Psychological Assessment:
A Journal of Consulting and Clinical Psychology
1989, Vol. 1, No. 3, 181-185

1040-3590/89/$00.75

An Examination of Some Diagnostic Strategies Involving the Wechsler Intelligence Scales

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

This report examines a number of Rapaport, Gill, and Schafer's (1968) diagnostic hypotheses, including differences between specific Wechsler subtests and clinical presentation, the magnitude of such a difference necessary for clinical significance, and overall indices of performance variability. *Ss* included 422 psychiatric patients comprising 3 diagnostic groups (affectives, psychotics, and schizophrenics) plus 19 normal controls. All *Ss* had either a Wechsler Adult Intelligence Scale (WAIS) or WAIS-R protocol. Only 2 of Rapaport et al.'s 12 hypotheses showed significant effects, and only 1 of these effects was moderated by the magnitude of difference between scales. When these 2 decision rules are translated into applied terms, the increase in diagnostic efficacy is low. There was no evidence for the utility of the overall scatter indices.

Since their introduction, the Wechsler scales have formed a cornerstone of psychological testing. The value of their contribution has been evidenced in the assessment of both intellectual capabilities and personality dynamics. It was the work of Rapaport, Gill, and Schafer (1968) that pioneered the application of the Wechsler scales in this latter area. Their work was influential in shaping clinical assessment, and their diagnostic hypotheses, although originally intended for the Wechsler-Bellevue scales, have been readily extended to later revisions of the instrument (Granick, 1963; McMullen & Rogers, 1984; Wiens, Matarazzo, & Gaver, 1959). However, much of this research has been limited to linking a few hypotheses with a diagnostically narrow sample, and not always successfully (Griffith, Estes, & Zerof, 1962; Kliajic, 1984; Wentworth-Rohr & Macintosh, 1972). The purpose of this article is to examine a wide range of Rapaport et al.'s diagnostic suggestions in a nosologically heterogeneous sample. Our aim is to evaluate these strategies empirically, documenting their strengths and weaknesses, in an attempt to refine their application and enhance their efficacy.

Working from the assumption that any performance index bears the characteristic features of a person's make-up, Rapaport et al. (1968) developed psychological rationales believed to underlie performance on each of the Wechsler subscales. Of particular clinical interest is the comparison of performance scores on various subscales; scores standing in a certain relationship to one another are believed to have diagnostic value (see Allison, Blatt, & Zimet, 1988, for many good examples). Because such intersubscale scatter plays a pivotal role in diagnostic classification, we selected a representative sample of 12 such hypotheses to evaluate. For example, Information > Comprehension is considered indicative of psychosis; as such, we should expect more psychotics and schizophrenics evidencing such a pattern than either affectives or normals. These hypotheses and their diagnostic implications are presented in Table 1.

These hypotheses merely state that one scale has a higher score than another, but a second issue that emerges concerns how much of a difference is necessary for clinical significance. Traditionally a 3-point difference between two scales was seen as significant (Matarazzo, 1972; Wechsler, 1955), although Rapaport et al. (1968) saw significance in from as few as 2 points to as many as 6 points. However, there has been criticism that such general rules of thumb are fallacious because they are not based on any distribution of difference scores one would expect to find for any given pair of scales (Jones, 1956). In fact, McNemar (1957) presented the distribution characteristics for each of the 55 possible difference scores among the 11 subtests of the Wechsler Adult Intelligence Scale (WAIS). He noted that the standard deviations for such difference scores ranged from 1.8 to 3.6 ($M = 2.88$). Piedmont, Sokolove, and Fleming (1989) noted similar difference-score variability as well as low reliabilities for such differences in the revised WAIS (WAIS-R). Clearly, a 3-point threshold may not always be sufficiently extreme to be diagnostically useful. In an attempt to determine clinically meaningful differences, we evaluated the above hypotheses in terms of 3, 4, 5, 6, 7, and 8-point differences between scales.

The final area of interest concerns the overall amount of intertest scatter or variability in the test protocol. Rapaport et al. (1968) assumed that increases in variability in performance are associated with increased psychopathology. Although such procedures have generated much controversy and little empirical documentation, they continue to be used in clinical interpretations (Allison et al., 1988). Three methods for assessing overall scatter are included in this study. The first method (referred to as scatter) calculates intertest scatter as the variance of an

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. Special thanks are extended to Paul T. Costa and Robert R. McCrae for many insightful comments.

