Psychological Assessment:
A Journal of Consulting and Clinical Psychology
1989, Vol. 1, No. 2, 155-159

1040-3590/89/$00.75

On WAIS-R Difference Scores in a Psychiatric Sample

Ralph L. Piedmont, Robert L. Sokolove, and Michael Z. Fleming
Boston University

This report examines the psychometric integrity of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) subscales, and the differences between them, in a sample of 229 psychiatric patients from 2 community mental health centers (ages 16 to 85). The results verify the overall alpha and split-half reliabilities of the instrument and indicate that greater caution needs to be exercised in clinically evaluating difference scores. Cutoff values presented in the manual appear too low to be of any statistical or diagnostic merit. Distributions for each of the 55 possible difference scores found in this sample are presented and provide a better guide for making nosological determinations.

Subscale scores on the Wechsler Adult Intelligence Scale-Revised (WAIS-R) are frequently used in nosological classification (Allison, Blatt, & Zimet, 1988; Piedmont, Sokolove, & Fleming, in press-b; Widiger, 1982), and there exist theoretical rationales that link certain performance outcomes to relevant psychological dynamics (Rapaport, Gill, & Schafer, 1968). Clinical attention is usually paid to differences between particular subtest scores. Although the recent WAIS-R manual (Wechsler, 1980) provides cutoff values for determining whether any difference between subscales is significant, it does not provide the norms for, or the reliabilities of, any of the 55 difference scores possible among the 11 subtests. Furthermore, the critical values that are presented are at a significance level (15%) not commonly accepted by psychologists and are based on estimates from a nonpsychiatric sample. The purpose of this article is to report the internal consistency of the WAIS-R subscales and their difference scores in a clinical sample in order to determine the psychometric integrity of the instrument with such individuals and to estimate the magnitude of difference necessary to attain statistical, and—it is hoped—clinical, significance.

Method

Subjects were 229 psychiatric patients (146 men, 83 women) receiving either in- or outpatient care from the Massachusetts Mental Health Center or the Boston University Medical Center. The average age at the time of testing was 32.7 years (SD = 12.1; low = 16, high = 85). Included were 144 Caucasians, 8 Hispanics, 2 Asians, 62 Blacks, and 13 persons with unknown ethnic background. Sixty percent were unemployed, 22.7% were involved in manual, clerical, or technical pursuits, 6% had white-collar positions, 8% were students, 3% were housewives, and the occupations of the remaining percentage were unknown. Overall performance was in the low average range: Mean Verbal IQ was 93.6 (SD = 17.6), mean Performance IQ was 86.4 (SD = 15.1), and mean Full Scale IQ was 89.6 (SD = 16.3). Diagnoses of these subjects represented a wide range of psychopathology including organic disorders (n = 10), schizophrenia (n = 82), affective disorders (n = 46), psychoses (n = 20), anxiety syndromes (n = 5), substance abuse (n = 4), adjustment disorders (n = 13), and personality disorders (n = 49). The diagnostic process has been documented elsewhere (cf. Piedmont, Sokolove, & Fleming, in press-b). For the purposes of this report all reliability estimates were made collapsing over age and diagnosis.

Results and Discussion

The split-half and alpha reliability estimates were calculated for each of the subscales with the exception of Digit Symbol and Digit Span (reliabilities for these scales are usually calculated by a test-retest method, and such information was not available in our sample). Although the alpha coefficients appeared to parallel nicely those given in the last column of Table 10 in the manual (low: Object Assembly = .77; high: Vocabulary = .96), the split-half values were generally lower than the normative coefficients (low: Arithmetic = .72; high: Vocabulary = .89). Nonetheless, the WAIS-R remains psychometrically robust in a clinical sample.

Table 1 presents the average intercorrelations between the subscales. It is interesting to note that these correlations are generally higher than the overall values presented in the manual. Of the 55 coefficients, 39 are larger and 13 are significantly so. It is unclear why the subscale scores are more correlated in this sample (only the correlation of Digit Span with Comprehension is significantly lower than its respective normative value). This may possibly be an artifact of the clinical sample. One quality underlying psychopathology is the return of the individual to a more concrete, object-literal level of interaction; higher cognitive functioning becomes impaired. Although this may create lower performance scores normatively, it may create greater consistency in scores from scale to scale.

