JOURNAL OF PERSONALITY ASSESSMENT, 1993, *60*(1), 153–180

The Impact of Response Frequency on the Rorschach Constellation Indices and on Their Validity With Diagnostic and MMPI–2 Criteria

Gregory J. Meyer
Department of Psychiatry
University of Chicago Hospitals

I first examined the effects of response frequency (*R*) on the Comprehensive System's constellation indices (Exner, 1991). *R* is significantly associated with 26 of the 29 constellation criteria that are based on raw numbers and significantly correlated with total scores on each of the constellations. I then examined how *R* impacts the external validity of the constellations. The ability of the Schizophrenia Index and the Suicide Constellation to discriminate diagnostic groups appears to be impaired when protocols deviate from average length, whereas the Hypervigilance Index (*HVI*) appears most diagnostic of a paranoid condition when it is relatively elevated in brief records. *R* also clearly mediates the relationship between the Rorschach and Minnesota Multiphasic Personality Inventory–2 (MMPI–2) criteria. In lengthy protocols, many of the cross-method convergent correlations between appropriate MMPI–2 scales and the Depression Index, Obsessive Style Index, and *HVI* are quite high and approach the maximum found in personality research. In brief protocols, there are strong negative correlations between these constellations and self-reports of depression and interpersonal distress. Implications from these findings for the integration of assessment methods are discussed.

Exner's (1986a, 1991) Comprehensive System has fostered significant advances in the clinical utility of the Rorschach as a nomothetic tool. The most recent achievement in this nomothetic evolution has been the development of scoring constellations to identify suicide potential (*S-Con*), depression (*DEPI*), hypervigilant scanning (*HVI*), schizophrenia (*SCZI*), social and emotional coping deficits (*CDI*), and obsessive cognitive processing (*OBS*). Each of these variable constellations were empirically derived to predict certain clinical phenomena and to assist in differential diagnosis.

Outside of the initial studies, there have only been a handful of efforts to cross-validate the constellations (Archer & Gordon, 1988; Ball, Archer, Gordon, & French, 1991; Exner, 1986b; Lipovsky, Finch, & Belter, 1989; Shapiro, Leifer, Martone, & Kassem, 1990; Spigelman & Spigelman, 1991; Weiner, 1992). Despite the experimental sophistication that went into their development, strong evidence for their external validity has not yet been demonstrated.

Patients diagnosed with schizophrenia have been found to score higher on the *SCZI* than patients with a diagnosis of conduct disorder, dysthymia, or borderline personality disorder (Archer & Gordon, 1988; Exner, 1986b). However, the *SCZI* has not discriminated patients with schizophrenia from patients with major depression or a mixed assortment of personality disorders (Archer & Gordon, 1988) and has not been able to differentiate patients with schizophrenia from patients with a schizotypal personality disorder (Exner, 1986b). In addition, the *SCZI* has not correlated ($r = .11$) with the Schizophrenia scale (Scale 8) from the Minnesota Multiphasic Personality Inventory (MMPI; Archer & Gordon, 1988).

The *DEPI* has been higher in sexually abused girls (Shapiro et al., 1990) and in children from divorced families (Spigelman & Spigelman, 1991) than in control samples. However, the *DEPI* has not demonstrated a significant association to inpatient diagnoses of depression or dysthymia (Ball et al., 1991); has not been higher in patients with major depression when compared to patients with schizophrenia, dysthymia, personality disorder, or conduct disorder (Archer & Gordon, 1988); and has not been higher in inpatients with a depressive condition (i.e., major depression, dysthymia, or adjustment disorder with depressed mood) when compared to patients without a depressive condition (Lipovsky et al., 1989).

The *DEPI* has also displayed mixed evidence of association with self-reports or parental reports of depression. Lipovsky et al. (1989) found the index was not correlated with the Children's Depression Inventory, although it was correlated with the Depression scale (Scale 2) from the MMPI ($r = .29$). However, the latter finding was contradicted by Archer and Gordon (1988) who found that the *DEPI* was not correlated with Scale 2 ($r = -.06$). Ball et al. (1991) found the *DEPI* was not related to parental ratings of depression in their sample of children, and Shapiro et al. (1990) found no relationship between the *DEPI* and the Children's Depression Inventory or the Internalizing scale from the Child Behavior Checklist.

There are several limitations in the current literature on the validity of the constellations. First, to date, studies have only examined the *DEPI* and *SCZI*. The *S-Con*, *CDI*, *HVI*, and *OBS* have not yet been subject to cross-validation, and these indices may find greater support. Second, the studies published prior to 1991 did not use the most recent versions of the *DEPI* and *SCZI*. This may account for some of the negative findings, although even the revised *DEPI* has

not fared better than the original (Ball et al., 1991). Third, even though the constellations were developed for use with adults, most validation studies have used samples of children or adolescents. Even with these limitations in mind, however, the evidence so far does not provide compelling support for the utility of the *DEPI* or the *SCZI* as specific diagnostic indicators.

The nomothetic advances that have been made with the Comprehensive System have placed an increasing primacy on the raw scores that quantify the data obtained from the test. Although the test has been scored and used in a quasi-nomothetic fashion since its inception, it has historically been used in a more idiographic fashion wherein comparison of scores to norms has been less emphasized than the sequential interplay of formal scores and content within an individual record. Currently, the raw numbers, ratios, and percentages obtained from the Rorschach are interpreted from a rather strict normative perspective (Exner, 1991), and precise scores determine which criteria on any of the constellations are positive. Because of this emphasis, it is now more necessary to ensure that formal scores are rigorous and valid representations of their underlying psychological constructs. Factors that cause the frequency of scores to vary in ways that are unrelated to their underlying constructs may compromise the nomothetic use of the Rorschach, as well as the diagnostic accuracy of the constellations.

I initiated this study to determine if response frequency (*R*) is a factor that clouds the diagnostic accuracy of the constellations. One of the most consistent findings in Rorschach research is that many scores are highly correlated with *R* (e.g., Fiske & Baughman, 1953; Meyer, 1992a). Further, the most consistent finding from factor analytic research is that *R* defines the largest factor within the data of the Rorschach (Meyer, 1992a, 1992b). This *R* factor is also defined by most of the other scores from the test when they are expressed in terms of their frequency of occurrence (e.g., number of determinants within a determinant category, number of detail location areas, number of contents, number of special scores, number of occasions of organizational activity, etc.). This factor indicates that the raw frequencies for most scores vary in tandem with *R*. A protocol with relatively few responses will generally have relatively few determinants, detail location areas, efforts at organization, special scores, and so on, whereas a protocol with many responses will have a higher frequency of scores across the scoring categories.

Because of this, I hypothesized (Meyer, 1992a) that *R* would be related to each of the 29 constellation criteria that are based on raw numbers. However, I expected that *R* would be unrelated to the 14 criteria that are based exclusively on ratios or percentages because these scores already control for *R*. Based on expectations for each of the specific criteria, I also predicted that the *SCZI*, *DEPI*, *HVI*, and *OBS* would be positive significantly more often in lengthy protocols than in brief records. This was expected because each of these constellations have a relatively high proportion of criteria that are based on the raw frequency

of scores (e.g., $M- > 1$, $FD > 2$, $Zf > 12$, etc.). Similarly, I expected the overall *CDI* to be positive more often in brief records than lengthy ones because four of the five criteria that comprise this index are based on scores occurring with a low frequency (e.g., $EA < 6$, $COP < 2$, etc.).

If these differences were evident, the next question I wanted to address was how *R* impacts the validity of the constellations. Currently, it is unclear what drives the correlation between *R* and other scores on the test. It may be that some underlying psychological characteristics cause both *R* and the frequency of certain scores to occur at a particular level. For example, a patient's degree of organizational strivings may dictate both a particular number of *Zf* scores within the test and a particular number of responses to the test as a whole. From this perspective, hypervigilant or obsessional tendencies could cause an increased frequency of scores to be observed on the *HVI* and *OBS* criteria and also result in an increased *R* to the inkblots. Conversely, internal coping deficits could validly result in there being a low frequency of scores on the *CDI* criteria (making the index more positive) and also result in an inability to muster many responses to the test.

However, there are instances when *R* fluctuates in response to examiner expectations or to the interpersonal climate that has been established between the patient and the tester (Exner, 1986a; Exner, Armbruster, & Mittman, 1978; Guilford, 1947, cited in Cronbach, 1949). When these factors, or other factors such as resistance or misunderstanding, are operative during the test, fluctuations in *R* may cause the level of many scores to vary in a fashion that makes them less precise indicators of the patient's underlying psychology. That is, if *R* varies because of extraneous factors, the formal relationship between *R* and score frequencies will result in many scores also varying for extraneous reasons. (The term *extraneous* is used here to refer to conditions that are not measured by the data of the Rorschach, and it does not necessarily imply conditions divorced from the patient's psychology.) For example, a compliant patient may have the misguided notion that the examiner wants many responses and will thus produce a 35- or 40-response protocol. Other patients may narcissistically or aggressively enjoy having the examiner in the role of a "scribe" and produce protocols of similar length. When patients produce long protocols for these reasons, they should also have higher scores on the *DEPI*, *SCZI*, *HVI*, and *OBS*, even though depressive, schizophrenic, hypervigilant, or obsessional processes may have nothing to do with the reasons for their elevated *R*. Similarly, patients who are malingering or afraid of the Rorschach for any of a variety of reasons may produce very brief records. These patients are prone to having high and potentially erroneous scores on the *CDI*. Because *R* can vary in response to these extraneous factors and result in potentially biased constellations, I hypothesized that constellations from average-length protocols would demonstrate the greatest external validity.