Correspondence concerning this article should be sent to Ralph L. Piedmont, who is now at the National Institute on Aging, National Institutes of Health, Gerontology Research Center, Francis Scott Key Medical Center, 4940 Eastern Avenue, Baltimore, Maryland 21224.

Table 1
Specific Intersubscale Hypotheses and Predicted Experimental Findings

Subscale comparison	Diagnostic indication	Relative incidence
1. Information > Comprehension	Psychosis	P, S > A, N
2. Comprehension > Information	Hysteria	A > P, S, N
3. Vocabulary > Digit Span	Anxiety, Depression	A > S, N
4. Vocabulary > Arithmetic	Schizophrenia, Depression	A, S > N
5. Vocabulary > Digit Symbol	Depression, Anxiety	A > P, S, N
6. Vocabulary > Picture Arrangement	Depressive Psychosis	P > A, S, N
7. Arithmetic > Digit Span	Psychosis	P, S > A, N
8. Picture Completion > Vocabulary	Paranoia	P > A, S, N
9. Similarities > Vocabulary	Paranoia	P > A, S, N
10. Vocabulary > Comprehension	Psychosis, Schizophrenia	P, S > A, N
11. VIQ > PIQ	Depression, Neurological	A > P, S, N
12. PIQ > VIQ	Affective trend	A > P, S, N

Note. Diagnostic groups: A = affectives, P = psychotics, S = schizophrenics, N = normal controls. VIQ = Verbal IQ, PIQ = Performance IQ.

individual's distribution of subtest scores. The second method (referred to as range) is the difference between a person's highest and lowest subscale scores. The final method (referred to as verbal scatter) evaluates the subscales relative to the Vocabulary score. This is done by subtracting the Vocabulary score from each subscale and then computing the variance of the resulting 10 difference scores. We determined whether these scatter indices varied over different nosological groups.

Method

Subjects

Subjects consisted of 422 psychiatric patients from the Massachusetts Mental Health Center and the Boston University Medical Center plus 19 nonpsychiatric controls. There were 257 men and 184 women ranging in age from 15 to 85 years ($M = 29.95$, $SD = 11.35$), with an average of 12.7 years of education ($SD = 2.9$). These individuals received either the WAIS ($n = 290$) or the WAIS–R ($n = 151$) as part of their psychological evaluation at these two institutions or as part of a graduate testing course. Both WAIS and WAIS–R protocols were aggregated in this study because no distinctions between these measures have been made vis-à-vis applying Rapaport et al.'s (1968) hypotheses. The average Verbal IQ (VIQ) for this sample was 100.76 ($SD = 18.13$), and the mean Performance IQ (PIQ) was 92.1 ($SD = 16.55$). Ethnicity was as follows: 69.3% Caucasian, 2.9% Hispanic, 1.3% Asian, 24.6% Black, and 1.9% unreported. At the time of testing, 47% of the subjects were inpatients, 41% were outpatients, and the remaining 12% represented both the normal controls and those with an unknown disposition.

Diagnosis

During the course of their psychiatric treatment, the clinical subjects received a diagnosis either from their supervising clinician or from a psychologist evaluating their psychological data (e.g., Minnesota Multiphasic Personality Inventory, Rorschach Inkblot Test). The WAIS protocols were not involved in determining diagnoses. These diagnoses either were formed after treatment was well under way or were discharge diagnoses, which may have enhanced their reliability, although given the retrospective nature of the data set, no specific psychometric evaluations were possible. Where no explicit diagnosis was made, a subject was retained in our sample only if sufficient background information was present in his or her file to make such a determination. In such instances (approximately 10% of the cases), the relevant information was reviewed by two of the authors (R.L.P. and R.L.S.) and a consensual diagnosis was reached. In other circumstances, individuals who had a *DSM-II* (second edition of the *Diagnostic and Statistical Manual of Mental Disorders*) diagnosis were retained in our sample only if a distinct, comparative *DSM-III* (third edition) diagnosis could be made (see Appendix C of *DSM-III*, American Psychiatric Association, 1980). Although there are limitations inherent in using these types of diagnoses, the value of this approach has been documented elsewhere (Morey, Blashfield, Webb, & Jewell, 1988; Piedmont, Sokolove, & Fleming, in press-a, in press-b).