Table 2 presents the reliabilities and distribution characteristics for differences between subscales. The standard deviations of the difference scores, σ_{DD} values, are the Arabic numbers

We would like to thank Arje Latz and June G. Wolf for providing access to their testing files and James Bernhard, Naomi Brunner, Timothy Scott, and Deborah Streeter for their assistance in collecting data.

Correspondence concerning this article should be addressed to Ralph L. Piedmont, who is now at the National Institute on Aging, National Institutes of Health, Gerontology Research Center, Laboratory of Personality and Cognition, 4940 Eastern Avenue, Baltimore, Maryland 21224.

Table 1
Intercorrelations Among Subscales of the Wechsler Adult Intelligence Scale–Revised

Subtest	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Information										
2. Comprehension	.74**									
3. Arithmetic	.73***	.64*								
4. Similarities	.70	.73	.62*							
5. Digit Span	.63***	.53	.66	.58						
6. Vocabulary	.83	.78	.70**	.77*	.63**					
7. Digit Symbol	.46	.47	.50	.46	.41	.46				
8. Picture Completion	.49	.48	.48	.53	.43	.47	.50			
9. Block Design	.56	.50	.61	.55	.53**	.57	.53	.61		
10. Picture Arrangement	.49	.46	.48	.52	.46*	.46	.50**	.64***	.64***	
11. Object Assembly	.40	.43	.40	.44	.36	.40	.51***	.62**	.70**	.59***

Note. Asterisks denote that correlation is significantly higher than normative value. Tests for differences were done using Fisher's r to Z transformations. See Cohen and Cohen (1975, p. 51).
* $p < .10$. ** $p < .05$. *** $p < .01$; one-tailed.

above the diagonal. Note that these values vary from 2.2 to 3.7 ($M = 3.1$) scaled score points. Such values are not given in the manual. The reliability coefficients (r_{DD}), calculated by a formula given by Cohen and Cohen (1975, p. 64), for the 55 sets of difference scores are presented below the diagonal. These vary from a low of .35 to .83, with a median of .67. Only 20 of the 55 reliability coefficients (36%) exceed the marginal value of .70.

Errors of measurement for the difference scores ($\sigma_{e(d)}$) are set in italics above the diagonal in Table 2. They were calculated according to the formula outlined by McNemar (1957). Given the marginal reliabilities, these errors of measurement are expectedly large relative to their respective standard deviations. As such, a sizable portion of a given difference-score variance is attributable to errors of measurement. The implications of these values can be seen when one looks for meaningful differences between performance subtests. Because such differences always tend to capitalize on chance, it is necessary to apply stringent criteria in evaluating such differences. For example, a 5-point difference between Information and Object Assembly would be "significant" at the .01 level as far as error of measurement is concerned. However, the standard deviation of 3.6 indicates that 17% of a psychiatric sample should have a difference of such magnitude between these scales. As such, McNemar

Table 2
Reliabilities (Below Diagonal) and Standard Deviations for Difference Scores (Roman Type) and Errors of Measurement of the Difference Scores (Italics)

