Advancing the Science of Psychological Assessment: The Rorschach Inkblot Method as Exemplar

Irving B. Weiner University of South Florida

This article comments on a series of 5 articles, concerning the utility of the Rorschach Inkblot Method (RIM). Two of the articles provide extensive empirical evidence that the RIM has been standardized, normed, made reliable, and validated in ways that exemplify sound scientific principles for developing an assessment instrument. A 3rd article reports a meta-analysis, indicating that the RIM and the Minnesota Multiphasic Personality Inventory have almost identical validity effect sizes, both large enough to warrant confidence in using these measures. The other 2 articles adduce sketchy data and incomplete literature reviews as a basis for questioning the psychometric soundness of Rorschach assessment. Unwarranted skepticism should not be given credence as an adequate platform from which to challenge abundant evidence that the RIM works very well for its intended purposes.

Because psychology is a behavioral science and assessment is a field of psychology, the credentials of psychological assessment instruments should be measured against scientific principles for advancing knowledge. Data sets used to evaluate the clinical utility of the Rorschach Inkblot Method (RIM) should accordingly be thorough and current, not incomplete or obsolete; they should be accurate and relevant to the issue at hand, not misleading, taken out of context, or beside the point; and they should be derived from adequately designed research studies. The information provided by these data sets should address the following six questions concerning the psychometric adequacy of an assessment instrument: (a) Does the RIM comprise standardized procedures for administration and scoring? (b) Can trained examiners agree reasonably well in their scoring of the RIM's variables? (c) Do reliability estimates indicate that the RIM's scores provide reasonably accurate information by virtue of closely approximating actual or true scores? (d) Are there adequate normative reference data concerning the descriptive statistics of the RIM's variables among various populations? (e) Does the RIM demonstrate corollaries that identify purposes for which it is reasonably valid? and (f) Can the RIM show incremental validity when applied in decision-making situations?

This article addresses these six questions with respect to Rorschach assessment, comments on the five articles previously published in this Special Series on Rorschach utility, and reaches the following conclusion: By virtue of the manner in which the RIM has been conceptualized, standardized, normed, and validated over the past 20 years, the available data identify this inkblot method as an exemplar of sound science in the development and application of a psychological assessment instrument.¹

Standardization

The 50 years following Rorschach's (1921/1942) original publication of the monograph that describes his test gave rise to numerous disparate schools of Rorschach methodology and countless individual variations in Rorschach technique. Despite a wealth of creative thinking and extensive research efforts, the unsystematic nature of these early endeavors prevented Rorschach assessment from establishing much claim to scientific status. Since the early 1970s, however, with the advent of the Rorschach Comprehensive System (Exner, 1991, 1993; Exner & Weiner, 1995), Rorschach assessment has evolved into a carefully standardized procedure. Exner (1993, chapter 3) prescribed an invariant set of instructions to be used in all Rorschach administrations, strict adherence to which ensures that Rorschach data are collected in the same manner from all respondents and can be cumulated for research purposes. Exner (1993, chapters 4-10) also formulated detailed guidelines for coding individual Rorschach responses and combining these codes into various percentages, ratios, and indices. Standardization of an instrument resides in its being composed of specific stimuli, methods of administration, instructions to respondents, and coding criteria to be used without exception, and differences in these components compromise the generalizability of findings from one person and one situation to the next. Strict adherence to Exner's prescribed Comprehensive System ensures the consistent comparability of data obtained from it under any circumstances.

Over the years, the Comprehensive System has become the most widely used approach to Rorschach assessment. The vast majority of students who are learning the Rorschach in graduate programs are being taught the Comprehensive System: It is by far the method most commonly used in published research studies, and abroad it is the prevalent approach in almost every country in

I thank Mark J. Hilsenroth for his helpful suggestions concerning the preparation of this article.

Correspondence concerning this article should be addressed to Irving B. Weiner, 31716 Halliford Drive, Tampa, Florida 33624. Electronic mail may be sent to iweiner@hsc.usf.edu.

¹ The final version of this manuscript was submitted on October 6, 1999. For this reason, with the exception of some in press material that was made available, it does not include reference to any publications appearing subsequent to that data.

which the Rorschach is used (see Hilsenroth, 2000; Weiner, 1998b). As the only carefully standardized approach to collecting and codifying Rorschach data, the Comprehensive System is the primary Rorschach method that should be considered in evaluating the scientific status of the instrument. Idiosyncratic and individually tailored ways of administering and coding the Rorschach will inevitably lead as in the past to inconsistent research findings. Inconsistent research findings attributable to nonstandard administration and coding have often provided grist for the mill of Rorschach critics, but in fact they have no bearing at all on the psychometric adequacy of Rorschach assessment. In a similar vein, misguided and unwarranted interpretations of Rorschach responses can lead to errors in clinical practice and generate anecdotes that depict Rorschach assessment as a dangerous procedure. Like failure to follow standard procedures, however, illinformed interpretations reflect clinician incompetence and constitute an abuse of the instrument; they say nothing about the soundness and value of the RIM when it is used properly.

It should nevertheless be acknowledged that the procedural standards intrinsic to the Comprehensive System do not encompass the interpretation of the Rorschach. Exner (1991, chapters 5-10) provided some detailed guidelines for Rorschach interpretation, but these interpretive strategies constitute a recommended way of working with the obtained data and are not an essential feature of the Comprehensive System approach. Other contemporary Rorschachists have endorsed Comprehensive System methods of administration and coding while elaborating interpretive strategies that diverge from or extend beyond those proposed by Exner (see, e.g., Gacono & Meloy, 1994; Meloy, Acklin, Gacono, Murray, & Peterson, 1997; Smith, 1994; Weiner, 1998a). The RIM resembles other widely used tests in this respect, with standardization residing in administration and in scoring procedures and various different interpretive approaches being used to determine what the obtained data signify. Thus, responses to the Wechsler Intelligence Scales are used by some psychologists solely as indications of intellectual functioning, whereas other psychologists draw inferences about personality characteristics as well from how respondents behave and express themselves while taking the test. In interpreting the second version of the Minnesota Multiphasic Personality Inventory (MMPI-2), some clinicians focus mainly on individual scale scores and other clinicians focus on code types, and differences of opinion persist concerning how much attention should be given to the subtle items in the test. Such individual variations and personal preferences in ways of approaching the interpretive process do not detract from the basic standardization of these instruments with respect to their administration and coding, not for the RIM any more than for the Wechsler or the MMPI.