METHOD

Subjects

The subjects were 90 adult patients, of whom 54 (60%) were female and 70 (78%) were inpatients. The average age was 33.5 (range = 18 to 65 years). These subjects came from a larger pool of patients and were selected on the basis of how many responses they produced to the Rorschach in order to form three groups. Clinical samples tend to have an average *R* of about 22 or 23, with a standard deviation of approximately 6 or 7 (Exner, 1990). My initial goal was to select high-*R* and low-*R* groups that were 1 standard deviation or more from the mean. However, because protocols of less than 14 responses are readministered or discarded, it was impossible to select a sufficient number of subjects who were 1 full standard deviation below the mean. Therefore, criteria for the low-*R* group were relaxed slightly to include all patients with 14 to 17 responses ($n = 27$). The average-*R* group consisted of all patients with 21 to 24 responses ($n = 31$), and the high-*R* group consisted of all patients with 29 or more responses (range = 29 to 65, mode = 30, median = 35, M = 38.4; $n = 32$).

Measures

All Rorschachs were administered and scored according to the most recent guidelines for the Comprehensive System (Exner, 1990). Scoring reliability was computed by having four protocols from each of the three groups randomly selected and blindly rescored by a research assistant. Percentage of exact agreement was determined across the following categories: location and space (94%), developmental quality (88%), movement (90%), shading and achromatic color (94%), color (95%), form dominance of determinants (89%), color shading or shading blends (97%), all determinants (74%), form quality (82%), pairs (95%), content (85%), populars (97%), organizational activity scores (88%), special scores that form the weighted sum of six special scores (82%), and other special scores (88%).

Two main sources of validation criteria were utilized. The first were diagnoses made according to the most recent version of the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., rev. [*DSM-III-R*]; American Psychiatric Association, 1987). For inpatients, the criteria were discharge diagnoses determined by treatment teams. These teams were not blind to assessment information, but they incorporated this data along with information from life history and daily observation of the patient on the unit. For outpatients, the diagnoses selected for validation criteria were based on a comprehensive diagnostic interview by the referring clinician before testing was initiated. An independent or quasi-independent diagnosis was not available for 9 patients.

To establish validation diagnoses that were relevant to the Rorschach constellations, diagnoses were arranged into the following categories: (a) major depression (n = 33; 37% of sample); (b) depressive disorders, including major depression, bipolar disorder with depressed mood or mixed affective features, depressive disorder not otherwise specified (NOS), and dysthymia (n = 48; 53% of sample); (c) psychotic disorders, including schizophrenia, schizoaffective disorder, bipolar disorder with psychotic features, major depression with psychotic features, borderline personality disorder, delusional disorder, and psychotic disorder NOS (n = 39; 43% of sample); (d) hypervigilant disorders, including paranoid schizophrenia, paranoid personality disorder, and paranoid delusional disorder (n = 8; 9% of sample); and (e) obsessive–compulsive disorder or obsessive–compulsive personality disorder (n = 3; 3% of sample). There are no *DSM–III–R* diagnoses that correspond directly to the *CDI*. However, an approximation was formed by including patients with a diagnosis of major depression, dysthymia, depressive disorder NOS, any personality disorder, or a substance abuse disorder (n = 45; 50% of sample).

To establish diagnostic criteria that would have some, albeit limited, applicability to the *S-Con*, an 8-point suicide severity scale was developed to rate each patient. I used this scale to blindly rate admission notes from the medical record (for inpatients) or presenting information (for outpatients) for each subject. This scale was rated as follows: 0 = *no depressive symptoms at the time of admission/evaluation* (n = 16), 1 = *depressive symptoms without suicidal ideation* (n = 19), 2 = *suicidal ideation without intentionality* (n = 23), 3 = *an attention seeking suicidal "gesture"* (n = 10), 4 = *suicidal ideation with intentionality* (n = 10), 5 = *a depression/hopelessness based attempt without serious intentionality* (n = 5), 6 = *a serious attempt with last-minute efforts to reverse the action* (n = 5), and 7 = *a serious attempt with full intentionality and no efforts to reverse the action* (n = 2).

The second set of validation criteria were MMPI–2 scale scores. At the time the Rorschachs were administered, 81 patients also completed the MMPI–2. Criteria for the *DEPI* were the basic depression scale (Scale 2); the Harris and Lingoes subscale of Subjective Depression (*D1*; cited in Butcher, Dhalstrom, Graham, Tellegen, & Kaemmer, 1989); Social Introversion (which includes a strong element of dysphoria; Scale 0); and the content scales for Depression (*DEP*), Low Self-Esteem (*LSE*), Social Discomfort (*SOD*), and Work/Interference (*WRK*). The criteria for the *CDI* were essentially the same as for the *DEPI* and included Scale 2, Scale 0, *LSE*, *SOD*, *DEP*, and *WRK*. The *SCZI* criteria were Scale 8, two of the Scale 8 Harris and Lingoes subscales—Lack of Cognitive Ego Mastery (*Sc3*) and Bizarre Sensory Experiences (*Sc6*)—and the content scale for Bizarre Mentation (*BIZ*). Criteria for the *HVI* were Scale 6; the Harris and Lingoes subscale of Persecutory Ideas (*Pa1*); *BIZ*; and the content scales for Cynicism (*CYN*) and specific, phobic-like Fears (*FRS*). The first three of these scales focused on the more pathological elements of hypervigilance, whereas the latter two focused on potentially more normal variants. The *OBS* criteria

focused on symptoms of anxiety, depression, and social withdrawal and included the basic Anxiety scale (Scale 7), the Anxiety content scale (*ANX*), *FRS*, the content scales for Obsessiveness (MMPI–*OBS*), Scale 2, Scale 0, *DEP*, and *SOD*. The MMPI-2 does not have scales to assess suicidal ideation or suicide potential. Therefore, the *S-Con* criteria focused on symptoms of internal disarray and included Scale 2, Scale 7, *ANX*, *OBS*, *DEP*, *LSE*, and *WRK*.

RESULTS

Relation of *R* to Specific Criteria and the Overall Constellation Indices

To assess the relationship between *R* and the constellation criteria a series of chi-square tests of association were carried out. Because these analyses utilized ordinal-level data and had a relatively large total number of subjects, the Mantel–Haenszel chi-square statistic was utilized to test the significance of the observed relationship when specific predictions were made. The Pearson chi-square was used in other instances. I expected 34 of the 49 tests within the constellation indices to be significantly associated with *R*. Twenty-nine of the 34 predicted relations were for specific criteria, whereas 5 were for scores on the overall constellations. Table 1 demonstrates that 29 of the 34 hypotheses were supported. Of the 5 that were not, 1—the expectation that the overall OBS would be positive more frequently in high *R* protocols—could not be determined because no patient in this sample was positive on the *OBS*. Of the other 4, 1 criteria demonstrated a trend (*DEPI* 5, $p = .103$) in the expected direction, and 2 (*CDI* 3 and 4) were mixed criteria, such that at least one other element within the criteria was expected to be unrelated to *R*.

Within the *S-Con*, as expected, longer protocols are significantly more likely to have more than two vista or form-dimensional scores, a color-shading blend, more than three morbid scores, and more than three space responses. The finding that there was also a relationship between *R* and the X+% was unexpected. It occurred because only four patients in this sample had an X+% greater than .70, and all of them had brief records. As expected, brief protocols are significantly more likely to have less than two pure human scores and, almost by definition, less than 17 responses. As anticipated, the overall *S-Con* is not significantly related to *R*.