Subtest	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Information		2.7	2.6	2.8	3.0	2.2	3.3	3.4	3.1	3.5	3.6
		1.6	*1.7*	*1.7*	*1.7*	*1.2*	*1.6*	*1.6*	*1.6*	*1.8*	*1.8*
2. Comprehension	.63		2.9	2.6	3.3	2.4	3.2	3.4	3.2	3.5	3.5
			1.8	*1.8*	*1.8*	*1.4*	*1.7*	*1.7*	*1.7*	*1.9*	*2.0*
3. Arithmetic	.56	.60		2.9	2.6	2.8	2.8	3.1	2.6	3.2	3.3
				1.9	*1.8*	*1.7*	*1.7*	*1.7*	*1.7*	*1.9*	*1.9*
4. Similarities	.65	.52	.59		3.0	2.4	3.1	3.1	3.0	3.2	3.3
					1.8	*1.5*	*1.7*	*1.7*	*1.9*	*1.8*	*1.9*
5. Digit Span	.68	.69	.50	.63		3.0	3.1	3.2	2.8	3.2	3.4
						1.6	*1.7*	*1.7*	*1.7*	*1.8*	*1.9*
6. Vocabulary	.68	.64	.65	.61	.72		3.5	3.5	3.2	3.7	3.7
							1.6	*1.4*	*1.8*	*1.7*	*1.7*
7. Digit Symbol	.77	.72	.65	.70	.70	.80		2.8	2.6	2.9	2.7
								1.6	*1.6*	*1.7*	*1.7*
8. Picture Completion	.79	.75	.70	.70	.73	.83	.68		2.6	2.7	2.6
									1.6	*1.8*	*1.8*
9. Block Design	.74	.72	.58	.61	.65	.77	.64	.62		2.6	2.2
										1.8	*1.8*
10. Picture Arrangement	.75	.72	.66	.67	.68	.80	.64	.56	.53		2.8
											2.0
11. Object Assembly	.75	.69	.67	.67	.69	.78	.58	.51	.35	.50	

(text continues on page 158)