These observations on Rorschach standardization have three important implications for the basic nature and utility of the RIM. First, Rorschach data are not theory bound and can be interpreted within the framework of diverse ways of conceptualizing personality processes: Provided that one believes in the existence of personality states and traits, the implications of Rorschach responses can be couched in whatever theoretical language one prefers. Second, the Rorschach stimuli and instructions are essentially cultural free, and the instrument can be administered in a standard manner independent of a respondent's age, gender, ethnicity, nationality, or other demographic characteristics; crosscultural and cross-national linguistic differences may influence the meanings that attach to the verbal content of responses, but no translation of the test stimuli is required to obtain a valid protocol from anyone, anywhere in the world. Accordingly, quantitative and qualitative research is necessary to determine ethnic differences in obtained scores and their interpretive significance, but no modification in basic Rorschach procedures is required to obtain meaningful Rorschach data from any group of respondents, except occasionally for young children.

Third, Rorschach data lend themselves both to nomothetic interpretation based on comparisons with available normative data and to idiographic interpretation that allows for individualized case conceptualization. The previously published articles by Stricker and Gold (1999) and by Viglione (1999) elaborate each of these advantages of Rorschach assessment; the meta-analysis by Hiller, Rosenthal, Bornstein, Berry, and Brunell-Neuleib (1999) does not address these points; and the articles by Hunsley and Bailey (1999) and by Dawes (1999) are silent with respect to them.

Intercoder Agreement

Assessment psychologists generally concur that test scores can serve useful purposes only if examiners can agree on what these scores should be. Without substantial agreement concerning the scores that should be assigned to test variables, these variables hold little promise for demonstrating internal consistency, temporal stability, or consistent extratest corollaries. Opinions vary, however, concerning the most appropriate way of evaluating intercoder agreement, especially in the case of assessment instruments that involve multiple complex coding categories. McDowell and Acklin (1996), Meyer (1997a, 1997c), and Wood, Nezworski, and Stejskal (1996, 1997) among others have engaged in lively debate, concerning such matters as (a) the relative appropriateness of calculating percentage agreement versus Kappa coefficients in determining intercoder agreement for Rorschach responses and (b) the necessity of using entire responses versus response segments as the unit on which to base these calculations. By far the most extensive and methodologically sophisticated work on this topic has been conducted by Meyer (1997a), who reported a metaanalytic study of interrater reliability data published in Rorschach research articles, using Comprehensive System variables that appeared in the Journal of Personality Assessment from the beginning of 1992 to the end of 1995. Using Kappa to assess these data for the chance-corrected interrater reliability of 10 response segments (e.g., Location, Determinants, Form Quality, etc.), Meyer obtained coefficients for individual segments, ranging from 0.72 to 0.98, with a mean value of 0.88. As Meyer noted, Kappa values greater than 0.75 are generally considered to demonstrate excellent beyond-chance agreement.

The present status of knowledge that concerns whether the Rorschach can be coded reliably is addressed in two of the five series articles under discussion here, those by Viglione (1999) and by Hunsley and Bailey (1999). In light of the previously stated criteria for a sound scientific approach to advancing knowledge, the comments of these authors should be introduced with some attention to the data sets on which they relied in drawing conclusions about the psychometric adequacy of the Rorschach method. As noted by Meyer (1999b) in his introduction to this Special Series, the authors of each contribution were provided with abstracts of 445 Rorschach articles identified by PsycLIT as having been published during 1977-1997 in five journals: Assessment, Journal of Clinical Psychology, Journal of Consulting and Clinical Psychology, Journal of Personality Assessment, and Psychological Assessment. Viglione (1999) indicated that he sifted through these and some additional articles appearing elsewhere; that he put aside reviews of research, reports of surveys, case studies, theoretical and methodological discussions, and other secondary or nonempirical contributions; and that he focused his attention on original research studies. His reference list contains 195 items, of which 139 are asterisked to identify them as empirical studies that he evaluated in preparing his article. Hunsley and Bailey (1999) did not identify the procedures they followed in examining the relevant literature. Their reference list comprises 128 items, of which just 28 appear to be original research reports. The remaining 102 items consist of commentaries, surveys, and secondary sources, that is, reviews of research published by others.

Turning to the substantive matter at issue, Viglione (1999) came to the following conclusion concerning intercoder agreement on the Rorschach:

In preparing this article, I had the opportunity to review hundreds of interscorer reliability practices and coefficients for all types of scores with all sorts of base rates; with dichotomous, multiple categorical, or rating-scale distributions; completed in Asia, North America, and Europe; and with all sorts of reliability statistics. The fact is that just about everybody reports adequate interscorer agreement. (p. 252, footnote 4)

Hunsley and Bailey (1999) did not seriously question whether the Rorschach can be coded reliably, nor did they review any evidence that suggested that it cannot. They limited their discussion on this matter to raising two points for consideration. First, they argued that some resolution must be reached, concerning the most appropriate methodology for assessing intercoder agreement on the Rorschach. They may be correct that some one method works best for this purpose and should be the one always used, although varied approaches to most problems in psychology have usually proved scientifically enriching rather than disadvantageous. More to the point, however, it appears not to matter at all which methods are used in assessing Rorschach scoring reliability. As Viglione (1999) has pointed out, adequate intercoder agreement has been demonstrated virtually wherever, whenever, and by whomever such demonstration has been attempted and with whatever reasonable methods have been used.

