The Rorschach and Structured Tests of Perception as Indices of Intellectual Development in Mentally Retarded and Nonretarded Children

JOHN R. WEISZ
Cornell University

DONALD M. QUINLAN
Yale University

PATRICK O’NEILL AND PAMELA C. O’NEILL
Acadia University

In a study of the developmental significance of certain perceptual activities, the Rorschach and four structured tests of perception were administered to five groups of children at various CA, MA, and IQ levels. All three Rorschach measures were significantly related to MA, as were all four of the structured tests. The Children’s Embedded Figures Test was the best predictor of MA: it accounted for 52% of the variance, while the six remaining measures yielded nonsignificant increments to a multiple regression equation. In a factor analysis the four structured tests loaded on one factor (51% of total variance), while the three Rorschach variables loaded on a second (17%); MA loaded on both (.669 and .447, respectively). In an additional finding, Zigler’s “developmental” hypothesis that level of development and not IQ determines cognitive competence, received partial support from 13 of 14 statistical tests.

What characteristics of the person can be reliably and validly measured by the Rorschach test? This question has concerned experimental psychologists for more than two decades and has been the focus of literally thousands of studies (See Achenbach, 1974). With but a few exceptions (See Suinn & Oskamp, 1969), research findings have generally failed to support either the reliability or the validity of personality descriptions based on Rorschach responding. Reviewers of the Rorschach literature (e.g., Suinn & Oskamp, 1969; Zubin, Eron, & Schumer, 1965) have commented on the irony that an instrument with so little empirical support

The authors thank Mr. David Erichsen, Mrs. Kate McGraw, and principals and teachers of the Ansonia, Connecticut school system for their generous assistance, and Thomas M. Achenbach, Susan Harter, and William Kessen for their thoughtful critiques of an earlier draft of this paper. Requests for reprints should be addressed to John R. Weisz, Department of Human Development and Family Studies, Cornell University, Ithaca, NY 14853.
might continue to stimulate such a massive outpouring of empirical literature.

As the usefulness of the Rorschach in generating broad personality descriptions has been called into question, some investigators have turned to the prediction of more specific, carefully circumscribed aspects of the person. Of particular interest to developmental psychologists have been efforts to predict level of intellectual development from Rorschach responses. Scoring systems have been developed for both Form Accuracy (Mayman & Holt—as cited in Holt, Note 1, and O’Neill, O’Neill, & Quinlan, 1976) and Response Complexity (O’Neill et al., 1976), two aspects of Rorschach responding which appear to improve with maturity (cf. Hemminger, 1960). In addition, building upon Werner’s (1940) developmental theory, Friedman (1952) has devised a system which scores the accuracy and location of responses in the light of an hypothesized developmental sequence moving from poorly defined whole blot responses to articulation of details to integration of well-defined details. The three scoring systems are detailed in Table 1.

Despite the assertion by Goldfried, Stricker, and Weiner (1971) that, “Friedman’s scoring of the Rorschach results in a good measure of the developmental level of functioning,” the evidence in support of Rorschach developmental scoring systems has actually been quite limited (O’Neill et al., 1976). There has been an undersupply of evidence on interjudge reliability of these measures and no evidence on their validity as predictors of the most rigorous index of general intellectual level, i.e., psychometric mental age (MA). Furthermore, it has been unclear whether these Rorschach measures reflect any aspects of cognitive development that are not accounted for by simpler, more structured, tests of perception.

To provide such evidence, the present investigation and a companion study (O’Neill et al., 1976) were undertaken, using the same population of 60 children. In the O’Neill et al. report, reliability and validity evidence were presented. Correlations between the ratings of two scorers working independently were .93 for Form Accuracy, .96 for Complexity, and .90 for Developmental Level (DL). All three Rorschach measures showed significant \( p < .01 \) effects of MA.

With the reliability and validity of these Rorschach measures thus supported, the objective of the present study was to confront the general issue of whether the measures reflect any aspects of cognitive development that are not reflected in more structured perceptual tests. This general issue was explored by an inquiry into two specific questions: (a) In predicting intellectual level (i.e., MA), do the Rorschach measures add significantly to the predictive power of structured perceptual tests?, and (b) Does the factor structure underlying intercorrelations among the Rorschach and structured measures suggest that the two types of measures tap overlapping or separate sources of variance?
### TABLE 1
**SCORING WEIGHT AND CATEGORY DESCRIPTIONS FOR FORM LEVEL, DEVELOPMENTAL LEVEL, AND COMPLEXITY**