Table 3

Percentage of Cases Having a Particular Difference Score

Difference between	Point difference																
	−8+	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8+
Information–Similarities	0.4	0.0	0.4	0.9	4.8	7.4	10.5	8.7	18.8	12.7	10.9	8.7	7.0	5.2	2.2	0.9	0.4
Information–Digit Span	0.4	0.4	1.3	2.6	4.4	9.6	10.9	13.5	17.5	10.0	8.3	6.6	6.6	3.9	2.6	0.4	0.8
Information–Comprehension	0.4	0.4	0.0	1.3	1.7	7.0	12.2	16.6	16.6	14.4	9.6	7.9	3.9	3.9	1.3	1.7	0.8
Information–Arithmetic	0.0	1.3	0.0	.4	1.8	2.2	11.0	11.0	17.6	15.0	13.2	12.8	6.6	2.2	3.1	0.9	0.9
Information–Vocabulary	0.0	0.4	0.4	1.3	4.0	5.3	13.2	16.7	22.5	16.7	10.1	4.4	2.2	1.8	0.9	0.0	0.0
Information–Digit Symbol	0.0	0.0	0.4	0.4	1.8	4.4	7.0	3.1	11.0	11.5	13.7	12.3	9.7	5.3	6.6	4.0	8.8
Information–Picture Completion	0.0	1.3	0.0	1.8	3.9	4.8	7.0	8.8	12.7	11.8	9.6	8.3	8.8	8.8	4.4	3.5	4.4
Information–Block Design	0.0	0.4	1.3	1.3	3.1	4.4	12.3	8.3	14.5	11.4	13.6	5.3	7.9	8.3	3.1	1.8	3.1
Information–Picture Arrangement	0.4	0.0	1.3	2.2	4.0	4.0	11.1	6.6	12.8	11.1	10.6	8.4	8.0	6.2	4.9	4.9	3.6
Information–Object Assembly	0.0	0.9	0.0	2.2	5.3	6.1	6.6	7.9	11.4	10.1	12.3	7.9	7.5	7.9	5.7	2.2	6.2
Comprehension–Similarities	0.0	0.4	1.3	.9	4.8	7.4	10.9	14.8	17.8	10.9	10.0	9.6	5.7	3.9	1.7	0.0	0.0
Comprehension–Digit Span	2.5	0.0	1.7	5.7	5.7	7.0	12.6	11.3	13.5	11.7	8.3	7.8	3.9	4.8	1.7	1.7	0.0
Comprehension–Arithmetic	0.4	1.3	0.9	1.8	1.8	3.9	9.6	15.4	19.7	13.2	9.2	8.3	5.7	3.1	3.1	1.3	0.9
Comprehension–Vocabulary	0.4	0.4	2.2	3.5	5.7	5.3	15.4	17.1	13.6	15.8	11.8	5.7	1.8	1.3	0.0	0.0	0.0
Comprehension–Digit Symbol	0.0	0.0	0.4	0.4	2.6	2.6	7.0	9.2	11.4	11.8	12.3	10.5	9.2	8.3	4.4	3.9	5.2
Comprehension–Picture Completion	0.0	0.0	2.2	2.2	6.1	3.9	7.9	9.2	10.8	14.4	13.1	6.6	7.9	4.4	5.2	2.6	4.4
Comprehension–Block Design	1.2	1.3	0.9	1.7	2.6	3.9	13.5	10.5	11.8	11.4	15.3	7.9	6.1	3.9	3.9	1.7	2.1
Comprehension–Picture Arrangement	0.8	0.0	3.1	0.9	3.5	8.8	6.6	11.0	11.9	8.8	10.6	7.9	10.6	7.9	3.5	0.9	3.0
Comprehension–Object Assembly	0.0	0.4	1.7	3.5	4.4	5.2	8.7	7.0	11.4	12.2	10.9	10.5	7.0	5.7	4.4	2.2	4.8
Arithmetic–Similarities	1.7	0.4	2.2	2.2	5.7	7.5	10.1	14.5	18.4	14.9	9.6	6.6	1.8	1.3	1.3	1.8	0.0
Arithmetic–Digit Span	0.8	0.0	0.9	4.4	6.6	12.7	13.6	18.9	11.4	14.0	7.5	4.4	3.1	0.4	0.9	0.0	0.4
Arithmetic–Vocabulary	0.9	1.3	3.1	6.6	5.8	9.7	15.9	12.4	14.6	14.2	6.6	4.0	2.7	1.3	0.9	0.0	0.0
Arithmetic–Digit Symbol	0.0	0.0	0.9	0.9	2.2	1.3	6.6	9.3	15.5	17.7	13.3	10.2	7.5	5.3	4.4	0.4	4.4
Arithmetic–Picture Completion	0.4	0.0	1.8	2.2	3.1	7.5	10.6	11.0	12.8	13.2	12.8	7.0	4.8	4.8	3.5	3.5	0.9
Arithmetic–Block Design	0.0	0.0	1.3	4.4	2.2	5.3	11.9	11.9	18.5	17.6	10.6	4.8	5.7	2.6	2.2	0.4	0.4
Arithmetic–Picture Arrangement	0.8	0.4	2.7	3.5	4.0	4.0	7.5	11.1	15.9	16.8	10.6	8.4	4.0	4.0	2.7	1.8	1.8
Arithmetic–Object Assembly	0.4	0.9	1.3	3.1	4.4	4.8	7.5	12.3	15.4	10.6	13.2	11.5	4.4	4.0	1.3	0.9	3.8
Similarity–Digit Span	0.8	1.7	3.5	4.8	5.2	6.1	11.3	12.2	15.2	13.9	10.9	6.5	4.8	1.3	0.9	0.9	0.0
Similarity–Vocabulary	0.0	0.0	2.6	4.8	6.1	7.9	17.1	13.6	14.5	14.5	10.1	3.9	4.4	0.4	0.0	0.0	0.0
Similarity–Digit Symbol	0.0	0.9	0.0	0.9	1.8	4.4	3.9	7.0	12.3	18.0	13.6	9.2	8.8	6.1	6.1	1.8	5.1
Similarity–Picture Completion	0.4	0.0	1.3	1.7	4.8	5.7	6.1	8.7	14.4	13.5	14.4	12.2	4.8	3.5	3.9	2.2	2.1
Similarity–Block Design	0.9	0.4	2.6	1.3	0.4	7.9	8.7	15.3	17.0	10.5	7.9	10.5	7.4	3.9	3.1	0.9	1.3
Similarity–Picture Arrangement	0.8	0.9	2.2	1.8	4.0	4.8	7.9	9.7	12.3	14.5	12.8	7.9	8.4	5.3	3.5	2.2	0.8
Similarity–Object Assembly	0.4	1.3	1.3	1.3	5.7	4.8	5.7	7.4	17.5	11.8	12.2	11.4	4.8	6.1	3.9	0.9	3.4
Digit Span–Vocabulary	0.8	0.4	1.8	3.5	5.3	9.6	13.2	13.2	13.2	11.0	11.4	6.1	3.5	3.9	1.8	0.9	0.4
Digit Span–Digit Symbol	0.0	0.0	0.4	0.4	1.8	3.1	5.7	5.3	11.4	13.2	9.2	15.8	9.6	9.2	6.1	3.9	4.9
Digit Span–Picture Completion	0.0	0.9	1.3	0.4	3.5	5.2	6.6	8.7	10.9	13.1	13.5	10.5	8.7	7.0	4.4	2.6	2.6
Digit Span–Block Design	0.4	0.0	1.7	1.3	2.2	6.1	7.4	8.3	10.5	18.8	13.5	14.4	6.1	4.8	1.3	1.3	1.8