In the final sample, four diagnostic groups were created. Group 1 consisted of 140 individuals having a *DSM-III* Axis 1 diagnosis of an affective disorder, including bipolar disorder (296.4 to 296.6), major depression (296.2 and 296.3), cyclothymic (301.13) and dysthymic (300.40) disorders, and atypical depression (296.70 and 296.82). Those having an affective disorder with psychotic features were not included in this category. Group 2 consisted of 65 individuals whose *DSM-III* diagnosis was in the psychotic range (297.10, 297.30, 297.90, 298.80, and 298.90), with the exception of the schizophrenic disorders (individuals having affective disorders with psychotic features—296.X4—were also included in this category). The 217 people in Group 3 all had a diagnosis of schizophrenia, including schizoaffective and schizophreniform disorders (295.10 to 295.95). Finally, Group 4 consisted of the 19 nonpsychiatric controls.

Results

Intersubscale Comparisons

We used regression analyses to evaluate each hypothesis. A difference score was created by subtracting scores on the second scale on each hypothesis from scores on the first (with the exception of Hypothesis 2, Comprehension > Information, because only its twin hypothesis, Information > Comprehension, was used). This score directly reflects the magnitude of difference between scales, and such discrepancies should be related to diagnostic grouping. The independent variable was diagnostic category, which was effects coded (see Cohen & Cohen, 1975, pp. 188–195). The resulting three variables were entered simultaneously into the regression equation. Given the unequal *n*s per diagnostic group, we saw regression analyses as the most

appropriate and robust method for analyzing the data. As Cohen and Cohen (1975) noted, such cell inequities have no effect on the resulting regression coefficients.

Of the 12 regression analyses performed, significant effects were found only in 2. The first was Information > Comprehension (and, conversely, Comprehension > Information), $R = .23$, $F(3, 421) = 8.1$, $p < .0001$. An examination of the regression coefficients showed both schizophrenics ($t = 4.26$, $p < .0001$) and psychotics ($t = 2.19$, $p < .05$) to have higher Information scores relative to Comprehension scores, whereas normals had higher Comprehension scores relative to Information scores ($t = -5.9$, $p < .0001$). The second effect was for Vocabulary > Comprehension, $R = .23$, $F(3, 421) = 7.55$, $p < .001$. The regression coefficients showed that schizophrenics had significantly higher Vocabulary scores relative to Comprehension scores ($t = 4.56$, $p < .0001$), whereas normals had significantly higher Comprehension scores relative to Vocabulary scores ($t = -5.88$, $p < .0001$).

Systematically for each of the three hypotheses (Information > Comprehension, Comprehension > Information, Vocabulary > Comprehension) we selected subjects whose scores differed in the hypothesized direction from 3 or more to 8 or more points. Chi-square analyses were then performed to determine whether certain diagnostic groups were overrepresented in the subsamples in order to evaluate various cutoff values. Because of multiple tests for each hypothesis, we calculated a Bonferroni correction to control for familywise error rates; the alpha level per analysis was .0083. The results are presented in Table 2.

As can be seen in Table 2, the Information > Comprehension hypothesis appears to have no discriminative power, regardless of the magnitude of scatter. A significant effect emerges with the Comprehension > Information hypothesis. Consistent with predictions, there is an underrepresentation of psychotics and schizophrenics and an overrepresentation of affectives. However, there is also an overrepresentation of normals, which is inconsistent with predictions. Rather than being an index of a hysterical trend, this hypothesis may be a more general indicator of psychological deterioration. Finally, Vocabulary > Comprehension also presents mixed results. Consistent with predictions there is a lower incidence of normals and affectives but an overrepresentation only of schizophrenics; psychotics do not show the anticipated increase. Thus this hypothesis may be specific only to a schizophrenic process. Furthermore, maximal discriminative power is found when there is a 5-point difference between the subscales.

Intersubtest Scatter

Intersubtest scatter was calculated via the three different methods mentioned above. Separate regression analyses were performed for each scatter variable. The results revealed no significant effects for scatter ($R = .08$, *ns*), range ($R = .12$, *ns*), or verbal scatter ($R = .12$, *ns*).