<table>
<thead>
<tr>
<th>Weight</th>
<th>Form Level (Mayman &amp; Holt)</th>
<th>Developmental Level (Friedman)</th>
<th>Complexity (O'Neil et al.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Successful combination of imagination and reality congruence; realistic responses that are not popular or near popular.</td>
<td>Wholes on broken blots (II, III, VII, VIII, X) or details combining discrete blot areas with good form level (W+, D+).</td>
<td>Combinations; two or more conceptually independent objects, integrated in a meaningful relation.</td>
</tr>
<tr>
<td>6</td>
<td>Popular and near-popular responses.</td>
<td>Wholes on broken blots (II, III, VII, VIII, X) or details combining discrete blot areas with good form level (W+, D+).</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Unconvincing form, which is not arbitrary; general shape of blot does not clash with response.</td>
<td>Wholes on broken blots (II, III, VII, VIII, X) or details combining discrete blot areas with good form level (W+, D+).</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Vague and amorphous responses, where the response implies so little that goodness-of-fit is virtually irrelevant.</td>
<td>Vague wholes (Wv).</td>
<td>Responses containing a single major object (even if in plural form) with two or more features articulated.</td>
</tr>
<tr>
<td>3</td>
<td>Weak form, some clash between the contour of the blot and the response.</td>
<td>Details which are vague, poor form, amorphous, or confabulations.</td>
<td>Single-object responses in which no feature beyond gross shape is articulated.</td>
</tr>
<tr>
<td>2</td>
<td>Spoiled responses; some important oversight or striking distortion spoils an otherwise good response.</td>
<td>Details which are vague, poor form, amorphous, or confabulations.</td>
<td>Single-object responses in which no feature beyond gross shape is articulated.</td>
</tr>
<tr>
<td>1</td>
<td>Wholly arbitrary responses.</td>
<td>Wholes which are poor form, amorphous, or confabulations, Fabulized combinations and contaminations.</td>
<td>Vague and amorphous responses.</td>
</tr>
</tbody>
</table>

*Note: Assignment of weights is the work of the present authors, not that of Mayman, Holt, or Friedman. Across the three columns, categories assigned the same numerical weight should not be regarded as corresponding or equivalent.*

The particular structured tests selected were designed to gauge the perceptual activities of "differentiation" and "integration," thought by many Rorschach developmentalists (e.g., Friedman, 1952) to reflect the ontogenetic processes of the same name which were central to Werner's (1940) developmental theory. Included were three structured tests designed by their authors to measure perceptual integration and one test designed by its authors to measure perceptual differentiation. The latter test, the Children's Embedded Figures Test (CEFT; Witkin, Oltman, Raskin, & Karp, 1971) was the structured measure of greatest interest, since earlier evidence has demonstrated that Witkin's tests of field-dependence-independence have developmental properties (cf. Witkin et
al., 1971) and relate significantly to both the form accuracy and the complexity of Rorschach responses (Gardner, Holzman, Klein, Linton, & Spence, 1959; Lipton, Kaden, & Phillips, 1958; Weiss, Stein, Atar, & Melnik, 1967; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962). Without the present study, however, the evidence is only suggestive in that it is drawn from studies using Witkin measures other than the CEFT, relying on adult populations, and not systematically varying level of intellectual development.

Because of the use of structured perceptual tests, particularly the CEFT, the selection of a psychometric instrument for assessing MA and IQ required care. To have assessed MA and IQ by means of the Stanford-Binet or the Wechsler Intelligence Scale for Children (WISC), both involving subtests that are in part structured tests of perception, might have introduced a bias in favor of the MA predictive superiority of the structured perceptual tests over the Rorschach measures. In fact, Witkin and his colleagues (e.g., Witkin et al., 1971) acknowledge a link between field-dependence-independence and certain performance subtests as are found in the Binet and WISC, but they deny such a link with the verbal component of intelligence. For these reasons, the Peabody Picture Vocabulary Test (PPVT; Dunn, 1965) which is heavily verbal in character, and which is correlated from .82 to .86 with 1960 Stanford-Binet MAs (Dunn, 1965), was used to assess MA and IQ in the present study.

Earlier developmental investigations of Rorschach responding (e.g., Hemmingdinger, 1953, 1960) have focused on relationships with CA. The present study was designed to tease apart the effects of CA and MA; thus these two variables were the logical factors in the experimental design. Because of the unique interrelation among the variables CA, MA, and IQ, a design which treats any two as factors must necessarily tolerate the third as a partially confounded variable. However, the present study demonstrates that it is not necessary to ignore effects due to the third variable. In fact, a probe for IQ effects formed a second general purpose of this study—i.e., to evaluate Zigler's (1969) "developmental" hypothesis that intellectual level, or MA, determines an individual's competence at formal cognitive processes, regardless of the individual's IQ. The opposing "difference" theorists insist that IQ will exercise an independent effect on cognitive activity. Prior to the present study the extensive research generated by this theoretical controversy (See Zeaman & House, 1967; Zigler, 1966; Zigler, 1969) has centered on learning and conceptual processes, leaving perception virtually unexplored.