For the *SCZI*, results demonstrate that as *R* increases so do the opportunities for observing more than six special scores or a weighted sum of special scores greater than 17, more than one Level 2 *Fabcom* within the context of other Level 2 special scores, and more than one M−. There is also an unexpected trend for high *R* protocols to have an X+% below the cutoff on Criteria 1 of the *SCZI*. This is probably due to the fact that high *R* protocols use more unusual detail

TABLE 1
Constellation Criteria and Chi-Square Significance of Association With *R* Groups

Constellation and Criteria	*Significance*	*Constellation and Criteria*	*Significance*
Suicide Potential		Schizophrenia Index	
1. *FV + VF + V + FD* > 2	.004[a]	1. X+% < .61	
2. *Color-Shading Blend* > 0	.026[a]	and *S*−% < .41;	
3. *3r + (2)/R* < .31 or > .44	.612	or X+% < .50	.056
4. MOR > 3	.015[a]	2. X−% > .29	.477
5. *Zd* > +3.5 or < −3.5	.122	3. *FQ*− > or = *FQu*	
6. *es* > *EA*	.978	or FQ− > *FQo* + FQ+	.221
7. *CF + C* > *FC*	.991	4. *Level 2 Sp. Sc* > 1	
8. X+% < .70	.008	and *FAB2* > 0	.001[a]
9. *S* > 3	.000[a]	5. *Raw Sum6* > 6	
10. *P* < 3 *or* > 8	.770	or *WtSum6* > 17	.016[a]
11. *Pure H* < 2	.014[b]	6. M− > 1 or X−% > .40	.000[a]
12. *R* < 17	.000[b]	Positive on 4 or more	.181[a]
Positive on 8 or more	.977	Coping Deficit Index	
Depression Index		1. *EA* < 6 or *AdjD* < 0	.000[b]
1. *sum V* > 0 or *FD* > 2	.001[a]	2. COP < 2 and AG < 2	.000[b]
2. *C-Sh-Bl* > 0 or *S* > 2	.000[a]	3. *WSmC* < 2.5 or *Afr* < .46	.000[b]
3. *3r + (2)/R* > .44 and *r* = 0		4. *Passive* > *Active* + 1	
or *3r + (2)/R* < .33	.686	or *Pure H* < 2	.157[b]
4. *Afr* < .46 or *Blend* < 4	.000[b]	5. *Sum T* > 1	
5. *SumSh* > *FM + m*		or *Isolate/R* > .24	
or *SumC'* > 2	.103[a]	or *Food* > 0	.228[a]
6. MOR > 2		Positive on 4 or more	.000[b]
or *2xAB + Art + Ay* > 3	.000[a]	Obsessive Style Index	
7. COP < 2 or *Isolate/R* > .24	.016[b]	1. *Dd* > 3	.000[a]
Positive on 5 or more	.005[a]	2. *Zf* > 12	.000[a]
Hypervigilance Index		3. *Zd* > 3.0	.300
1. *Sum T* = 0	.663	4. *Populars* > 7	.025[a]
2. *Zf* > 12	.000[a]	5. FQ+ > 1	.040[a]
3. *Zd* > 3.5	.387	Positive by criteria	None[a]
4. *S* > 3	.000[a]		
5. *H + (H) + Hd + (Hd)* > 6	.000[a]		
6. *(H) + (A) + (Hd) + (Ad)* > 3	.000[a]		
7. *H + A:Hd + Ad* < 4:1	.069		
8. *Cg* > 3	.001[a]		
Positive on 1 and 4 others	.001[a]		

Note. *N* = 90.

[a]Criteria hypothesized to be positive more frequently in the high *R* group. [b]Criteria expected to be positive more frequently in the low *R* group.

locations, which also generally receive either an unusual or minus form quality score. Despite the fact that three of the six *SCZI* criteria are positive more frequently in high *R* protocols, the overall *SCZI* is not.

For the *DEPI*, longer protocols are significantly more likely to have a vista score or more than two form-dimensional scores, a color-shading blend or more than two space responses, and more than two morbid scores or an Intellectualization Index greater than 3. There is also a trend to observe more than two achromatic color scores. As expected, both Criteria 4 (partially defined by less than four blends) and Criteria 7 (partially defined by less than two cooperative movement scores) were positive significantly more often in low-*R* protocols. Despite these latter two criteria being negatively related to *R*, the overall *DEPI* is positive significantly more often in longer protocols as predicted.

For the *CDI*, brief records are significantly more likely to have a summed *Erlebnistypus* less than 6, less than two cooperative or aggressive movement scores, and a weighted sum of color scores less than 2.5. In addition, as was expected, low-*R* protocols are significantly more likely to be positive on the overall *CDI*.

Consistent with predictions, the overall *HVI* is positive significantly more often in longer protocols. This is because longer protocols are significantly more likely to have more than 12 instances of organizational activity, more than three space responses, more than six human or quasi-human percepts, more than three quasi-human and quasi-animal percepts, and more than three clothing scores. There is also an unanticipated trend for longer protocols to have a higher proportion of detail human and animal contents.

For the *OBS*, longer protocols are significantly more likely to have more than three unusual detail locations, more than 12 instances of organizational activity, more than seven popular responses, and more than one plus form quality response. However, because of the generally poor level of reality testing in this patient sample, no subject was positive on the overall *OBS*.

To further assess the relationship between *R* and the constellations, *R* was correlated with total scores from each index. I expected total scores for each constellation to be significantly correlated with *R* because each constellation was expected (and observed) to have more than one criteria associated with *R*. Table

TABLE 2
Correlations Between *R* and Each of the Constellation Total Scores Within the Full Sample

	S-Con	*SCZI*	*DEPI*	*CDI*	*HVI*	*OBS*
R	.25*	.26*	.30**	−.37***	.60***	.56***

Note. $N = 90$.
*$p < .05$. **$p < .01$. ***$p < .001$. All one-tailed tests.

2 indicates that *R* correlates significantly with the total scores on each of the constellations, although the magnitude of the correlations have considerable range.

Validity of the Constellations—Diagnostic Criteria

To assess the validity of the *DEPI*, *CDI*, and *SCZI* in relation to diagnostic criteria, I conducted *t* tests to determine if total scores differed significantly across patients with and without the target diagnosis (e.g., total scores on the *DEPI* across patients with and without a diagnosis of depression). These *t* tests were first conducted in the full sample and then within each level of *R* to determine if diagnostic discrimination varied as a function of *R*. I hypothesized that total scores would be significantly different across patient groups in the average-length protocols but not in the high- and low-*R* protocols.

DEPI total scores did not differ significantly between patients with and without a diagnosis of depression. This held when both narrow and broad diagnostic classifications were utilized, and this was the case within the full sample of patients and within each level of *R*. A similar lack of diagnostic discrimination occurred for the *CDI* in the total sample and each *R* group.

In the full sample of patients, *SCZI* total scores were significantly higher in patients with a psychotic disorder than in those without such a diagnosis. Examination of Table 3 indicates that this was due to the average-*R* group, for which scores on the *SCZI* significantly and strongly differentiated patient groups. In low- and high-*R* protocols, *SCZI* total scores were not higher in patients with a psychotic disorder. Thus, although the *SCZI* is effective overall, the power behind its discriminative ability comes from patients with average-length protocols. This can be seen most readily when total scores on the *SCZI* are correlated with a psychotic diagnosis. In the total sample, there is a low but significant correlation (Spearman $r = .22$, $p = .024$, one-tailed); in the low- and high-*R* protocols, there are nonsignificant correlations (Spearman *r*s = .13 and .07, respectively), but in the average-length protocols there is a robust correla-

TABLE 3
t Test of Mean *SCZI* Scores in Patients With and Without a Psychotic Disorder Diagnosis

SCZI	*no Dx*	*Dx*	*t*	*Probability*
Full sample	2.40[a]	3.18[b]	2.07	.042
Low *R*	2.17[c]	2.77[d]	.82	.422
Average *R*	1.73[e]	3.21[f]	2.95	.006
High *R*	3.27[g]	3.58[h]	.48	.638

Note. Probabilities are two-tailed tests.
[a]$N = 42$. [b]$N = 39$. [c]$n = 12$. [d]$n = 13$. [e]$n = 15$. [f]$n = 14$. [g]$n = 15$. [h]$n = 12$.

tion ($r = .48$, $p = .004$, one-tailed). These findings indicate that salient variations from an average *R* impair the effectiveness of the SCZI as a diagnostic tool.

The number of patients who made serious suicide attempts or who were given a hypervigilant or obsessive diagnosis were too limited to conduct *t*-test analyses within the *R* groups. However, tentative correlational analyses were performed to assess the *S-Con* and *HVI* in relation to diagnostic criteria. The Spearman rank-order correlation between total scores on the *S-Con* and a suicide attempt was .00 in the total sample, .09 in the low-*R* group, .21 ($p = .130$, one-tailed) in the average-*R* group, and $-.34$ ($p = .060$, two-tailed) in the high-*R* group. Although none of these correlations reach statistical significance, post hoc comparison of the difference between the correlations revealed that the correlation in the average-*R* group was significantly higher than the correlation in the high-*R* group ($z = 2.10$, $p = .036$, two-tailed). This is a very tentative finding, but it suggests that the *S-Con* is more predictive of suicide attempts in an average-*R* protocol than in a high-*R* protocol. This is probably because longer protocols have some *S-Con* criteria that are positive simply because there are more responses available to score.

Total scores on the *HVI* were not significantly correlated with a hypervigilant diagnosis in the total sample, the average-*R* protocols, or the high-*R* protocols (Spearman *r*s = .13, $-.10$, and .20, respectively). However, there was a robust correlation between total *HVI* scores and a hypervigilant diagnosis in the low-*R* protocols (Spearman $r = .57$, $p = .004$, two-tailed). This was an unanticipated finding that should again be considered quite tentative. It occurred because the few patients who had both a low-*R* protocol and a hypervigilant diagnosis scored relatively high on the *HVI*. There were no low-*R* protocols that were positive on the overall *HVI*, which is not surprising given the strong correlation between *HVI* scores and *R*. However, the robust correlation in the brief protocols suggests that it is quite diagnostic when a patient displays hypervigilant tendencies on the Rorschach despite giving very few responses.