Discussion

Overall, the results of this study do not offer encouraging support for a number of Rapaport et al.'s (1968) hypotheses relating Wechsler intertest scatter and clinical presentation. Before discussing these results in some detail, we need to mention a few caveats. First, Rapaport et al.'s diagnostic schemata were originally based on the Wechsler-Bellevue scales, whereas later revisions of this instrument (i.e., the WAIS and the WAIS–R) were the major data-collecting measures in this study. Although Rapaport et al.'s hypotheses have been readily extended to these later scales (e.g., Allison et al., 1988), it may be possible that differences in content and style between the original and revised scales may be responsible, in part, for the predicted relationships' not being supported by the data.

Furthermore, Rapaport et al.'s original hypotheses were formulated on 217 patients who represented both in- and outpatient cases at the elite Menninger Clinic and Institute. The majority of our subjects came from community mental health centers and probably differ sharply on demographic characteristics from the original sample. However, if this is a limiting factor, then one must question the general applicability of these hypotheses. Moreover, our cases represented the more chronic and deteriorated diagnostic ranges, and there were few of the less severe, more neurotic disorders such as the obsessive-compulsive, hysterical, and generalized anxiety disorders. Thus hypotheses that make predictions about these disorders cannot be evaluated with as much certainty as those concerning the more severe pathology groups.

Nonetheless, that only 2 of the 12 hypotheses found some support in our data brings into question the validity and clinical value of these hypotheses. The theoretical formulations that underlie them appear more compelling than the data that support them. At best, rather than reflecting specific psychopathological processes, some of these hypotheses may be better used as general indicators of pathological processes (e.g., psychotic–nonpsychotic). This appears to be the case with the Comprehension > Information and Vocabulary > Comprehension hypotheses. In the former case we find an overrepresentation of nonpsychotics, whereas the opposite pattern is found with the latter case. However, given the lack of significant findings in this study, this suggestion is tentative and in need of further empirical support.

Two hypotheses tested in this report clearly stand out as those most used in diagnostic prediction: Information > Comprehension indicating schizophrenia and Vocabulary > Digit Symbol indicating depression. The fact that the former hypothesis was supported and specifically predicted schizophrenia and not psychoses in general is worthy of note. The Comprehension subtest is generally recognized as a test of social judgment. Impaired social judgment is noted in *DSM-III* and the revised *DSM-III* as a particular attribute of schizophrenia and may reflect aspects of prefrontal lobe dysfunction putatively ascribed to patients with this disorder (Weinberger, 1987).

Similarly, it is equally worthy of note that the Digit Symbol hypothesis was not confirmed. The clinical thinking from the time of Rapaport et al. (1968) on has suggested that lower Digit Symbol scores reflect the psychomotor concomitance of clinical depression. Negative findings may imply either the broad motoric sequalae of many psychiatric disorders or the heterogeneity of the concept of clinical depression itself.

For those two hypotheses that found some support, two issues emerge. The first concerns the magnitude of difference between

Table 2
Chi-Square Analyses Examining the Diagnostic Utility of Specific Intertest Comparisons and the Magnitude of Such Scatter

Diagnostic group	Scatter					
	3 pt	4 pt	5 pt	6 pt	7 pt	8 pt
	Information > Comprehension					
Affectives	28.0	20.6	18.6	18.5	13.3	—
Psychotics	15.0	19.1	16.3	22.2	20.0	—
Schizophrenics	56.0	58.8	62.8	59.3	66.7	—
Normals	1.0	1.5	2.3	0.0	0.0	—
$\chi^2(3)$	4.93	7.20	4.89	4.50	3.58	—
N	100	68	43	27	15	—
	Comprehension > Information					
Affectives	41.5	44.1	50.0	—	—	—
Psychotics	13.8	11.8	0.0	—	—	—
Schizophrenics	32.4	29.4	25.0	—	—	—
Normals	12.3	14.7	25.0	—	—	—
$\chi^2(3)$	18.09*	14.18*	22.67*	—	—	—
N	65	34	16	—	—	—
	Vocabulary > Comprehension					
Affectives	30.6	26.5	16.2	17.6	18.2	—
Psychotics	15.3	13.2	8.1	5.9	0.0	—
Schizophrenics	54.1	60.3	75.7	76.5	81.8	—
Normals	0.0	0.0	0.0	0.0	0.0	—
$\chi^2(3)$	7.19	6.30	11.76*	5.48	5.24	—
N	111	68	37	17	11	—

