceptualized as a sequence of contrasts resembling a design matrix in the general linear model. Thus, the set of comparisons among the subscales involve computing total score or profile level using weights of (1, 1, . . . 1); a Verbal-Performance contrast using weights of (1 1 1 1 -1 -1 -1 -1 -1); and a set of contrasts comparing each subscale with the total using weights of (1, -1/9, -1/9, . . .), (-1/9, 1, -1/9, . . .), The strategy offered thus minimizes capitalization on chance differences by restricting the number of comparisons and simultaneously offers a series of relatively uncorrelated comparisons

allowing simpler interpretation. By comparison, the often implicitly and explicitly recommended procedure of comparing all individual scales to one another is analogous to doing all possible t -tests, a procedure known to capitalize on chance differences and one which necessarily involves correlated comparisons.

In order to facilitate the use of specific contrasts, Table 8 provides standard errors of measurement for each scale by age. Simultaneously, it is recommended that contrasts among all possible pairs of subtests not be done; and users of the WISC-R are cautioned *not* to use Table 12, p. 35, of the WISC-R manual (Wechsler, 1974) because it does not take into account (1) the number of comparisons; (2) intercorrelations among comparisons; and (3) variation in reliabilities across ages.

The results obtained in these analyses indicate that the WISC-R differs across age groups in terms of the reliability of specified contrasts and in terms of which composites are most reliable. These results are seemingly in opposition to the verbal inferences about the factor structure invariance of the WISC-R across ages (Kaufman, 1975). The series of analyses undertaken in this investigation, however, indicates that (1) studies

Table 8
Standard Errors of Measurement of Specific Contrasts
in Scale Score Units by Age

Test	Age Group										
	6½	7½	8½	9½	10½	11½	12½	13½	14½	15½	16½
Information	1.8 (1.9)	1.4 (1.5)	1.4 (1.5)	1.4 (1.5)	1.3 (1.4)	1.2	1.2	1.2	1.2	1.1 (1.2)	1.1 (1.2)
Similarities	1.2 (1.5)	1.3 (1.5)	1.5 (1.6)	1.5 (1.6)	1.5	1.4	1.3	1.5	1.4	1.6	1.3 (1.4)
Arithmetic	1.5 (1.6)	1.6 (1.7)	1.8	1.4 (1.5)	1.5 (1.6)	1.4	1.4	1.4	1.6 (1.7)	1.4	1.6
Vocabulary	1.6 (1.7)	1.7 (1.8)	1.2 (1.3)	1.2 (1.3)	1.3 (1.4)	1.2 (1.3)	1.2	1.1 (1.2)	1.0 (1.2)	1.1 (1.2)	1.0 (1.1)
Comprehension	1.8	1.7 (1.8)	1.6 (1.7)	1.5 (1.6)	1.7	1.3 (1.4)	1.2 (1.3)	1.4	1.4	1.7	1.5
Picture Completion	1.3	1.4 (1.5)	1.3	1.5 (1.6)	1.8 (1.9)	1.4 (1.5)	1.6 (1.7)	1.6 (1.7)	1.7	1.8 (1.9)	1.6 (1.7)
Picture Arrangement	1.5 (1.6)	1.7	1.8	1.6 (1.7)	1.7 (1.8)	1.6 (1.7)	1.5 (1.6)	1.7 (1.7)	1.6 (1.7)	1.6 (1.7)	1.7 (1.8)
Block Design	1.4 (1.5)	1.4 (1.5)	1.3 (1.4)	1.4 (1.6)	1.2 (1.4)	1.1 (1.2)	1.2 (1.4)	1.2 (1.4)	1.3 (1.4)	1.3 (1.4)	1.1 (1.3)
Object Assembly	1.6	1.6 (1.7)	1.8 (1.9)	1.7 (1.8)	1.9 (2.0)	1.7	1.9 (2.0)	1.7	1.7	1.8 (1.9)	1.7 (1.8)
Mazes	1.4	1.4 (1.5)	1.5 (1.6)	1.7 (1.8)	1.8 (1.9)	1.6 (1.7)	1.9 (2.0)	1.8 (1.9)	1.7	1.8 (1.9)	2.0 (2.1)

Note: The top row for each test is the SEM for contrast with the average of all scales. When the SEM for contrast with the average of other subtests of the same type (i.e., verbal/performance) differs from the SEM for the average of all scales, it is presented in parentheses in the bottom row for each test. The SEM's provided in this table can be compared to Table 10, p. 30 of the WISC-R Manual (Wechsler, 1974); SEM's for contrasts with average exceed univariate SEM's by .1 or more for 57% of the entries (otherwise they are equal).

of the factor structure of the WISC-R should demonstrate rather comparable across-age factors (i.e., a null hypothesis that correlation matrices differ across ages was not rejected) but that (2) although the same structure may exist, the saliency of factors may vary from age to age, and the reliability of factors and other composites do vary from age to age (i.e., the null hypothesis that univariate reliabilities and the reliability of multivariate contrasts are the same across ages was clearly rejected).

A question remains about the meaning of the obtained differences. Quite striking correlations exist between reliabilities and age attesting to a systematic rather than a random phenomenon. Scales such as Information, Vocabulary, and Comprehension are all highly saturated with school-related questions. Perhaps educational experiences enhance individual differences (true score variances) on these scales with an end result of greater reliability at older ages. Alternatively, these scales may be tapping different in-

tellectual factors at different ages, or performance may be more random for younger children (increasing error variance). Block Design, however, also shows increasing reliability with age. This Performance scale obviously is not education related and suggests that within-test format differences (e.g., four Blocks versus nine Blocks, bonus points for time) or increased individual differences would be the source of increasing reliabilities.