To address the question of whether *R* may be low because of coping deficits or high because of hypervigilant or obsessional tendencies, *R* was correlated with each of the diagnostic categories used in this study. The only significant correlation occurred in conjunction with an obsessional diagnosis. However, the direction and magnitude of the correlation (Spearman $r = -.23$, $p = .038$, two-tailed) was opposite of what would be predicted. A trend was also observed to find *R* correlated with a coping deficit disorder (Spearman $r = .20$, $p = .074$); however, the direction was again opposite of what would be predicted. The other correlations between *R* and diagnostic categories were near 0 (e.g., with a hypervigilant diagnosis, Spearman $r = .03$). Thus, these data provide no support for the notion that *R* consistently varies as a function of diagnostic condition.

Validity of the Constellations–Self-Report Criteria

To assess the validity of the constellations in relation to the self-report MMPI–2 criteria, I performed two types of analyses. First, I conducted a series of correlations between constellation total scores and their appropriate MMPI–2 scales. These correlations were done with the total sample and for each level of *R*. As suggested by previous research with the *SCZI* and *DEPI*, I anticipated that correlations would be nonsignificant in the full sample of patients. However, I hypothesized that there would be positive and significant correlations between Rorschach scores and self-report measures in average-length protocols. The second set of statistics were one-way analyses of variance (ANOVAs) for only those patients who were positive on a constellation. The independent variable was *R* at three levels, whereas the dependent variables were the appropriate MMPI–2 scales. I expected that MMPI–2 scales would be highest in average-length protocols. Because there were an insufficient number of subjects positive on the *HVI* and the *OBS* in each of the *R* groups, only the *SCZI*, *DEPI*, *S-Con*, and *CDI* were examined in the one-way ANOVAs. Findings for each constellation are discussed in turn.

The *SCZI* was not correlated with MMPI–2 scales in the total sample and within each level of *R* (e.g., with Scale 8, *rs* = −.02, −.19, .05, −.00, in the total sample, and low-, average-, and high-*R* protocols, respectively). Further, there was no clear pattern to the correlations across the *R* groups. Finally, when examining just the patients who were positive on the *SCZI*, there were no significant differences in MMPI–2 scale scores across the *R* groups.

The *CDI* was negatively correlated with its MMPI–2 criteria in the full sample, although only the correlation with the *DEP* scale reached statistical significance ($r = -.22$, $p = .048$). Within each of the *R* groups, the *CDI* did not correlate with MMPI–2 scales. Analysis of MMPI–2 scores across the *R* groups for only those patients positive on the *CDI* revealed no significant differences.

The *DEPI* was not correlated with self-report measures in the full sample of subjects. However, within each level of *R*, there were dramatic differences in the nature of the *DEPI*'s relationship to MMPI–2 scales (see Table 4). In general, within the low-*R* group there were salient negative correlations between the *DEPI* and validation criteria, whereas in the high-*R* group there were significant positive correlations. To determine whether these correlations were significantly different from each other, they were first transformed to Fisher's *z* statistic, and then *z* scores were computed (Hays, 1981). Table 4 demonstrates that the nature and magnitude of the correlations for each of the MMPI–2 criteria are significantly different between the low- and high-*R* groups. *DEPI* scores contradict self-reports of depression in brief protocols (and to some extent, in average-length protocols), but in longer Rorschach records there is clear correspondence between the *DEPI* and appropriate MMPI–2 scales.

Furthermore, Table 5 demonstrates that the patients who were positive on

TABLE 4
Correlations Between the *DEPI* and MMPI–2 Validation Criteria for the Total Sample and Three Levels of *R*

Criteria	*Total*[a]	*Low* R[b]	*Average* R[c]	*High* R[d]	*z* score[e]
Scale *2*	−.08	−.40	−.24	.35	2.63**
D1	−.14	−.40	−.27	.37*	2.71**
DEP	.03	−.27	−.11	.40*	2.34*
LSE	.14	−.11	.00	.46*	2.03*
Scale *0*	−.07	−.48*	−.26	.38*	3.08**
SOD	−.11	−.44*	−.23	.19	2.22*
WRK	.08	−.12	−.22	.51**	2.28*

Note. MMPI–2 scales are as follows: Scale *2* = basic Depression scale; *D1* = Harris and Lingoes subscale of Subjective Depression; *DEP* = content scale for Depression; *LSE* = Low Self-Esteem; Scale *0* = Social Introversion; *SOD* = Social Discomfort; *WRK* = Work Interference.

[a]$N = 81$. [b]$n = 22$. [c]$n = 29$. [d]$n = 30$. [e]*z* score compares the difference between the correlations from the low- and high-*R* groups.

*$p < .05$. **$p < .01$. All two-tailed tests.

TABLE 5
ANOVAs of Mean MMPI–2 Scale Scores Across the *R* Groups for Those Patients Positive on the *DEPI*

Criteria	*Low* R[a]	*Average* R[b]	*High* R[c]	*F Ratio*	*Probability*
Scale *2*	18.0	27.1	30.3	7.30	.002
D1	6.6	12.7	16.2	5.81	.007
DEP	5.2	13.6	15.6	3.64	.038
Scale *0*	22.6	32.4	37.6	4.41	.020
WRK	7.4	11.3	16.0	3.23	.052
SOD	5.2	10.4	12.9	3.13	.057
LSE	4.2	9.0	10.1	2.70	.082

Note. MMPI–2 scales are as follows: Scale *2* = basic Depression scale; *D1* = Harris and Lingoes subscale of Subjective Depression; *DEP* = content scale for Depression; Scale *0* = Social Introversion; *WRK* = Work Interference; *SOD* = Social Discomfort; *LSE* = Low Self-Esteem.

[a]$n = 5$. [b]$n = 13$. [c]$n = 17$.

the *DEPI* had significantly different mean MMPI–2 scores across the *R* groups on scales *2*, *D1*, *DEP*, and *0*. For each scale there is a clear linear increase across the *R* groups. Those patients who are positive on the *DEPI* but have low-*R* protocols report very little depressive symptomatology. For these patients, almost all of the MMPI–2 raw score means translate into *T*-scores of less than 50. On the other hand, those patients who are positive on the *DEPI* but have high-*R* protocols report considerable depressive symptomatology. For these patients, the raw score means for Scale *2*, *D1*, and *DEP* all translate into *T*-scores that are greater than 70. It is worth noting that the same pattern of findings held under more stringent conditions when the *DEPI* was considered positive only after six of the seven criteria were met. Under these circumstances, the MMPI–2 scores were

generally more extreme, with the brief protocols having lower scores and the lengthy protocols having higher scores.

The *S-Con* was unrelated to MMPI–2 criteria within the full sample of patients. In the brief and average-length protocols the *S-Con* had negative but nonsignificant correlations with criteria, whereas in the lengthy protocols the *S-Con* was positively correlated with criteria. For example, the correlation between the *S-Con* and *DEP* was −.11 in the full sample, −.34 in the low-*R* group, −.31 in the average-*R* group, and .21 in the high-*R* group. However, the only correlations to reach statistical significance were those with Scale 7 and *ANX* in the high-*R* group (*r*s = .36 and .42, respectively, $p < .05$, two-tailed).

Despite the fact that there were very few patients who were positive on the *S-Con* in each of the *R* groups, which results in low statistical power, Table 6 indicates that the means for four of the seven MMPI–2 scales differ significantly across the *R* groups. The pattern is essentially the same as it was for the patients who were positive on the *DEPI*. In general, patients who have a positive *S-Con* but a brief record report very few symptoms of internal distress. These patients have raw scores on the MMPI–2 scales equivalent to *T*-scores of approximately 40. However, patients who have a positive *S-Con* and a lengthy record report considerable internal distress. These patients have raw scores on the MMPI–2 scales that are generally equivalent to *T*-scores greater than or equal to 65.

Correlational data for the *HVI* are presented in Table 7. In the full sample, *HVI* total scores are significantly correlated with Scale 6 of the MMPI–2. However, when the different *R* groups are examined, an interesting pattern again emerges. In the low-*R* group the *HVI* tends to be negatively correlated with criteria, whereas in the high-*R* group the *HVI* has a positive correlation with reports of paranoid symptomatology. Furthermore, it is clear that the significant correlation between the *HVI* and Scale 6 in the total sample is primarily the result of the robust correlation that occurs within the high-*R* protocols ($r = .50$).

TABLE 6
ANOVAs of Mean MMPI–2 Scale Scores Across the *R* Groups for Those Patients Positive on the *S-Con*

Criteria	*Low R*[a]	*Average R*[b]	*High R*[c]	*F Ratio*	*Probability*
ANX	.5	5.3	15.2	17.62	.003
Scale *7*	3.0	6.3	25.0	10.80	.010
DEP	1.0	3.3	12.5	6.88	.028
MMPI-OBS	.0	3.0	7.5	6.38	.033
WRK	1.5	4.7	15.5	4.82	.056
LSE	1.0	3.7	9.0	2.48	.164
Scale 2	16.0	22.7	27.3	2.41	.170

Note. MMPI–2 scales are as follows: *ANX* = content scale for Anxiety; Scale *7* = basic Anxiety scale; *DEP* = content scale for Depression; *MMPI-OBS* = Obsessive Ruminations; *WRK* = Work Interference; *LSE* = Low Self-Esteem; Scale *2* = basic Depression scale.