METHOD

Subjects

The four older groups shown in Table 2 formed the principal experimental design of the study, a $2 \times 2$ (MA9 and 12 vs. CA9 and 12)
design in which MA and CA were orthogonal. The youngest group (CA6-MA6) together with the CA9-MA9 and CA12-MA12 groups, formed a second, ancillary design—overlapping partially with the first—with three levels of intellectual maturity; in this design, to be used for only one analysis on each dependent variable, equating MA and CA permitted an examination of 'pure developmental effects,' with IQ held constant. The CA9-MA12 group contained five boys and seven girls, all other groups six boys and six girls. CA12-MA9 contained two black children, CA9-MA9 one black child; all other subjects were white. All attended school in a lower middle-class community.

**Tasks in the First Session**

A female experimenter unaware of the purposes of the study, individually administered the PPVT and the Rorschach according to standardized procedures (see O’Neill et al., 1976).

**Tasks in the Second Session**

In a second session, about 4 weeks later, a male experimenter unaware of subjects’ performance levels in the first session, administered four perceptual tests. One of these, the CEFT, is correlated from .70 to .86 with the adult Embedded Figures Test (Witkin et al., 1971), which according to its originators, requires ‘the ‘breaking up’ of an organized field in order to
separate out a part of it (p. 5)." The test requires the subject to locate a specific two-dimensional figure (a tent or a house shape) within embedding contexts composed of designs in a variety of color schemes and degrees of complexity.

Three measures were employed to tap perceptual synthesis, or integration. The Gestalt Completion Test (Street, 1931) consists of 15 incomplete, achromatic pictures which are to be identified by the subject, presumably through the use of "figural closure." The Closure Speed Test was developed by Thurstone and Jeffrey as a measure of Thurstone's (1944) "first closure factor," defined as "the capacity to construct a whole picture from incomplete or limited material (Kantor, 1966, p. 1)." The measure consists of 27 incomplete, achromatic pictures. These were exposed one at a time by the experimenter. The subject was allowed 3 seconds to examine the picture, asked if he had an idea what the picture was, then given 3 more seconds. In Gollin's (1960) test for Visual Recognition of Incomplete Objects, materials consist of simple line drawings of 23 everyday objects and familiar animals. Each item is drawn five times in increasing degrees of completeness, so that card one in each series showed only a few portions of the figure's outline while card five showed a complete drawing. Each series formed a flip chart, with each drawing shown for 3 seconds. All subjects recognized all 23 objects at or before card five. A subject's score was the mean number of cards required for recognition. To control for order effects, half of the male and half of the female subjects in each experimental group received the CEFT first, followed by the three tests of perceptual integration. The remainder received the CEFT last.

RESULTS

Initial analyses revealed no significant sex or order effects on any of the measures.

MA and CA Effects

To assess independent CA and MA effects a $2 \times 2$ (CA level $\times$ MA level) ANOVA was performed on all seven dependent measures, using the four groups of the CA–MA orthogonal design ($r_{CA \times MA} = -.05$). In this analysis all Rorschach variables and perceptual measures improved significantly with MA ($F > 8.0, df = 1/44, p < .01$, for all seven analyses; see means in Table 2). The CA effect, however, only approached significance with

---

1 It should be noted that reviews referring to the tests of "perceptual synthesis, or integration" used in the present study indicate that most available evidence supports the validity of these measures, but not all investigators agree that the three tests tap perceptual synthesis or integration. (See, e.g., Gollin, 1965; Guilford, 1967; Kantor, 1966; and Wohlwill, 1960.)
respect to Rorschach Complexity, where scores decreased with CA, \( F(1,44) = 7.49, p < .01 \). Table 3 (above diagonal) reveals that, within the CA–MA orthogonal design, all Rorschach and perceptual measures were significantly correlated with MA, while their correlations with CA were low and nonsignificant in all but one instance. An apparently independent MA effect can be enhanced by the positive correlation between MA and IQ which is inevitable in any CA–MA orthogonal group structure (expected \( r \) in such a 2 × 2 design = .71; in the present design \( r_{MA\times IQ} = .75, p < .001 \). For this reason it is appropriate to test Zigler’s hypothesis that if groups are matched for MA there will be no cognitive differences attributable to IQ. The CA9-MA9 and CA12-MA9 groups were compared, as were the CA9-MA12 and CA12-MA12 groups, with respect to the three Rorschach and four structured measures. Although all six comparisons on the Rorschach measures indicated better performance by higher IQ groups, only one of the total of 14 t tests approached significance; this test compared the two MA12 groups on Complexity, \( t(22) = 2.54, p < .05 \).