(table continues on page 158)

Table 3 (*continued*)

Difference between	Point difference																
	−8+	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8+
Digit Span–Picture Arrangement	1.3	0.4	2.2	0.9	1.8	5.3	4.0	8.8	14.5	11.5	15.9	11.0	8.4	5.7	3.5	2.6	2.1
Digit Span–Object Assembly	0.0	0.0	1.3	2.6	3.9	6.1	4.4	7.4	10.9	12.2	17.0	10.0	7.0	5.7	4.8	3.1	3.4
Vocabulary–Digit Symbol	0.0	0.0	0.4	0.9	1.8	3.1	8.0	6.2	9.7	9.3	12.4	11.9	8.0	6.6	7.1	5.8	8.7
Vocabulary–Picture Completion	0.0	0.0	0.9	3.1	2.2	6.2	7.9	9.7	10.1	8.8	9.7	15.4	4.8	6.6	5.3	3.5	5.7
Vocabulary–Block Design	0.0	0.9	0.9	2.2	3.5	4.4	4.8	15.4	13.2	12.3	7.5	10.6	7.9	6.6	4.8	2.6	2.2
Vocabulary–Picture Arrangement	0.9	0.4	0.4	4.4	3.1	3.6	7.6	10.2	10.7	9.6	8.9	8.4	12.0	6.7	4.9	4.0	4.0
Vocabulary–Object Assembly	0.0	0.9	0.4	4.0	3.1	5.7	8.4	4.8	10.6	13.7	9.3	8.8	8.4	5.3	7.0	4.4	5.2
Digit Symbol–Picture Completion	0.8	2.2	3.1	4.8	4.8	10.5	15.4	15.8	11.4	11.8	10.5	3.1	1.8	2.2	1.8	0.0	0.0
Digit Symbol–Block Design	1.7	1.8	3.9	3.9	8.8	10.5	13.6	18.0	16.7	8.8	6.6	2.2	2.6	0.4	0.4	0.0	0.0
Digit Symbol–Picture Arrangement	2.5	1.3	3.5	3.5	5.3	12.4	11.5	16.8	15.9	11.9	4.9	4.9	3.1	0.4	1.8	0.0	0.0
Digit Symbol–Object Assembly	1.3	1.3	1.3	6.6	9.2	7.5	10.1	15.8	17.1	10.5	10.1	4.8	3.1	0.4	0.9	0.0	0.0
Picture Completion–Block Design	0.9	0.4	0.9	2.2	7.9	9.6	13.5	11.4	17.9	13.5	10.9	3.9	4.8	0.0	1.7	0.4	0.0
Picture Completion–Picture Arrangement	0.4	0.4	3.1	3.1	1.8	7.9	12.3	15.0	17.6	13.7	10.6	7.0	4.4	1.3	0.4	0.4	0.4
Picture Completion–Object Assembly	0.0	1.3	0.9	1.7	6.1	4.4	15.3	10.9	16.2	12.2	12.7	10.0	5.2	2.6	0.4	0.0	0.0
Block Design–Picture Arrangement	0.0	1.3	1.8	0.4	3.5	4.8	10.1	15.0	17.6	16.3	13.7	6.2	4.0	1.3	3.5	0.0	0.4
Block Design–Object Assembly	0.0	0.0	0.0	0.0	3.9	5.7	9.2	12.2	21.8	16.2	14.8	7.9	3.5	3.9	0.0	0.4	0.4
Picture Arrangement–Object Assembly	0.0	0.4	1.3	1.3	5.7	4.0	12.8	14.5	18.5	12.3	12.3	6.6	4.0	3.1	0.4	0.9	1.7

(1957), who found similar results with the WAIS, recommended that any obtained difference be about 2.5 times the relevant error of measurement before being judged significant. Based on these data, such differences would need to range from 3 to 5 points, values much higher than those recommended in the WAIS–R manual.