Second, Hunsley and Bailey (1999) expressed concern that reliable Rorschach coding depends on the coding skills of individual examiners and that there remains "the significant question of how reliably the Rorschach is scored in routine clinical practice (i.e., field reliability)" (p. 268). Indeed, the adequacy with which clinical practitioners use methods of evaluation and treatment must always be of concern among helping professionals. However, poor practices reflect on the competence of the clinicians responsible for them, not necessarily on the soundness of the methods they are attempting to use. As a case in point, I do not recall ever having seen a field study of the reliability with which Wechsler Adult Intelligence Scale (WAIS)/WAIS-R/WAIS-III examiners assigned 0, 1, or 2 points to responses on the Comprehension, Similarities, and Vocabulary subtests. However, the Wechsler manuals provide explicit criteria and scoring examples for 0-, 1-, and 2-point answers, and conscientious examiners who study the manual and receive some training should be able to achieve good agreement in this scoring. On this basis, it is generally assumed that clinicians in practice are scoring the Wechsler tests reliably. If they are not, the shame is theirs, and the WAIS/WAIS-R/WAIS-III is not held accountable for their incompetence.

So let it be with the Rorschach. The Comprehensive System workbook (Exner, 1995) provides detailed coding criteria, numerous examples, and an extensive set of practice exercises; and the abundant research noted by Meyer (1997a) and Viglione (1999) demonstrates that examiners can readily be trained to achieve adequate intercoder agreement. Whatever grievous errors individual practitioners may commit in coding Rorschach protocols will speak to their lack of diligence and give no fault to the intrinsic psychometric adequacy of the instrument and its coding criteria. These considerations do not obviate continued effort to improve the specificity of coding criteria as a means of promoting intercoder agreement, without which any instrument, including the Wechsler scales as well as the Rorschach, will have limited validity and utility.

Finally with respect to intercoder agreement, those who remain unconvinced of the reliability of the Rorschach in this respect, even while failing to muster any data that suggest that such reliability is elusive, are paying insufficient attention to the implications in this regard of test-retest findings with the RIM. As discussed next in considering the retest reliability of the instrument, the vast majority of variables coded in the Comprehensive System have consistently demonstrated substantial stability coefficients over intervals, ranging from 7 days to 3 years. As McCann (1998) and Meyer (1997c) have pointed out, coding accuracy is always a nested component of a test-retest design. Only if both the first and second set of scores in a retest study have been assigned reliably can these scores correlate highly with each other. Hence, the substantial stability coefficients repeatedly demonstrated for numerous Rorschach variables bear irrefutable witness to good interrater reliability among the many persons who participated in coding these records.

Reliability

The reliability of Rorschach data coded, according to the Comprehensive System, has, as just noted, been demonstrated in a series of retest studies that are summarized by Exner and Weiner (1995, pp. 21-27). Almost all of the variables coded in the system that relate to trait characteristics of individuals have shown substantial short-term and long-term stability in adults. Most of these variables demonstrate retest correlations above 0.75, and some of these correlations (e.g., the Affective Ratio and the Egocentricity Index) approach 0.90. The only variables in the system that show low stability coefficients among adults are inanimate movement (m) and diffuse shading (Y), both of which are conceptualized as indexes of situational distress and are expected to prove unstable over time. Among children, 3-week retest studies identify stability coefficients similar to those found in adults. However, as would be predicted from the evolving nature of personality during the developmental years, nonpatient young people do not begin to show adult levels of Rorschach stability over a 2-year period until they reach age 14 (Exner, Thomas, & Mason, 1985).

Consistent with this evidence that the RIM is a scientifically sound instrument with respect to its reliability, Viglione (1999) concluded from his literature review that "the great majority of Rorschach comprehensive system (CS) variables and configurations have shown impressive consistency reliability" (p. 252). Hunsley and Bailey (1999) similarly stated that "it does appear that many of the scales central to the Comprehensive System can have adequate reliability... Evidence ... generally supports the reliability (often test-retest reliability) of scales commonly used in research that are scored by trained raters" (p. 268). With respect to Hunsley and Bailey's reference to scoring by trained raters, and with further relevance to the previous discussion of field reliability, one would hope that neither researchers nor clinicians would attempt any coding of Rorschach responses without having been trained to do so.

Normative Reference Data

None of the five articles being discussed addresses the availability of normative standards for Rorschach variables, even though adequate norming is identified as a requisite in test construction and evaluation in the Standards for Educational and Psychological Testing (American Psychological Association [APA], 1985). With respect to the psychometric adequacy of the RIM in this regard, the development of the Comprehensive System included the compilation of descriptive statistics for each of its codes and summary scores on a sample of 700 nonpatient adults, ages 18 to 70. As described by Exner (1993, chapter 12), this sample was randomly selected from a larger group of 1,332 volunteer participants and was stratified to include an equal number of males and females and 140 respondents from each of five geographic areas across the United States (Northeast, South, Midwest, Southwest, and West). The normative reference sample is also generally representative of United States demographic patterns of marital status, level of education, socioeconomic status, and urban-suburban-rural residence, and it includes 19% of African American, Hispanic, and Asian American respondents. Reference data are also available for 1,390 nonpatient young people, ages 5 to 16, and for groups of adult psychiatric patients, including 320 schizophrenic inpatients, 315 depressed inpatients, 440 psychiatric outpatients, and 180 patients with diagnosed character disorder.