Note. Diagnostic groups: Affectives, 31.7% of sample; psychotics, 14.8% of sample; schizophrenics, 49.2% of sample; normals, 4.3% of sample.
* $p < .0083$ (= .05/6; Bonferroni correction).

subscales. Only with Vocabulary > Comprehension does this appear to be an issue: A 5-point difference is necessary to capture diagnostic differences. Thus, the 3-point rule may not always provide the optimal cutting point. Therefore, in examining intersubtest differences, one should be familiar with the actual distributional characteristics of such difference scores. Such information is available for both the WAIS and the WAIS–R (McNemar, 1957; Piedmont et al., 1989). The second issue concerns the clinical applications of these decision rules. Given these results, it is clear that they should not be taken as necessary and sufficient diagnostic indicators. For example, take the hypothesis that Comprehension > Information is indicative of an affective disorder. As predicted, with a 3-point difference, there was an overrepresentation of affectives in this sample. However, only 27 of 140 subjects with an affective disorder (19%) manifested this pattern. Thus, one may have the disorder but not manifest the hypothesized difference. Conversely, that there were schizophrenics and normals evidencing such a difference indicates that one may manifest the difference pattern but not have the disorder. Clearly, this low level of sensitivity underscores the need for clinicians to buttress insights garnered in this manner with other relevant information.

After 40 years, Rapaport et al.'s (1968) hypotheses continue to influence clinical interpretations of test data. The results presented here bring into question the value of this practice. This is not to say that the Wechsler scales cannot provide insights into psychological functioning. Rather, we believe that more empirical methods need to be used in developing and validating such decision rules. For example, using discriminant function analyses, we have been successful in developing valid algorithms for discerning among affective and psychotic states as well as between various clusters of personality disorders (Piedmont et al., in press-a, in press-b). Given the many complexities of psychological functioning, these more sophisticated statistical techniques and methods may be what is needed to provide predictive accuracy.

References

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

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

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

Granick, D. (1963). Comparative analysis of psychotic depressives with matched normals on some untimed verbal intelligence tests. *Journal of Consulting Psychology, 27,* 439–443.

Griffith, R. M., Estes, B. W., & Zerof, S. A. (1962). Intellectual impair-

ment in schizophrenia. *Journal of Consulting Psychology, 26,* 336–339.

Jones, H. G. (1956). The evaluation of the significance of differences between scale scores on the WAIS: The perpetuation of a fallacy. *Journal of Consulting Psychology, 20,* 319–320.

Kliajic, A. N. (1984). The predictive utility of a significantly lower WAIS PIQ with psychiatric patients. *Journal of Clinical Psychology, 40,* 571–576.

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

McMullen, L. M., & Rogers, D. L. (1984). WAIS characteristics of nonpathological obsessive and hysteric styles. *Journal of Clinical Psychology, 40,* 577–579.

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

Morey, L. C., Blashfield, R. K., Webb, W. W., & Jewell, J. (1988). MMPI scales for DSM-III personality disorders: A preliminary validation study. *Journal of Clinical Psychology, 44,* 47–50.

Piedmont, R. L., Sokolove, R. L., & Fleming, M. Z. (1989). On WAIS-R difference scores in a psychiatric sample. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2,* 155–159.

Piedmont, R. L., Sokolove, R. L., & Fleming, M. Z. (in press-a). The classification of personality disorders: A comparison of three approaches. *Journal of Personality Assessment.*

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. (1955). *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corporation.

Weinberger, D. R. (1987). Implications of normal brain development for the pathogenesis in schizophrenia. *Archives of General Psychiatry, 44,* 160–169.

Wentworth-Rohr, I., & Macintosh, R. (1972). Psychodiagnosis with WAIS intrasubtest scatter of scores. *Journal of Clinical Psychology, 28,* p. 68.

Wiens, A. N., Matarazzo, J. D., & Gaver, K. D. (1959). Performance and verbal IQ in a group of sociopaths. *Journal of Clinical Psychology, 15,* 191–193.

Received January 9, 1989
Revision received April 26, 1989
Accepted May 3, 1989 ■