Table 3 presents the actual percentages of subjects in this sample manifesting a particular difference between subscales for each of the 55 possible combinations. A negative difference indicates that the score on the second scale is larger than the score on the first scale, whereas a positive difference reveals the first scale's score to be the larger of the two. Although this sample is too small to serve as a normative reference, these values are included to provide readers with some empirical guide for applying confidence levels appropriate for their situations. Furthermore, the degree to which several of these distributions are not normally defined undermines the utility of the summary statistics presented in Table 2. With regard to the above example, fully 25% of our sample had a difference of 5 or more points between their Information and Object Assembly scores. It is also interesting to note that Information tended to be the higher scale score. Thus not only is the magnitude of a difference important, but the direction of that difference may also hold clinical import because some outcomes appear more likely than others.

Therefore, merely finding a statistically significant difference between subscales may *not* be indicative of anything clinically meaningful. Piedmont, Sokolove, and Fleming (in press-a) have shown that some significant differences may not be diagnostically revealing and that in other circumstances differences need to be even larger than those at the .05 threshold to be of clinical value. Although more systematic research linking subscale scores with nosological considerations is necessary, the data presented here can serve as an important guide in the making of such inferences from WAIS–R protocols.

It should also be pointed out that this sample is extremely heterogeneous in terms of age, ability level, and diagnosis. Such heterogeneity may be masking individual patterns between various groups. To more fully explore the utility of the WAIS–R with clinical populations, future research should use samples representative of more homogeneous populations. Another issue concerns the role of age. Although we did not use age-corrected scores in this study, it would be of interest to determine the impact, if any, that such adjustments would have on the presented distributions. Given that such a correction would more than likely affect the scores of those over age 50 and that only 10% of our sample fell into that category, this study cannot provide any clear insights into this important issue.

In summary, the WAIS–R remains internally consistent with a psychiatric sample. However, such psychometric integrity does not always readily translate into the difference scores between subscales; much of any such difference is due to chance. As these data indicate, clinicians need to be conservative in judging differences as being due to nonchance factors. On the basis of these findings, one should be skeptical of the cutoff values presented in the WAIS–R manual because they appear too small to be of any diagnostic or clinical value. Rather, the σ_{DD}

values and the actual distributional patterns may provide a more efficacious guide for making such interpretations.

References

Allison, J., Blatt, S., & Zimet, C. N. (1988). *The interpretation of psychological tests.* New York: Hemisphere.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

McNemar, Q. (1957). On WAIS difference scores. *Journal of Consulting Psychology, 21,* 239–240.

Piedmont, R. L., Sokolove, R. L., & Fleming, M. Z. (in press-a). An examination of some diagnostic strategies involving the Wechsler intelligence scales. *Psychological Assessment: A Journal of Consulting and Clinical Psychology.*

Piedmont, R. L., Sokolove, R. L., & Fleming, M. Z. (in press-b). Discriminating psychotic and affective disorders using the WAIS–R. *Journal of Personality Assessment.*

Rapaport, D., Gill, M. M., & Schafer, R. (1968). *Diagnostic psychological testing* (rev. ed.). New York: International Universities Press.

Wechsler, D. (1980). *Manual for the Wechsler Adult Intelligence Scale–Revised.* New York: Psychological Corporation.

Widiger, T. A. (1982). Psychological tests and the borderline diagnosis. *Journal of Personality Assessment, 46,* 227–238.

Received January 6, 1989
Revision received March 8, 1989
Accepted March 22, 1989 ■