The compilation of these reference data has helped to establish the RIM as a standardized assessment instrument and has aided assessment clinicians considerably in their efforts to identify personality assets and liabilities in persons they examine. However, normative standards remain an unfinished aspect of the Comprehensive System. Because of the manner in which Exner's nonpatient volunteer respondents were recruited, they may constitute a relatively well-adjusted rather than a generally representative sample of United States adults. With respect to utilization of the Comprehensive System abroad, moreover, some adjustments in coding criteria and some modifications in normative expectations may be required in order to apply the Comprehensive System effectively in certain countries or with certain cultural groups. On the other hand, there is good reason to believe that general principles of Rorschach interpretation are universally applicable and only slightly affected by cross-cultural differences (Butcher, Nezami, & Exner, 1998; Weiner, 1998a, chapter 2; 1999). Preliminary results from a multinational collaborative normative study involving 2,250 respondents examined in 12 different countries have, in fact, identified remarkably similar ranges of values for

many Comprehensive System variables (Erdberg & Shaffer, 1999).

Discussions of test development have tended in general to overlook the importance of normative data not only for enhancing the clinical utility of assessment instruments but also for demonstrating aspects of their construct validity. Demographic characteristics, such as a person's age or gender, can be particularly helpful in this regard, because they constitute observed variables that have little, if any, error variance and are completely independent of test findings. Should demographic differences on personality test variables emerge precisely as would be predicted by well-documented aspects of personality theory, normative data can provide powerful construct validation. The Comprehensive System normative data for young people, ages 5 to 16, do in fact provide such validation by mirroring closely certain expected maturational changes, during childhood and adolescence.

Weiner (1996) has previously called attention to three age differences of this kind. First, consistent with the ordinary tendency of personality characteristics to become increasingly stable during the developmental years, longitudinal findings indicate steadily increasing retest correlations for Rorschach variables as young people age from 8 to 16 years. Second, the normal maturational tendency for children over time to become emotionally less intense and more reserved is reflected in a steadily decreasing ratio from age 6 to 16 of the number of color-form (CF) responses, which are considered a corollary of relatively unmodulated patterns of emotionality, to the number of form-color (FC) responses, which are considered a corollary of relatively modulated patterns of emotionality. Third, children are typically observed to become less self-centered as they mature and are increasingly capable of focusing their attention on other people as well as themselves, and the Egocentricity Index, which is considered to signify the extent to which a person is self-focused, diminishes in a steady linear fashion from age 5 (M = 0.69) to age 16 (M = 0.43). In a similar data analysis based on the Comprehensive System age-group norms, Wenar and Curtis (1991) demonstrated longitudinal Rorschach changes consistent with predictions from developmental psychological data, concerning increases over time in cognitive complexity, precision of thinking, and conformity to socially acceptable ways of thinking.

Validity

In the final analysis, even the most carefully standardized, highly reliable, and thoroughly normed assessment instrument serves little purpose unless it can be demonstrated to have meaningful correlates. The nature and extent of an instrument's corollaries define its validity and utility, and the five articles under discussion are addressed mainly to this feature of the RIM. Three of the articles (Stricker & Gold, 1999; Hiller et al., 1999; Viglione, 1999) address methodological considerations as well as substantive findings in assessment validity research, and a brief review of three such considerations can help to introduce comments on the substantive findings.

Methodological Considerations

Assessment and measurement psychologists generally concur with respect to three considerations in designing and evaluating empirical studies of test validity. First, the validity of assessment instruments that are multidimensional in nature cannot be captured by a single numerical value or narrative statement. Each of the scale scores generated by such instruments has its own validity coefficients, and each scale's coefficients vary with the purposes for which it is used. Depending on the correlates of their individual scales, most instruments demonstrate greater validity for some purposes than for other purposes; depending on their nature, some instruments demonstrate greater validity for certain purposes than other instruments. The RIM is by nature a measure of personality processes, and Rorschach assessment should accordingly be expected to measure variables that are determined mainly by personality characteristics; it should not be expected to produce significant correlations with phenomena or events that are attributable largely to nonpersonality variance, nor should its validity be impugned when it fails to do so.

Second, the validity of assessment instruments should be judged primarily from their correlations with observed rather than inferred variables. Observed variables comprise directly noted features of how people think, feel, and act; inferred variables consist of hypotheses about how people are likely to think, feel, and act that are derived from indirect sources of information. Because personality assessment instruments are inferential measures, their correlations with each other provide little information that concerns their validity for explaining or predicting observed behavior; it is possible for a set of inferential measures to correlate substantially with each other while showing no consistent relationship to observable behavior. Viglione (1999) elaborated this consideration in his article and stressed the importance of validating Rorschach variables against "behavioral, real-life criteria," as did Hilsenroth, Fowler, Padawer, and Handler (1997) in research that demonstrated how well the RIM can identify accurately pathological expressions of narcissism.

Correlations between inferential measures are especially limited in their significance for validity when the measures involve different methods of approach. The RIM, for example, is a largely indirect assessment method, involving a relatively unstructured set of test materials and instructions to which people respond with little conscious awareness of what their answers might signify. The MMPI and similar self-report inventories, by contrast, are largely direct assessment methods, involving relatively structured materials and instructions; by virtue of how their clinical scales were developed, these inventories may include subtle as well as obvious items, but people respond to them for the most part with considerable awareness of what their answers might signify. An extensive conceptual and empirical literature documents that method variance can produce substantial differences between Rorschach and MMPI findings, even on variables that are assumed to measure similar constructs, without invalidating either instrument as a way of measuring certain aspects of these constructs (see Bornstein, 1999; Ganellen, 1996a; McClelland, Koestner, & Weinberger, 1989; Masling, 1997; Meyer, 1997b). Stricker and Gold (1999) concluded a thorough discussion of this literature by noting that "it is not good science to study the validity of the Rorschach in convergent validity studies that use self-report instruments as target criteria" (p. 242).