With the relative influence of CA and MA thus established, single classification ANOVAs were calculated for each of the seven variables, using the three groups which formed the ‘developmental’ design. The developmental effect was significant at the .05 level on Rorschach

### Table 3

**Intercorrelations Among All Measures**

<table>
<thead>
<tr>
<th></th>
<th>CEFT</th>
<th>Gestalt completion</th>
<th>Closure speed</th>
<th>Visual recognition***</th>
<th>Rorschach D.L.</th>
<th>Rorschach form accuracy</th>
<th>Rorschach complexity</th>
<th>Mental age</th>
<th>Chronological age</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEFT</td>
<td></td>
<td>.52***</td>
<td>.64***</td>
<td>-.35</td>
<td>.54***</td>
<td>.49***</td>
<td>.30***</td>
<td>.72***</td>
<td>.03***</td>
</tr>
<tr>
<td>Gestalt completion</td>
<td>.56***</td>
<td></td>
<td>.45**</td>
<td>-.44**</td>
<td>.24</td>
<td>.24</td>
<td>.31**</td>
<td>.42**</td>
<td>-.21**</td>
</tr>
<tr>
<td>Closure speed</td>
<td>.69***</td>
<td>.59***</td>
<td>- .01***</td>
<td>-.61***</td>
<td>.35**</td>
<td>.39***</td>
<td>.33**</td>
<td>.59***</td>
<td>.05***</td>
</tr>
<tr>
<td>Visual recognition***</td>
<td></td>
<td>-.57***</td>
<td>-.65***</td>
<td>-.76***</td>
<td>-.17</td>
<td>-.26</td>
<td>-.13</td>
<td>-.41**</td>
<td>-.10**</td>
</tr>
<tr>
<td>Rorschach D.L.</td>
<td>.60***</td>
<td>.38*</td>
<td>.49**</td>
<td>-.49**</td>
<td>.89***</td>
<td>.48***</td>
<td>.45**</td>
<td>.17</td>
<td>-.17***</td>
</tr>
<tr>
<td>Rorschach form accuracy</td>
<td>.70***</td>
<td>.47**</td>
<td>.53***</td>
<td>-.54***</td>
<td>.90***</td>
<td>.42**</td>
<td>.49***</td>
<td>-.18**</td>
<td>-.36**</td>
</tr>
<tr>
<td>Rorschach complexity</td>
<td>.30</td>
<td>.40*</td>
<td>.52**</td>
<td>-.51**</td>
<td>.53***</td>
<td>.41*</td>
<td>.32*</td>
<td>-.36*</td>
<td></td>
</tr>
<tr>
<td>Mental age</td>
<td>.83***</td>
<td>.56***</td>
<td>.79***</td>
<td>-.72***</td>
<td>.60***</td>
<td>.72***</td>
<td>.43**</td>
<td>-.05*</td>
<td></td>
</tr>
<tr>
<td>Chronological age</td>
<td>.84**</td>
<td>.54***</td>
<td>.80***</td>
<td>-.74***</td>
<td>.63***</td>
<td>.73***</td>
<td>.43**</td>
<td>.95***</td>
<td></td>
</tr>
</tbody>
</table>

*Note. Correlations above the diagonal are based on four groups (\( N = 48 \)) in which MA and CA are orthogonal. Correlations below the diagonal are based on three groups (\( N = 36 \)) in which IQ is constant and MA and CA are completely confounded.

* = \( p < .05 \).
** = \( p < .01 \).
*** = \( p < .001 \).
**** = Lower score indicates better performance.
Complexity, $F(2,33) = 4.3$, and at the .005 level on the remaining variables; all three variables increased with development, $F(2,33) > 7.0$.