[a]*n* = 2. [b]*n* = 3. [c]*n* = 4.

TABLE 7
Correlations Between the *HVI* and Self-Report Criteria for the Total Sample and Three Levels of *R*

Criteria	*Total*[a]	*Low R*[b]	*Average R*[c]	*High R*[d]	*z score*[e]
Scale 6	.26*	−.02	.17	.50*	1.90
Pal	−.04	−.18	−.05	.31	1.68
BIZ	.09	−.11	.15	.25	1.22
FEARS	−.11	−.13	.13	−.01	0.40
CYN	−.14	−.15	.05	−.17	0.07
Discriminant Correlations					
Scale 2	.04	−.48*	.08	.14	2.22*
Scale 0	.08	−.52*	.08	.13	2.36*
DEP	.18	−.41	.27	.29	2.45*
SOD	.07	−.64**	.16	.10	2.87**

Note. MMPI–2 scales are as follows: Scale 6 = Paranoia; *Pal* = Harris and Lingoes subscale of Persecutory Ideas; *BIZ* = Bizarre Mentation; *FRS* = Fears; *CYN* = Cynicism; Scale 2 = basic Depression scale; Scale 0 = Social Introversion; *DEP* = content scale for Depression; *SOD* = Social Discomfort.

[a]$N = 81$. [b]$n = 22$. [c]$n = 29$. [d]$n = 30$. [e]z scores contrast the correlations found in low-*R* protocols to the correlations found in high-*R* protocols.

*$p < .05$. **$p < .01$. All two-tailed tests.

The discriminant correlations between the *HVI* and other MMPI–2 scales are also noteworthy. Although these associations were not the direct focus of this investigation, the *HVI* had an intriguing pattern of correlations with reports of depressive symptomatology and interpersonal discomfort. The lower portion of Table 7 demonstrates that *HVI* scores in low-*R* protocols have strong and generally significant negative correlations with reports of depression and social discomfort. In high-*R* protocols these correlations are positive but not statistically significant. Nonetheless, the magnitude of the correlations in the brief records are significantly different from those in the longer records.

The *OBS* correlational findings are presented in Table 8. Scale 7 and *ANX* are significantly correlated with the *OBS* in the total sample, but it is again clear that these relationships are primarily due to the strong positive correlations that are found within the high-*R* protocols. In general, the correlations between the *OBS* and criterion measures are negative in low-*R* protocols, but positive in the high-*R* protocols. The pattern of correlational changes across levels of *R* is robust, and this is somewhat more pronounced for depressive and interpersonal symptomatology than it is for anxiety symptoms. In brief protocols, *OBS* scores tend to be the opposite of self-reported depression and social discomfort, whereas in longer protocols, *OBS* scores correspond directly to self-reports of anxiety, depression, and social discomfort.

The final analysis undertaken was a direct correlation between *R* and the various MMPI–2 scales used in this study (in addition to *L*, *F*, and *K*). Only 2 of

TABLE 8
Correlations Between the *OBS* and Self-Report Criteria for the Total Sample and Three Levels of *R*

Criteria	*Total*[a]	*Low R*[b]	*Average R*[c]	*High R*[d]	*z score*[e]
Scale *7*	.22*	−.29	.25	.39*	2.02*
ANX	.32*	−.26	.19	.43*	2.42*
FEARS	−.05	−.08	.12	.18	0.90
MMPI–OBS	.17	−.16	.22	.18	1.17
Scale *2*	.13	−.49*	.16	.44*	3.38***
Scale *0*	.20	−.44*	.28	.36	2.86**
DEP	.17	−.50*	.31	.37*	3.17**
SOD	.15	−.51*	.36	.18	2.51*

Note. MMPI–2 scales are as follows: Scale 7 = basic Anxiety scale; *ANX* = content scale for Anxiety; *FRS* = Fears; *MMPI–OBS* = Obsessive Ruminations; Scale 2 = basic Depression scale; Scale 0 = Social Introversion; *DEP* = content scale for Depression; *SOD* = Social Discomfort.

[a]$N = 81$. [b]$n = 22$. [c]$n = 29$. [d]$n = 30$. [e]z scores contrast the correlations found in low-*R* protocols to the correlations found in high-*R* protocols.

*$p < .05$. **$p < .01$. ***$p < .001$. All two-tailed tests.

the 21 correlations were significant. *R* was correlated with Scale *0* ($r = .22$, $p = .046$) and *ANX* ($r = .32$, $p = .003$), suggesting that patients who describe themselves as anxiously withdrawn or introverted produce more responses to the test.

DISCUSSION

In this study, I investigated how *R* influences the Rorschach constellation indices. I also made a preliminary exploration of how *R* affects the validity of the constellations with external criteria. This was accomplished by selecting three groups from a general patient population who differed in the number of responses they produced to the Rorschach. The low-*R* group consisted of patients with 14 to 17 responses ($n = 27$), the average-*R* group consisted of patients with 21 to 24 responses ($n = 31$), and the high-*R* group consisted of patients with more than 29 responses (median = 35, $n = 32$).

Internal Validity

As hypothesized, *R* is significantly associated with virtually all of the frequency criteria within the constellations (26 of 29), but it is essentially unrelated to criteria that are based on ratios or percentages (13 of 14). Further, *R* is significantly related to whether the overall *DEPI*, *CDI*, and *HVI* are positive using dichotomous decision rules. Presumably this would also hold for the *OBS* in samples with more intact reality testing abilities. Finally, as expected, total

scores on each of the constellations are significantly correlated with *R*. More frequent responses are associated with higher scores on the *S-Con*, *SCZI*, *DEPI*, *HVI*, and OBS, whereas fewer responses are associated with higher scores on the *CDI*.

These findings raise the question of whether *R* should be expected to have such a strong impact on the constellations. On logical grounds, it would be difficult to articulate reasons why high *R* should be related to suicide potential and schizophrenia and depression and hypervigilance and obsessional tendencies, each of which are distinct clinical conditions. In a similar fashion, although there are reasons to suspect that one's level of internal resources (*EA*), organizational strivings (*Zf*), or psychological complexity (blends) could cause variations in *R*, it would be difficult to articulate reasons why high levels of distorted perceptions involving human activity (M−), disorganized or bizarre thinking operations (special scores), oppositionality (space responses), pessimism (morbid), painful introspection (vista, form-dimensional scores) or ambivalent affect states (color-shading blends), among others, should all cause frequent responses to be made to the inkblots.

Therefore, the data can be understood most parsimoniously from a perspective that assumes a significant part of the correlations between *R* and raw score criteria are driven by *R* itself. This is essentially a formal cause matter that operates because scores are assigned whenever a response occurs. Consequently, the number of responses available to score "causes" variations to occur in the frequency of observed scores. When many responses are available there will be an increased frequency of scores across scoring categories. When there are few responses, scores will have a lower frequency. It is this formal relationship between the occurrence of a response and the assignment of scores that appears most responsible for the findings that virtually all frequency criteria (e.g., M− scores, *S* scores, vista scores, etc.), and total scores on all of the constellations are significantly related to *R*.

The correlations between *R* and total scores on the *S-Con* and *SCZI* are statistically significant and meaningful, but they are not very dramatic ($rs = .25$ and .26, respectively). On the other hand, the correlations between *R* and total scores on the *HVI* and *OBS* are of substantial magnitude ($rs = .60$ and .56, respectively). It is tempting to speculate that hypervigilant and obsessional tendencies are causing more responses to be made to the test, resulting in these psychological characteristics working synergistically with the formal scoring relationship to drive the correlations so high. However, this speculation appears premature for two reasons. First, the *HVI* and *OBS* are the only constellations that have individual raw frequency scores that comprise the majority of their criteria (75% for the *HVI* and 80% for the *OBS*, compared to 33% for the *S-Con*, 40% for the *CDI*, 43% for the *DEPI*, and 50% for the *SCZI*). Given this, the *HVI* and *OBS* should have the strongest correlations with *R* simply on logical grounds: The more a constellation is comprised of frequency criteria that are

scored in a single direction, the more the constellation should correlate with *R*. Second, the data from this sample do not support the proposition that hypervigilant and obsessional processes cause *R* to increase. *R* was not correlated with diagnostic (Spearman $r = .03$) or self-report criteria of paranoia (e.g., with Scale 6, $r = .14$). It was also negatively correlated with obsessional diagnostic criteria (Spearman $r = -.23$) and unrelated to self-reports of obsessionally ruminative activity (e.g., with MMPI–*OBS*, $r = .11$). In a related fashion, this sample did not provide evidence that social and emotional coping deficits result in low *R*. Instead, there was a trend for *R* to be positively correlated with coping deficit-type diagnoses (Spearman $r = .20$) and with self-report scales of interpersonal or social dysfunction (e.g., $r = .22$ with Scale *0* and .19 with both *SOD* and *WRK*).