From a clinical perspective, judging the validity of Rorschach scales from their correlations with self-report scales has some additional shortcomings. Because these measures differ in the degree to which respondents are consciously aware of what their answers might signify, they differ in their susceptibility to impression management. Respondents' test-taking attitudes and the impression they would like to give are consequently less likely to influence the nature and extent of what they reveal about themselves when they are doing the RIM than when they are doing the MMPI. Meyer (1999b) has presented a detailed analysis of data, demonstrating that the degree to which Rorschach and MMPI scales correlate is moderated by similarities and differences in the style of approach that test respondents bring to their task. Also of note is the common observation that divergence between relatively structured and relatively unstructured measures of similar phenomena can often be very informative, sometimes even more useful than convergence between the measures, and by no means invalidates either finding (see Finn, 1996; Ganellen, 1996b; Weiner, 1993). Such complementary use of tests is considered further in the final section of this commentary, which concerns incremental validity.

Third, in addition to addressing the use of specific scales for specific purposes and correlating these scales with observed variables, validation research with Rorschach scales should focus on particular observed variables that are conceptually linked to these scales in ways that would predict a consistently significant relationship between them. It is not sufficient merely to anticipate that Rorschach scores in general will correlate with personalitydetermined behavior in general. Instead, an informed conceptual approach to assessment research formulates predictions on the basis of identifying particular personality characteristics that are believed to account both for a particular test score that measures it and for a particular behavior that reflects it. A positive finding then goes beyond demonstrating what goes with what, which constitutes criterion validation, and provides a heuristic explanation of why a test score measures what it does, which is the essence of construct validation (see Weiner, 1995). In the absence of sufficient conceptualization to identify criterion variables with which test scales should be expected to correlate, researchers run the risk of giving as much weight to irrelevant as to relevant variables. Failure to avoid irrelevance in the selection of criterion variables can result in penalizing a test for not doing something in the absence of any reason to believe that it should.

Substantive Findings

Turning to the substantive conclusions about validity in these five articles, Stricker and Gold (1999) stated in their review of the empirical literature that, with respect to basic personality structure and the dispositions of people to behave in certain ways, "... in recent years there has been the accumulation of an impressive body of studies that have indicated that the Rorschach and other related methods reliably and validly tap into such processes" (p. 241). Viglione (1999) came to the following similar conclusion in his extensive analysis of published research:

Taking the time to read and to understand the empirical literature in refereed journals over the last 20 years leads to the conclusion that the Rorschach variables are useful for many purposes in clinical, forensic, and educational settings. This conclusion rests on a synthesis of the empirical literature emphasizing ecologically valid, behavioral, real-life criteria... The assumption that the Rorschach is not useful... is mistaken and contrary to the evidence. (p. 260)

A careful reading of Viglione's review and of the primary sources he cites should leave little doubt concerning the soundness of his conclusions. Lending considerable additional support to this conclusion is the Hiller et al. (1999) article mentioned only briefly thus far, which consists of a meta-analytic study, comparing criterion validity evidence for the RIM and the MMPI. Hiller et al. introduced their analysis by reviewing prior efforts of this kind, with particular attention to frequently cited meta-analyses by Atkinson (1986) and by Parker, Hanson, and Hunsley (1988) and to a methodological critique of these meta-analyses by Garb, Florio, and Grove (1998). The Garb et al. (1998) article was subsequently followed by a rebuttal from Parker, Hunsley, and Hanson (1999) and a reply from Garb, Florio, and Grove (1999). Hiller et al. delineated numerous methodological shortcomings in both the earlier meta-analytic reviews and in the Garb et al. (1998) reanalysis of the earlier data, and they described in detail the procedures they followed to correct these flaws and enhance the reliability of their findings.

Hiller et al. (1999) began their meta-analysis with 4,370 MMPI articles and 1,793 Rorschach articles appearing in the literature from 1977–1997. Using random sampling procedures and limiting their database to research studies in which there was at least one external criterion variable and to associations between variables that could be expected to be significant, they eventually examined studies involving 5,007 MMPI protocols and 2,276 Rorschach protocols.

The results they obtained led Hiller et al. (1999) to numerous important conclusions, of which the following four are especially germane to the issue of Rorschach validity:

1. The RIM and the MMPI have equivalent validity effect sizes. Hiller et al. (1999) found unweighted mean validity coefficients of 0.29 for Rorschach variables and 0.30 for MMPI variables, and they were able to demonstrate that these validity estimates for the two instruments do not differ significantly from each other. Hiller et al. commented as follows on the reliability of their findings: "The methodological features of this study, including random sampling from the published literature, expert judgments for inclusion of validity evidence, and the use of accepted effect size estimation techniques, lend greater credibility to these results compared with those from previous efforts" (p. 291).

2. The magnitude of the obtained effect sizes for both the RIM and the MMPI is substantial and warrants user confidence in applying both instruments for their intended purposes. Hiller et al. (1999) referred to Cohen's (1988) observation that correlation coefficients on the order of 0.30 are near-maximum outcomes in relating personality measures to real-life criteria, and they concluded that "validity for these instruments [Rorschach and MMPI] is about as good as can be expected for personality tests" (p. 291).

3. The RIM and the MMPI demonstrate areas of superiority to each other in the validity coefficients they generate. Average effect sizes found by Hiller et al. (1999) indicate that Rorschach variables are somewhat superior to MMPI variables in predicting objective criterion variables, such as unambiguous behavioral outcomes (e.g., treatment attendance) and discrimination of objectively different groups (e.g., patients with closed head injury vs. normal controls). The MMPI, on the other hand, shows somewhat higher effect sizes than the RIM in studies that use psychiatric diagnosis and self-report measures as criterion variables. Hiller et al. noted that shared method variance probably accounts for the relatively high correlations of the MMPI with self-reports and psychiatric diagnoses, which are based largely on self-reports. Likewise, in providing confirmation for the previously referenced conceptual and empirical literature concerning differences between these measures, their data may reflect the particular sensitivity of the RIM to persistent behavioral dispositions.