**Perceptual Tests and Rorschach Variables as Predictors of MA**

**Correlational comparisons.** To test the relative accuracy of Rorschach and structured measures in assessing intellectual level, the sizeable correlations of DL and Form Accuracy with MA were compared to the correlations of the four perceptual measures with MA, using the Pearson $r$s shown above the diagonal in Table 3. The CEFT was more highly correlated with MA than were DL or Form Accuracy, $t(45) = 3.05$ and 2.31, $p < .005$ and .05, respectively (cf. McNemar, 1962, p. 140), while Closure Speed scores were more strongly related to MA than was DL, at a borderline level, $t(45) = 1.35$, $p < .10$, two-tailed. The correlations of Gestalt Completion and Visual Recognition with MA did not differ significantly from the DL-MA correlation, or the Form Accuracy-MA correlation.

**Multiple regression and partial correlations.** Again using data from the CA-MA orthogonal design, a stepwise multiple regression of MA on the four structured tests was calculated. These four measures combined accounted for 56% of the variance in MA, $R(4,43) = .748$, $p < .001$. A similar analysis, including the four perceptual tests plus the three Rorschach measures as predictors yielded a nonsignificant ($F[3,40] = 1.1$) improvement in predictability, accounting for only an additional 3% of the MA variance, $R(7,40) = .769$, $p < .001$. In both analyses CEFT entered the regression equation first and accounted for 52% of the variance ($p < .001$), while the remaining variables yielded nonsignificant individual increments. When the relationship between MA and CEFT was partialed out, no other variable showed a significant relationship with MA (all partial $r$s < .10). On the other hand, when the Rorschach variables most significantly related to MA (Form Accuracy and DL) were partialed out, a significant relationship between MA and CEFT remained, partial $r$s = .64 and .72, respectively, $p < .001$.

**Factor analysis.** Again using the MA-CA orthogonal design, a principal components factor analysis was performed on the matrix of intercorrelations among the three Rorschach indices, the four perceptual measures, and MA. The two principal factors having eigenvalues greater than 1 were subjected to a varimax rotation. All four of the structured measures and MA had their highest loadings (.699 to .807) on the first factor, accounting for 51% of the total variance (highest Rorschach loading on Factor 1 = .199). All three of the Rorschach variables loaded highest (.635 to .936) on the second factor, accounting for 17% of the total variance. MA was correlated .669 with the first factor and .447 with the second.
DISCUSSION

These findings indicate that, within the age range studied, scores on the three Rorschach scoring systems are correlated with intellectual level as defined by MA; but this relationship appears to be eclipsed by the strong relationship between intellectual level and perceptual ability as assessed through structured tests, particularly the CEFT. Controlling for CEFT performance reduced the correlation between Rorschach measures and MA to nonsignificance, while partialing out Rorschach scores had little effect on the CEFT–MA correlation; this suggests that the perceptual activity measured by the CEFT is the more central correlate of intellectual level. One specific question this study was designed to answer was whether the Rorschach developmental measures add significantly to the MA-predictive power of structured perceptual tests. The answer, based on multiple regression analysis, is that the structured tests as a group give a strong accounting of developmental effects, and that MA predictability is not significantly enhanced by the Rorschach measures.

A second specific question addressed in this study was whether the Rorschach and structured measures tap overlapping or separate sources of variance. The answer, based on the factor analysis reported above, is that the three Rorschach measures and the four structured tests appear to tap two quite independent sources of variance. It should be noted that the factor structure which emerged from the analysis was apparently unrelated to the differentiation–integration dichotomy which Rorschach developmentalists (e.g., Friedman, 1952) have held to be important. It should also be noted that the ‘‘Rorschach factor’’ and the ‘‘structured test factor,’’ while unrelated to one another, were both highly correlated with MA. This suggests that the two types of measures may gauge two different aspects of cognitive development, and, consequently, that it would be premature to conclude that the Rorschach developmental measures add so little that they need not be preserved. Instead, these findings suggest the need to determine precisely what aspect(s) of intellectual development are reflected in these Rorschach developmental measures that structured tests do not tap.

The strong relationship found between CEFT performance and Rorschach responding, as well as MA, are in harmony with several other findings (cited earlier) based on adult samples and other Witkin measures. The failure to find CEFT sex differences, while discrepant with adult findings on field-dependence, is generally consistent with results obtained using children (cf. Witkin et al., 1971, pp. 5, 24). Finally, in 13 of 14 statistical tests there were no significant differences between groups matched on MA, but differing in IQ. This finding provides new evidence extending support for Zigler’s ‘‘developmental’’ theory into perceptual
domains. However, it should be noted that, with the small sample sizes employed, the power of these tests would only permit the detection of relatively large group differences. Thus, the support for the Zigler hypothesis generated by the present study should be regarded as modest.

REFERENCES


**REFERENCE NOTE**


**RECEIVED:** January 8, 1977; **REVISED:** March 29, 1977.