Note, however, that *R* is positively correlated with anxiety and withdrawn introversive tendencies ($rs = .32$ and .22, respectively). This suggests that *R* may increase in response to agitated ideational activity that is more diffuse and less symptomatic than overt obsessional or paranoid characteristics. However, these internal states only appear to account for a small proportion of *R*'s variability. A much larger proportion (i.e., at least 85%) appears to be due to other factors, many of which may be unrelated to the psychological characteristics measured by the data of the Rorschach. These factors may include guardedness, beliefs over what is expected, and the nature of the testing relationship, among others.

In summary, although obsessional or hypervigilant tendencies may occasionally cause *R* to increase, and coping deficits may at times result in low *R*, this study does not indicate these are consistent forces that influence *R* in the Rorschach. As such, *R* does not appear to vary in ways that enhance the internal validity of the constellations. This suggests that *R* should be controlled in some fashion in order to refine internal validity and the fidelity of the constellations. This issue is discussed in more detail later.

External Validity

Data from the second part of this study indicate that *R* plays a significant role in determining the validity of the *SCZI*, *DEPI*, *S-Con*, *HVI*, and *OBS*. However, the role of *R* is more complicated than I initially hypothesized, and it differs depending on the constellation and whether diagnostic or self-report criteria are examined.

Diagnostic criteria. Variations from an average *R* clearly compromise the diagnostic efficacy of the *SCZI* and probably the *S-Con* too. The *SCZI* is most able to differentiate patients with a psychotic disorder when protocols are of average length. If protocols have many or few responses, the *SCZI* is no longer effective because the formal cause nature of scoring principles results in frequency criteria that are less valid representations of their underlying constructs. Two M – scores from a 15-response protocol mean something different than two M – scores from

a 22-response protocol, and both mean something different than two M − scores from a 35-response protocol. Clinically, these data indicate that *SCZI* Criteria 4, 5, and 6 should be given less weight in a longer protocol. If the *SCZI* is positive because of these criteria, it would be prudent to exercise additional caution before drawing diagnostic inferences. Conversely, greater weight could be assigned to these criteria if they are positive in a brief protocol, because it is atypical to find them positive when few responses are given.

The *S-Con* appears more related to diagnostic criteria in average-length protocols than in longer protocols. This finding is quite tentative, but it indicates long records may result in *S-Con* criteria that are positive more because of the number of responses available to score, rather than because of extensive internal distress and instability. Clinicians should recognize that although the overall *S-Con* is not positive more frequently in longer protocols, many of the individual criteria are, and inferences should be modified accordingly.

Conclusions about the *HVI*'s relationship to diagnostic criteria are also very tentative, but the data suggest that this index is mediated by *R* in a unique fashion. The *HVI* appears to be most diagnostic of a paranoid condition when it is relatively elevated in a brief record. In average-length and longer records the *HVI* is not associated with diagnoses of paranoia, even though the *HVI* is positive significantly more often in longer protocols. In retrospect, this finding is easy to understand. Despite a brief and restricted Rorschach record, if a patient displays evidence of very synthetic cognitive operations in the context of oppositionality and attentiveness to the real and imagined details of human activity, it is quite evident he or she is demonstrating hypervigilant behaviors. In a lengthy protocol, the same Rorschach scores do not appear to denote the same psychological characteristics, most likely because their meaning is confounded by *R*.

The *DEPI* was not diagnostic of a depressive disorder in the total sample or any of the *R* groups. This is a surprising finding that deserves further investigation and clarification. However, its apparent lack of diagnostic specificity may have been due to the fact that depressive symptoms were so ubiquitous in this sample (i.e., occurring in all but 16 patients). This is likely to be a problem for most patient samples, and future research with the *DEPI* should include a clearly defined, nondepressed control group to more completely assess the diagnostic validity of this index.

Self-report criteria. Before discussing the relationship of the constellations to MMPI–2 criteria, it is worth reviewing the methods and types of information that are being contrasted in these analyses. Even though there should not necessarily be a direct correspondence between Rorschach and self-report data sources, authors frequently assume there will be (e.g., Greenwald, 1990, 1991; Meyer, 1992b; Weiner, 1992). Particularly in the past when correspondence was not evident, it was usually the validity of the projective technique that was

criticized. As an aside, it is interesting to note that similar demands and similar criticisms never developed for other complex but "nonprojective" performance-based measures. For example, I am not aware of any literature that argues scores on intelligence tests or neuropsychological tests must correspond to self-reported ability in order to demonstrate validity.

A number of authors (e.g., Leary, 1957; McClelland, 1980; McClelland, Koestner, & Weinberger, 1989; Meyer, 1992b; Weiner, 1992; Westin, 1991, in press) have discussed the distinctions between self-assessments and assessments based on projective data, or the general limitations of self-report data (e.g., Kagan, 1988; Wilson, Hull, & Johnson, 1981). Essentially, a self-report personality instrument requires a person to reflect on their character to make judgments about what they are like (i.e., Do I really enjoy reading mechanics magazines?; Do I generally feel blue?). This is a cognitive process that is dependent on conscious self-representations and awareness of internal experiences. For this process to result in valid information, subjects must be relatively insightful, conflict-free, and have stable, well-articulated, and differentiated self-schemata. In addition, self-report data is valid only if subjects are motivated to accurately share what they know or believe about themselves. Thus, unless more sophisticated configural analyses are utilized, self-report scales yield information that the patient is both aware of and willing to report in an accurate manner. Self-reports remain one of the best ways to ascertain specific information, such as whether or not one hears voices, has ever attempted suicide, has been in trouble with the law, or believes the FBI is plotting against them. However, for a patient population, in which rigid defensive structures are the rule rather than the exception and in which a substantial number of patients may overreport or deny symptomatology for various reasons, it is a tenuous assumption to consider self-report scales to be "true" markers of a patient's condition.

The kind of information obtained from a projective test such as the Rorschach is very different from that of a self-report measure. The Rorschach elicits a sample of actual behavior in response to the inkblots that is then coded along certain dimensions. Although this behavior is still affected by censoring to some degree, it is a demonstration of how one copes with and organizes the blot stimuli, and it yields information about stable psychological styles and processes, internal dynamics, and tacit self–other representations, all of which are not necessarily mediated by the patient's conscious schemata (Blatt, 1990; Exner, 1980). Furthermore, unlike a self-report inventory that relies on patients to accurately understand what is being asked of them and that relies on all patients to use roughly the same benchmarks for self-evaluation, accurate and consistent categorization of the different aspects of Rorschach behavior relies on the skills of the assessor.

In this study, there are generally no substantial correlations between the Rorschach constellations and the MMPI–2 scales when *R* is disregarded. This replicates the result of earlier investigations with the *DEPI* and *SCZI*. However,

when different levels of *R* are examined, it becomes clear that the number of responses a patient gives mediates the way the *DEPI, HVI, OBS,* and, to a lesser extent, *S-Con,* correlate with self-assessment data. *R* appears to function as a dimension of openness, willingness, or ability to acknowledge dysfunction and reveal symptomatology. This is suggested because patients who deliver many responses also readily acknowledge the presence or absence of anxious, depressive, or paranoid symptomatology in a fashion that directly corresponds to the respective Rorschach constellations. The picture is very different when patients give few responses. Under these circumstances, the MMPI–2 scales tend to directly contradict the Rorschach constellations.

In high-*R* protocols, there is clear and direct support for the validity of the constellations, and one can be fairly confident that Rorschach data will correspond to reports patients give of their condition. The magnitude of a number of the cross-method correlations in lengthy protocols is quite compelling. For example, the *HVI* correlated with Scale 6 of the MMPI–2 at .50; the *DEPI* correlated with the *DEP, LSE,* and *WRK* scales at .40, .46, and .51, respectively; the *S-Con* correlated with *ANX* at .42; and the *OBS* correlated with Scale 2 and *ANX* at .44 and .43, respectively.

A greater appreciation of the magnitude of these correlations can be gained by turning to the personality assessment literature on cross-method convergent validity. Much of this research has been concerned with self-report personality questionnaires or rating scales. It is not uncommon for studies to demonstrate convergent validity by having individuals first rate themselves on a personality questionnaire (i.e., true or false in response to questions) and then on a personality rating scale (i.e., Likert-type rating of characteristics). The content being rated by both measures is the same (e.g., traits of extraversion), and the person doing the rating is also the same; what differs is the format of the self-report instrument. In general, when well-constructed scales and questionnaires are used with bright adults or college students, correlations in the range of .5 to .7 can be obtained (e.g., Goldberg, 1992; McCrae & Costa, 1985). If one alters the methodology so that self-reports are now compared to ratings made by spouses or close friends, the magnitude of the cross-method validity coefficients drop into the range of .2 to .6, which is still considered substantial evidence of validity (e.g., Funder & Colvin, 1988; McCrae & Costa, 1987; Watson & Clark, 1991).