4. There is little likelihood that unpublished research studies contain findings that would alter the results and implications of this meta-analysis. Hiller et al. (1999) determined on the basis of statistical procedures that thousands of unpublished studies with negative results would have to be hidden away in file drawers to detract substantially from the effect sizes they obtained. Although they acknowledged that "some bias may be present because of unrepresented unpublished studies," they concluded that "the like-lihood that these unpublished studies would reduce the significance of results below traditional levels is very small" (p. 290).

This brings us to the article by Hunsley and Bailey (1999), who concluded with respect to the validity of the RIM that "there is little scientific evidence to support the clinical utility of the Rorschach [and] currently no scientific basis for justifying the use of Rorschach scales in psychological assessments" (p. 266). Hunsley and Bailey even went so far as to say that "the Comprehensive System, as a whole, does not meet the requirements set out in professional standards of practice such as the *Standards for Psychological and Educational Testing* ([APA,] 1985)" (p. 271).

Is it possible that these authors read the same literature that led Stricker and Gold to conclude that accumulating evidence provides an impressive demonstration of Rorschach validity, that led Viglione to conclude that Rorschach variables are useful for many purposes, and that resulted in Hiller et al.'s (1999) obtaining a highly respectable average validity coefficient for Rorschach scales, equivalent to the validity of MMPI scales? Foregoing speculation in this regard, let it be said that Hunsley and Bailey's sweeping indictment of Rorschach assessment as failing to meet professional standards of practice, ignoring as it does abundant evidence to the contrary, is without substance.

As an important case in point, let us examine the evidence cited by Hunsley and Bailey in support of their opinion that the RIM lacks validity. Hunsley and Bailey based their challenge of Rorschach validity largely on three lines of research. First, they found fault with the methodology of the previously mentioned metaanalytic studies by Atkinson (1986) and Parker et al. (1988) and observed that these studies have frequently been credited with providing support for the scientific merit of Rorschach assessment. As far as Hunsley and Bailey are concerned, "it is clear that current assessments of the general validity of the Rorschach rely almost entirely on invalid or extremely limited meta-analytic evidence [and] this meager evidence cannot be used to substantiate the clinical utility of the test" (p. 269). The methodological adequacy of the Parker et al. meta-analysis has also been challenged by Garb et al. (1998), whose opinions as previously noted evoked first a rebuttal (Parker et al., 1999) and then a reply (Garb et al., 1999). In contrast to the just quoted Hunsley and Bailey (1999) position that there is insufficient scientific basis to justify the use of Rorschach scales in psychological assessments, Parker et al. (1999) asserted that the Rorschach "is valid, stable, and reliable under certain circumstances" and that "psychologists should be prepared to accept that there may be domains in which the Rorschach has greater utility than the MMPI" (pp. 291 & 292).

However this may be, the extensive and methodologically sophisticated meta-analysis of Rorschach validity findings reported by Hiller et al. (1999) appear to render moot any disputes about the adequacy of the Atkinson and Parker et al. (1988) methodology. Moreover, the magnitude of the effect sizes found by Hiller et al. for Rorschach scales refutes any assertion that the clinical utility of the Rorschach lacks substantiation by adequate evidence. Hunsley and Bailey (1999) did not have the Hiller et al. data available to them when they prepared their article, and they cannot be faulted for not considering its implications. Nevertheless, the fact remains that the Hiller et al. data weaken considerably their argument that only limited or suspect meta-analytic data are available to support the validity of Rorschach assessment. Of further note in this regard are two other recent meta-analyses that have documented the predictive validity of specific Rorschach scales. Meyer and Handler (1997), examining 20 effect sizes involving 752 participants, found a population correlation of 0.44 between the Rorschach Prognostic Rating Scale and independent ratings of psychological treatment outcomes 1 year later. Bornstein (1999) examined 21 effect sizes, involving 538 respondents and found a validity coefficient of 0.37 in using the Rorschach Oral Dependency scale to predict independently observed dependency-related behaviors.

Second, Hunsley and Bailey (1999) argued that unresolved issues, concerning the effect of response frequency (R) on obtained scale scores, undermine any potential validity that the RIM might otherwise demonstrate. If this were indeed the case, then logical reasoning indicates how remarkable it is that the Rorschach has demonstrated substantial effect sizes in the total 1977-1997 corpus of research studies, despite such a handicap; logic suggests further that Rorschach validity coefficients would be even higher than those found by Hiller et al. if this imputed problem with R were corrected. As for the problem itself, however, there do remain some issues to be resolved. On the one hand, recent work by Meyer (1999a), using refined statistical modeling procedures, has demonstrated that R by itself does not moderate the convergent validity of Rorschach scales with the MMPI and that prior findings that suggested otherwise were the result of sampling error. On the other hand, Meyer (1998) has also reported an average correlation of .25 between R and 33 criterion scores that contribute to various Comprehensive System indices. Additionally, Weiner (1998a, p. 109) has observed that short records with 14 to 19 responses may be valid as far as they go but, as a consequence of the respondent's guardedness, may not fully reveal the extent of either the person's capacities or his or her concerns. Further research is needed to elucidate appropriate ways of taking unusual protocol length into consideration in the interpretive process.

Third, Hunsley and Bailey (1999) gave considerable weight to research, showing limited convergence between Rorschach and MMPI scales, from which they infer limited convergent validity on the part of the RIM. As noted earlier, attempts to validate one inferential personality measure against another such measure, in the absence of observable criterion variables, and especially to make such an attempt in the face of considerable method variance between them, constitutes simplistic and largely uninformative methodology. Hunsley and Bailey acknowledged these considerations to some extent and commented in particular on the differences in the level of conscious awareness with which respondent's give their answers to different kinds of tests. However, they gave no ground in this regard, concluding instead that "claims that the Rorschach's low-convergent validity reflects its virtues rather than its limitations must be supported by empirical evidence, not simply rhetoric" (p. 270).