Similarities, Picture Completion, and Mazes show decreasing reliabilities with age. Again, no educational source of variance would seem to account for this pattern. Is it the case that children become more alike in their performance on these subscales (reducing true score variance) or more random in performance (increasing error variance)? It seems unlikely that behavior should become more random with age, particularly in light of increasing reliabilities on four other scales. It seems more likely that measured individual differences are decreasing with age. The decrease could be due to either a reduction in individual differences for the domain covered by the items or may simply be due to ceiling effects.

No single reason for age-related individual differences can be definitively offered from these results, but it does appear that there are age-related individual differences that are manifest in univariate reliabilities and in both derived and rational contrasts. Additionally, 6½ and 7½ year olds seem most different from the remaining ages for both the specific comparisons and the structure of the two most reliable contrasts. The question is raised, Is the WISC-R for 6½ and 7½ year olds really the same test as the WISC-R for 8½ to 16½ year olds?

The results obtained in the analysis of the WISC-R also may be compared to those obtained by Conger and Conger (1975) in their analysis of the WISC. First, the WISC-R profiles are more reliable than the WISC profiles. The average WISC-R canonical reliability was .58, whereas the average WISC canonical reliability was only .51. Second, for the WISC, Conger and

Conger found only five canonical dimensions at each of three age groups which exceeded a reliability of .55; the WISC-R by comparison shows an average of six dimensions exceeding that reliability.

The structure of the WISC-R is also clearer. Although both the WISC and WISC-R show highly similar first and second unrotated dimensions, the WISC-R shows a strong pattern of specific contrasts not evident on the WISC. Users of the WISC-R can more confidently interpret specific scale contrasts (with total) than can WISC users. There remains, however, a question of whether or not something might have been lost. The WISC analysis did show some reliable composites similar in structure to composites that were used to measure Verbal Comprehension, Relevance, and Perceptual Organization (Robb et al., 1972). With some apparent changes in the structure of what is being reliably measured on the WISC-R versus the WISC, it seems necessary to investigate further both clinically and empirically derived diagnostic procedures based on the WISC in terms of their applicability to the WISC-R.

References

- Anastasi, A. *Psychological testing*. New York: Macmillan, 1968.
- Bock, R. D. Contributions of multivariate experimental designs to educational research. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology*. Chicago: Rand McNally, 1966.
- Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, 56, 81-105.
- Conger, A. J. Estimating profile reliability and maximally reliable composites. *Multivariate Behavioral Research*, 1974, 9, 85-104.
- Conger, A. J., & Conger, J. C. Reliable dimensions for WISC profiles. *Educational and Psychological Measurement*, 1975, 35, 847-863.
- Conger, A. J., & Lipshitz, R. Measures of reliability for profiles and test batteries. *Psychometrika*, 1973, 38, 411-427.
- Conger, A. J., & Stallard, E. Equivalence among canonical factor analysis, canonical reliability

- analysis, and principal components analysis: Implications for data reduction of fallible measures. *Educational and Psychological Measurement*, 1976, 36, 619-626.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. *The dependability of behavioral measurements*. New York: Wiley, 1972.
- Harman, H. H. *Modern factor analysis*. Chicago: University of Chicago Press, 1967.
- Hays, W. L. *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart, & Winston, 1973.
- Kaufman, A. S. Factor analysis of the WISC-R at 11 age levels between 6½ and 16½ years. *Journal of Consulting and Clinical Psychology*, 1975, 43, 135-147.
- Kaufman, A. S. A new approach to the interpretation of test scatter on the WISC-R. *Journal of Learning Disabilities*, 1976, 9, 160-168. (a)
- Kaufman, A. S. Verbal-Performance IQ Discrepancies on the WISC-R. *Journal of Consulting and Clinical Psychology*, 1976, 44, 739-744. (b)
- Littell, W. M. The WISC—A review of a decade of research. *Psychological Bulletin*, 1960, 57, 132-162.
- McNemar, Q. On WAIS difference scores. *Journal of Consulting Psychology*, 1957, 21, 239-240.
- Piotrowski, R. J., & Grubb, R. D. Significant subtest score differences on the WISC-R. *Journal of School Psychology*, 1976, 14, 202-206.
- Robb, G. P., Bernardoni, L. C., & Johnson, R. W. *Assessment of individual mental ability*. Scranton, PA: International Text Book Co., 1972.
- Sattler, J. M. *Assessment of children's intelligence* (Rev. ed.). Philadelphia: Saunders, 1974.
- Tellegen, A., & Briggs, P.F. Old wine in new skins: Grouping Wechsler subtests into new scales. *Journal of Consulting Psychology*, 1967, 31, 499-506.
- Timm, N. H. *Multivariate analysis with applications in education and psychology*. Monterey, CA: Brooks/Cole, 1975.
- Wallbrown, F. H., Blaha, J., Wallbrown, J. D., & Engin, A. The hierarchical factor structure of the Wechsler Intelligence Scale for Children-Revised. *The Journal of Psychology*, 1975, 89, 223-235.
- Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children*. New York: The Psychological Corporation, 1949.
- Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children-Revised*. New York: The Psychological Corporation, 1974.
- Zimmerman, I. L., & Woo-Sam, J. Research with the Wechsler Intelligence Scale for Children: 1960-1970. *Journal of Clinical Psychology*, 1972, Monograph Supplement No. 33.

Acknowledgments

This research was supported in part by a grant from the Department of Education, Purdue University, to Judith Cohen Conger and by funds from NIMH Training Grant No. 5 T01 MH14908-02.

Author's Address

Send requests for reprints or further information to Anthony J. Conger, Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907.