McCrae and Costa (1987) reported a study that compared the self-ratings of well-educated nonpatients on the "big five" personality dimensions to ratings made about them by several close friends on the same dimensions. In discussing their results, they stated:

> The magnitude of the correlations—generally .4 to .6—deserves some comment, because it was larger than typically reported . . . in part [this may be due] to the nature of the raters . . . [who] were very well acquainted with the subjects they

> rated, having seen them frequently in a wide variety of circumstances over a period of many years. . . . Given the qualifications of the raters in this study, it seems likely that the correlations seen here will be near the ceiling for self–other agreement. (p. 86)

Given that Rorschach data are obtained from a method wholly unlike a self-report inventory and that these data are scored and interpreted by a tester who has never seen the patient before the day of the assessment, the substantial convergent correlations seen within the high-*R* protocols must be recognized as powerful testaments to the Rorschach's validity.

The picture is less clear-cut, but perhaps it is just as compelling for patients who deliver few responses to the Rorschach. In low-*R* protocols, the constellations tend to directly contradict self-reports of symptomatology, as the *DEPI*, *HVI*, and *OBS* correlate with MMPI–2 scales of affective or interpersonal distress in the range from −.40 to −.64. On a superficial level, one could interpret these findings to mean the Rorschach constellations are actively invalid in brief protocols. However, the true scenario appears more complicated.

The strong negative correlations in the low-*R* protocols are probably facilitated by several factors, and one must consider patients who are high on the self-report scales but low on the Rorschach constellations, as well as patients who are low on the MMPI–2 scales but high on the constellations. Patients who are high on the *DEPI*, *OBS*, *HVI*, and *S-Con* but report minimal depressive, anxious, or interpersonal symptomatology are likely to have masked depressions, to be alexithymic, or to be feigning emotional health. As such, these patients are unable or unwilling to acknowledge affective symptoms and interpersonal problems. The restricted number of responses these patients give to the Rorschach can also best be understood as efforts to contain, hide, or restrict symptomatology. For example, in this study five patients with MMPI–2 data had positive *DEPI*s and low-*R* protocols (see Table 5). Each of these patients had MMPI–2 depression scales with a *T*-score less than 50. However, it was clear that their self-evaluations were inconsistent with their clinical picture. Two patients were admitted following serious suicide attempts with full intentionality, one had a long history of depression and was hospitalized following a manic episode, another was diagnosed with a brief psychotic reaction and admitted with guilt-based delusions and command hallucinations to kill himself, and the last was a forensic case in which the patient was attempting to present himself in an emotionally healthy fashion to further a lawsuit that claimed a physical injury caused his erectile dysfunction. Thus, similar to the *HVI*, it appears that when the *DEPI* is elevated in a brief record, it is a more accurate indicator of a patient's clinical condition than the MMPI–2 scales.

Data are less available to help understand those patients who produce low-*R* protocols, report considerable depressive and/or interpersonal symptomatology, and yet show low scores on the *DEPI*, *S-Con*, *OBS*, and *HVI*. It is likely that

these patients represent a heterogeneous group, and further empirical studies would be helpful to understand them. However, I speculate that some of these patients are exaggerating their reports of dysfunction for some reason (e.g., high *F* on the MMPI–2). Others may be accurately reporting distress but inhibiting its behavioral expression on the Rorschach so as not to overwhelm their ego functioning. Others may be accurately reporting extensive depression and distress but because of their withdrawal and lack of energy may be unable to marshall the cognitive and affective resources necessary to engage with the Rorschach (e.g., Hartlage, Alloy, & Vazquez, in press). This is most likely to be a factor with the *HVI* and *OBS* because these indices tap more active and synthetic cognitive operations than the *DEPI* and *S-Con*.

The CDI. In this study the *CDI* was unique among the constellation indices because it was unrelated to diagnostic criteria and negatively related or unrelated to self-report criteria. The *CDI* is frequently positive in brief protocols and/or in protocols where *Lambda* is high (e.g., the correlation between *Lambda* and total *CDI* scores in this sample is .32, $p = .003$). Beside suggesting coping deficits, low *R* and high *Lambda* can both indicate guarded defensiveness. It would be fruitful for other investigations to attempt to disentangle guardedness from legitimate social and emotional coping deficits to determine whether the *CDI* has validity when it is not potentially confounded by this factor. More importantly, however, other studies should use validity criteria that correspond more closely to the *CDI* than those used in this study.

Limitations

Both of the validation criteria used in this study have deficiencies for judging the validity of the Rorschach. The limitations of the MMPI–2 criteria have already been discussed, but the *DSM–III–R* diagnoses are also limited in several ways. First, the reason most of these patients were evaluated with psychological testing is because diagnostic uncertainty remained after an interview or after inpatient observation. Second, diagnoses were assigned during the course of regular clinical practice rather than through more strict and systematic review of symptomatology. As such, dichotomous decisions were made to quantify a broad and variable clinical presentation across many psychological dimensions. A consequence of this is that an appropriate diagnosis may not fully characterize a patient's condition. For example, a patient may have a diagnosis of schizophrenia but have psychotic symptoms well-controlled by medication and be admitted for increasing suicidal ideation. Even though the diagnosis of schizophrenia is still appropriate, this patient should not necessarily be positive on the *SCZI* but may well be positive on the *DEPI*. A third problem is that the *DSM–III–R* diagnoses do not cleanly correspond to some of the Rorschach

constellations. This is particularly the case for the *S-Con* and *CDI*, although it is also a problem for the *HVI* and, to some extent, the *OBS*.

Several other limitations to this study deserve note. First, a relatively small number of subjects were utilized, particularly for some analyses conducted within each of the *R* groups. In a related fashion, this study had few patients diagnosed with hypervigilant or obsessional conditions, which limits the generalizations that can be made concerning the *HVI* and *OBS*. An additional caution relates to the high number of analyses that were conducted within the overall experiment. Alpha levels were not adjusted to account for all of these, in part for fear of too severely restricting statistical power and in part because of the preliminary nature of this study.

Implications

Future research. As noted, the Rorschach constellations did not display equivalent patterns of validity with MMPI–2 criteria and diagnostic criteria. The *SCZI* is related to a psychotic diagnosis in average-length protocols, as are the appropriate MMPI–2 scales; however, the *SCZI* does not correlate with MMPI–2 scales in the total sample or within each of the *R* groups. The *HVI* appears most diagnostic of a paranoid condition in brief protocols, but it is most strongly related to self-reports of paranoia in longer protocols. The *S-Con* is more positively related to suicide criteria in average-length records than in longer protocols, but MMPI–2 scales of internal distress are most elevated when the *S-Con* is positive in longer protocols. The *DEPI* is highly associated with reports of distress in longer protocols, and reports of distress are associated with diagnostic criteria, but the *DEPI* is unrelated to diagnostic criteria within the total sample and each of the *R* groups.

These inconsistent findings are somewhat perplexing if the distinctions and limitations of each of these three data sources are not recognized. Self-reports of symptomatology in specific areas do not exactly correspond to observations codified in *DSM–III–R* diagnoses or to the tacit personality characteristics and dynamics assessed by the Rorschach, and tacit characteristics of personality and affective life do not exactly correspond to *DSM–III–R* diagnoses. To refine the science of personality assessment, these distinctions should be recognized and emphasis should be shifted to understanding what conditions lead these data sources to converge and diverge in theoretically understandable ways.

A precondition for this type of research is improvement in the diagnostic criteria used to judge the validity of the Rorschach. Two suggestions are offered. First, it would be optimal if criteria were founded on Spitzer's (1983) suggestion that expert clinicians form judgments about patients after evaluating them over time and after taking into account all available sources of information. Use of the Longitudinal Expert evaluation of All Data (LEAD) standard has been fruitfully applied in personality disorder research (e.g., Pilkonis, Heape, Ruddy,

& Serrao, 1991; Skodol, Rosnick, Kellman, Oldham, & Hyler, 1988) and would be ideal for future Rorschach investigations. Second, it would be optimal if ratings are made on dimensional criteria that directly correspond to the data of the Rorschach. The Rorschach yields unique kinds of information that are not necessarily available from standard rating scales. To maximize evidence for the Rorschach's validity, the conceptual link between Rorschach predictors and clinical criteria should be as close as possible. (Scales for rating patients on Rorschach dimensions are available by writing me.)

To refine understanding of how the Rorschach relates to self-ratings, additional steps can also be taken. In a general sense, the data from this study suggest that patients who are willing or driven to give many responses have greater congruence between their conscious self-representations (MMPI–2 scales) and the tacit conditions measured by the Rorschach. Patients who are unable, unwilling, or afraid of responding to the Rorschach have considerable dissynchrony between these data sources. With the exception of Scale 6, however, the pattern of consistency in longer records and disagreement in brief records was found with data related to emotional and interpersonal distress. This pattern was not evident when data related to psychotic symptoms were examined. Experiences of emotional and interpersonal distress are fairly easy to represent cognitively, but they are also frequently defended against. In contrast, and almost by definition, psychotic processes may be harder for patients to represent in a conscious and objective fashion. Because the data from the Rorschach cover a multiplicity of functional domains (e.g., coping styles, affect modulation, narcissism, oppositionality, reality testing, conceptual rigidity, abuse of fantasy, etc.), it would be useful for future research to determine which Rorschach information is more amenable to conscious representation, which is more likely to operate outside the realm of awareness, and which is more subject to defensive operations.