Such empirical evidence is ready to hand in the Hiller et al. (1999) meta-analysis, which as previously mentioned identifies types of real-life criterion variables that are predicted better by Rorschach than by MMPI scales. Should one still wish to embrace Rorschach–MMPI correlations as a valuable index of validity, why not conclude that limited convergence between these measures reflects badly on the validity of the MMPI? The answer to this question touches on the issue of the incremental validity of Rorschach scales, which is discussed by Hunsley and Bailey, Stricker and Gold (1999), and Viglione, and which is the central focus of the article in this Special Series by Dawes (1999).

Incremental Validity

The incremental validity of an assessment instrument consists of the extent to which its scores are likely to increase the accuracy of predictions derived from other sources of information, including scores on other tests. Not surprisingly, the articles under discussion arrive at quite different conclusions in this matter. Consistent with their blanket disbelief in the validity of the RIM, Hunsley and Bailey (1999) gave short shrift to its incremental validity: "The limited evidence bearing on this question to date does not support, in general, the incremental validity of Rorschach scales" (p. 271). They cited two sources of information in support of this conclusion. First, they referred to a conclusion reached by Garb (1984) that Rorschach data do not add to the accuracy of personality assessments based on demographic or self-report data. They continued on to acknowledge that none of the studies reviewed by Garb involved use of the Comprehensive System. Given that the Comprehensive System is the only adequately standardized and the most widely used Rorschach method, how can a 1984 review of research exclusive of Comprehensive System studies be taken seriously in 1999 as a basis for casting aspersions on the incremental validity of the Rorschach? Garb (1998, p. 20) has recently restated his conclusion that Rorschach assessment shows little or no incremental validity, but the publication dates of the references he cites in support of his opinion range from 1954 to 1982 and have no bearing whatsoever on the Comprehensive System.

The only other basis cited by Hunsley and Bailey for their negative opinion of the Rorschach's incremental validity comprises two studies in which the diagnostic efficiency of MMPI scales was reportedly not enhanced by combining them with Comprehensive System Rorschach indices (Archer & Gordon, 1988; Archer & Krishnamurthy, 1997). A close look at these two studies suggests that scholars should not be too quick to interpret them as challenging the utility of Rorschach indices. As shown in Table 2 of the Archer and Gordon study, the traditional cutoff score of >3 on the Rorschach Schizophrenia Index achieved a 69% hit rate in differentiating schizophrenic from nonschizophrenic patients in their sample, whereas the traditional cutoff score of >69 on MMPI Scale 8 (Schizophrenia [Sc]) achieved 60% accuracy. When more conservative cutoff scores were examined, the classification accuracy increased to 80% for the Schizophrenia Index > 4 and to 76% for Sc > 74. If there is any difference here, the Rorschach scale outperformed the MMPI scale, and it would seem legitimate to ask at which doorstep the problem should be laid if using both tests together did not enhance classification accuracy. As for the Archer and Krishnamurthy study, these authors found that the Rorschach Depression Index did not add incrementally to the MMPI in predicting a diagnosis of depression. On the other hand, their data indicate that two components of the Depression Index, the Vista response and the Affective Ratio, did enhance the identification of depression over and above information provided by the MMPI.

Dawes (1999) has contributed a methodological article to this series in which he describes and illustrates two data analysis techniques for assessing the incremental validity of a Rorschach variable. Using the Ego Impairment Index (EII; Perry, Viglione, & Braff, 1992) as his example, he found that it correlated significantly with severity of psychopathology, as operationalized by several diagnostic categories. It is noteworthy that Dawes, who has been widely quoted for writing in 1994 that the Rorschach is a "shoddy" instrument that "is not a valid test of anything" (Dawes, 1994, pp. 123 & 146), has demonstrated with his data analysis that it is very good indeed for something important and as good as the MMPI. Table 1 in Dawes's article shows correlations between the EII and two types of ratings for severity of psychopathology of .35 and .20, both significant at the 1% level of confidence. The table also shows almost identical correlations of .31 and .27 between these ratings and average scale elevation on the MMPI. The MMPI Goldberg index for differentiating between psychotic and neurotic disorder fared less well than the Rorschach EII in this analysis, with correlations of .17 and -.03 with the two ratings of severity of psychopathology.

With respect to incremental validity, however, Dawes (1999) concluded that the EII "had only slight incremental validity over and above the number of responses and form quality, and even less when the average MMPI elevation and L. R. Goldberg's (1965) formula for predicting psychosis versus neurosis were entered before these Rorschach variables" (p. 297). This conclusion is difficult to reconcile with the data shown in Table 3 of the Dawes article. When the R^2 values in this table are unsquared to indicate the multiple Rs that obtain, the prediction of maximum diagnostic severity from combining average clinical scale elevations and the Goldberg Index for the MMPI reaches .37. When Rorschach response total, form quality, and the EII are added, the R with maximum diagnostic severity increases to .49. This increased predictability from an R of .37 to an R of .49 is not in my opinion an instance of "slight incremental validity." Rather, Dawes's data appear to demonstrate substantially increased validity, with the Rorschach variables adding a respectable amount of information beyond what could be obtained from the clinical scales of the MMPI.

On the other side of this debate, Stricker and Gold (1999) and Viglione (1999) cited numerous research studies involving many different Rorschach scales and diverse criterion variables to warrant the conclusion that Rorschach assessment has considerable incremental validity and provides valuable information in personality evaluations beyond what is learned from interview and self-report methods. As a current example in this regard, Perry, Geyer, and Braff (1999) reported a study in which Rorschach indices of impaired reality testing (low X-%) and disordered thinking (elevated WSum6) predicted physiological correlates of schizophrenia (prepulse inhibition of the startle response) significantly and over and above predictions from scores derived from two structured

interviews, the Scale for the Assessment of Positive Symptoms and Scale for the Assessment of Negative Symptoms.