The control of R. Recently, a series of papers addressed the question of whether *R* should be controlled in the Rorschach to improve its clinical utility and psychometric foundation (Exner, 1992; Kinder, 1992; Lipgar, 1992; Meyer, 1992a). Each of these authors referenced a slightly different body of data and did not reach a consensus on this issue. For example, I presented data that suggested it would be useful to conduct research into an *R*-controlled format of administration in order to determine if this improved internal and external validity, Lipgar discussed the clinical and empirical evidence that supported the importance of retaining *R* as a variable, Exner provided evidence to indicate the issue is less important for the Comprehensive System than it was for earlier scoring systems, and Kinder drew a distinction between the clinical use of the Rorschach when he felt *R* should not be controlled and research applications where certain limited circumstances may require some control of *R*.

The data from this study indicate that methods to control the effects of *R* are

likely to have distinct consequences for different external criteria. It appears that controlling the effects of *R* would enhance the internal validity of all the constellations, as well as enhance the external validity of at least some of the constellations in relation to diagnostic criteria. However, controlling *R* would detract from the correlations that can be found between some constellations and self-report criteria. For example, although it may be helpful to partial *R* out of correlations with diagnostic criteria, this technique should result in near 0 correlations when most constellations are examined in relation to self-report criteria. In fact, if one wished to maximize the validity coefficient with self-report criteria, one should select only insightful individuals who are willing and able to provide many responses to the Rorschach. I believe this would be a misguided goal, however, because it would trivialize the value of the Rorschach for clinical purposes. Instead, to help settle the questions related to controlling *R*, carefully designed studies using the LEAD standard as ultimate Rorschach validity criteria would be most helpful.

CONCLUSION

In summary, the findings presented here highlight the significant impact *R* has on other Rorschach scores and on the Rorschach's relationship to external criteria. If these findings are replicated in other samples, the discovery of *R*'s complex role in the Rorschach should help clarify the conditions under which this instrument can be used as a valid and unique nomothetic tool for understanding personality. Simultaneously, however, these data underscore the need for examiners to be sensitively attuned to the dynamics of the patient and the testing relationship (Schafer, 1954) in order to understand what factors are causing *R* to vary. It is necessary to make use of this idiographic information in order to appropriately qualify standard nomothetic inferences regarding the constellations and frequency based scores (Exner, 1991) when protocols are brief or lengthy. Finally, this study indicates that when possible, research with the Rorschach should incorporate a methodology that allows the impact of *R* to be examined. These efforts will provide further understanding of the many factors that influence *R* and the validity of the Rorschach, and they will offer assistance when interpreting protocols of various lengths.

ACKNOWLEDGMENTS

An earlier version of this article was presented at the Annual Meeting of the Society for Personality Assessment, Washington, DC, March, 13, 1992.

I thank Theresa Kiolbasa for providing the reliability ratings, William Divane and Robert Faull for their assistance with data management, and Sharon Grob-Murphy for her helpful comments on a draft of this article.

REFERENCES

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd. ed., rev.). Washington, DC: Author.

Archer, R. P., & Gordon, R. A. (1988). MMPI and Rorschach indices of schizophrenic and depressive diagnoses among adolescent inpatients. *Journal of Personality Assessment, 52,* 276–287.

Ball, J. D., Archer, R. P., Gordon, R. A., & French, J. (1991). Rorschach depression indices with children and adolescents: Concurrent validity findings. *Journal of Personality Assessment, 57,* 465–476.

Blatt, S. J. (1990). The Rorschach: A test of perception or an evaluation of representation. *Journal of Personality Assessment, 55,* 394–416.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory (MMPI-2). Manual for administration and scoring.* Minneapolis: University of Minnesota Press.

Cronbach, L. J. (1949). Statistical methods applied to Rorschach scores: A review. *Psychological Bulletin, 46,* 393–429.

Exner, J. E., Jr. (1980). But it's only an inkblot. *Journal of Personality Assessment, 44,* 562–577.

Exner, J. E. (1986a). *The Rorschach: A comprehensive system: Vol. 1. Basic foundations* (2nd ed.). New York: Wiley.

Exner, J. E., Jr. (1986b). Some Rorschach data comparing schizophrenics with borderline and schizotypal personality disorders. *Journal of Personality Assessment, 50,* 455–471.

Exner, J. E., Jr. (1989). Searching for projection on the Rorschach. *Journal of Personality Assessment, 53,* 520–536.

Exner, J. E., Jr. (1990). *A Rorschach workbook for the Comprehensive System* (3rd ed.). Asheville, NC: Rorschach Workshops.

Exner, J. E., Jr. (1991). *The Rorschach: A comprehensive system* (Vol. 2, 2nd ed.). New York: Wiley.

Exner, J. E., Jr. (1992). *R* in Rorschach research: A ghost revisited. *Journal of Personality Assessment, 58,* 245–251.

Exner, J. E., Jr., Armbruster, G., & Mittman, B. (1978). The Rorschach response process. *Journal of Personality Assessment, 42,* 27–37.

Fiske, D. W., & Baughman, E. E. (1953). Relationships between Rorschach scoring categories and the total number of responses. *Journal of Abnormal and Social Psychology, 48,* 25–32.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55,* 149–158.

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 4,* 26–42.

Greenwald, D. F. (1990). An external construct validity study of Rorschach personality variables. *Journal of Personality Assessment, 55,* 768–780.

Greenwald, D. F. (1991). Personality dimensions reflected by the Rorschach and the 16PF. *Journal of Clinical Psychology, 47,* 708–715.

Hartlage, S., Alloy, L. B., & Vazquez, C. (in press). Automatic and effortful processing in depression. *Psychological Bulletin.*

Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart & Winston.

Kagan, J. (1988). The meaning of personality predicates. *American Psychologist, 43,* 614–620.

Kinder, B. N. (1992). The problems of *R* in clinical settings and in research: Suggestions for the future. *Journal of Personality Assessment, 58,* 252–259.

Leary, T. (1957). *Interpersonal diagnosis of personality.* New York: Ronald.

Lipgar, R. M. (1992). The problem of *R* in the Rorschach: The value of varying responses. *Journal of Personality Assessment, 58,* 223–230.

Lipovsky, J. A., Finch, A. J., & Belter, R. W. (1989). Assessment of depression in adolescents: Objective and projective measures. *Journal of Personality Assessment, 53,* 449–458.

McClelland, D. C. (1980). Motive dispositions: The merits of operant and respondent measures. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 1, pp. 10–41). Beverly Hills: Sage.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96,* 690–702.

McCrae, R. R., & Costa, P. C., Jr. (1985). Updating Norman's "adequate taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology, 49,* 710–721.

McCrae, R. R., & Costa, P. C., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52,* 81–90.

Meyer, G. J. (1992a). Response frequency problems in the Rorschach: Clinical and research implications with suggestions for the future. *Journal of Personality Assessment, 58,* 231–244.

Meyer, G. J. (1992b). The Rorschach's factor structure: A contemporary investigation and historical review. *Journal of Personality Assessment, 59,* 117–136.

Pilkonis, P. A., Heape, C. L., Ruddy, J., & Serrao, P. (1991). Validity in the diagnosis of personality disorders: The use of the LEAD standard. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3,* 46–54.

Schafer, R. (1954). *Psychoanalytic interpretation in Rorschach testing.* New York: International Universities Press.

Shapiro, J. P., Leifer, M., Martone, M. W., & Kassem, L. (1990). Multimethod assessment of depression in sexually abused girls. *Journal of Personality Assessment, 55,* 234–248.

Skodol, A. E., Rosnick, L., Kellman, D., Oldham, J. M., & Hyler, S. E. (1988). Validating structured DSM-III-R personality disorder assessments with longitudinal data. *American Journal of Psychiatry, 145,* 1297–1299.

Spigelman, A., & Spigelman, G. (1991). Indications of depression and distress in divorce and nondivorce children reflected by the Rorschach test. *Journal of Personality Assessment, 57,* 120–129.

Spitzer, R. L. (1983). Psychiatric diagnosis: Are clinicians still necessary? *Comprehensive Psychiatry, 24,* 399–411.

Watson, D., & Clark, L. A. (1991). Self-versus peer ratings of specific emotional traits: Evidence of convergent and discriminant validity. *Journal of Personality and Social Psychology, 60,* 927–940.

Weiner, I. B. (1992, March). *Self-report validation of Rorschach indices of adjustment difficulty.* Paper presented at the annual meeting of the Society for Personality Assessment, Washington, DC.

Westin, D. (1991). Social cognition and object relations. *Psychological Bulletin, 109,* 429–455.

Westin, D. (in press). The cognitive self and the psychoanalytic self: Can we put our selves together? *Psychological Inquiry.*

Wilson, T. D., Hull, J. G., & Johnson, J. (1981). Awareness and self-perception: Verbal reports on internal states. *Journal of Personality and Social Psychology, 40,* 53–71.

Gregory J. Meyer
Department of Psychiatry
University of Chicago Hospitals
5841 South Maryland Avenue, MC 3077
Chicago, IL 60637–1470

Received May 5, 1992