Stricker and Gold (1999) and Viglione (1999) also both argued that the determination of whether personality assessment instruments are incrementally valid should go beyond nomothetic approaches, which have implications for group averages and tendencies, and include as well idiographic approaches, which have implications for individual case conceptualization. To elaborate this nomothetic-idiographic distinction in the present context, let it be noted that Rorschach critics who have considered the instrument lacking in incremental validity have typically based their conclusions on nomothetic data analysis, as illustrated by MMPI-Rorschach studies to which Hunsley and Bailey referred and by the methodology proposed by Dawes. Lack of incremental validity from this nomothetic perspective means that combining one test with another, as in adding the RIM to an MMPI, or vice versa, does not significantly increase the percentage of a participant sample that is correctly classified as having some characteristic or condition.

Whatever import may attach to such a nomothetic lack of incremental validity, it bears little relevance to the assessment of psychological functioning in individual patients and clients seen in clinical practice. Most assessment clinicians who use the RIM and MMPI conjointly are familiar with cases in which (a) a normal range MMPI occurs together with Rorschach indications of substantial deviations from normative personality functioning (a common occurrence in custody, personnel selection, and other administrative evaluations in which respondents are trying to make a favorable impression); or (b) a valid but guarded Rorschach protocol, providing minimal information occurs together with a valid and highly informative MMPI (a common occurrence when respondents are disinclined to reveal very much about themselves on an open-ended task but are sufficiently cooperative to answer direct "yes" or "no" questions honestly). Generally speaking, impression management of both kinds, whether involving guardedness or an effort to "look good," increases the likelihood of seeming incongruities between Rorschach and MMPI results, whereas openness on the part of respondents is likely to promote congruent findings. Clinicians who examine respondents motivated by impression management are likely to learn substantially more about their personality characteristics from a conjoint RIM-MMPI assessment than they would have learned from using either test by itself (see Ganellen, 1996b).

Even if these two patterns of RIM-MMPI results occur only occasionally in clinical assessments, they send a clear message to practitioners who are trying to be helpful to their patients. There is good reason to believe that, at least in some cases, either the RIM or the MMPI may yield apparently valid data that are nevertheless misleading or incomplete, and there may be little way to anticipate or determine before the fact whether this will be the case. Clinicians who ill-advisedly restrict their personality testing to either just the RIM or just the MMPI will never know whether the other instrument would have given them complementary bits of information that could have sharpened their formulations and enhanced the utility of their conclusions and recommendations. Such useful complementarity may also involve Rorschach and MMPI findings that initially seem contradictory but begin to make sense after one ponders their implications in the individual case. These idiographic considerations in case conceptualization obtain independently of whatever level of incremental validity for Rorschach variables is demonstrated on a nomothetic basis.

These observations touch on the general topic of multimethod assessment, which has obvious implications for including the RIM in clinical test batteries but goes beyond the intended focus of the present commentary. Suffice it to say that, contrary to Garb's (1984) often cited indictment of test batteries for failure to demonstrate incremental validity for their component measures, a recent review by Meyer et al. (2001) adduced logical and empirical considerations that affirm the utility of multimethod batteries as a means of maximizing assessment validity:

We believe that there is a direct parallel between empirical research and applied clinical practice on this issue. In research *monomethod* bias and *monooperation* bias are critical threats to the validity of any investigation... The optimal methodology to enhance the construct validity of nomothetic research consists of combining data from multiple methods and multiple operational definitions... Just as effective nomothetic research recognizes how validity is maximized when variables are measured by multiple methods, particularly when the methods produce meaningful discrepancies ... the quality of idiographic assessment can be enhanced by clinicians who integrate the data from multiple methods of assessment. (p. 150)

Conclusion

Three of the five Special Series articles discussed in this commentary present extensive empirical evidence that the RIM has been standardized, normed, made reliable, and validated in ways that exemplify sound scientific principles for the development of an assessment instrument. The authors of the other two articles present sketchy data and incomplete literature reviews on the basis of which they question the psychometric foundations of Rorschach assessment and assert that adequate validating data to demonstrate its utility have not yet been generated. Scientific debate often proceeds with one side arguing that something is so or something works, whereas the other side argues that it is not so or does not work. If the proponents base their arguments on accumulating evidence of its efficacy, and the opponents base their arguments on being unconvinced by this evidence, there comes a time when being unconvinced exceeds the boundaries of appropriate skepticism. Those who take issue with the abundant and compelling evidence that Rorschach assessment works very well indeed for certain purposes for which it is intended must sooner or later present equally abundant and compelling evidence that it does not work for these purposes or else put their case to rest.

On the other hand, contemporary critics of Rorschach assessment are raising some appropriate concerns about the psychometric status of the instrument, and those expressed in this series of articles include (a) the need to promote and document, if possible, accurate response coding among Rorschach practitioners as well as researchers; (b) the importance of resolving uncertainties that involve the impact of response total (R) on various summary scores and their interpretation; and (c) the need for extending well-designed validity studies to many Rorschach variables that have thus far not been adequately examined. Along with recognizing the legitimacy of such concerns and their implications for continued commitment to good Rorschach training and research, Rorschach advocates should be aware of two additional limitations in present knowledge that also call for their attention. First, as

noted in the present article, the available normative data for Rorschach variables may not be sufficiently representative of the current United States population and does not yet fully encompass possible cross-cultural differences. Second, very few longitudinal data are available for either nonpatient or patient populations from which life cycle developmental changes and changes in clinical status over time can be adequately examined. Generalizability across diverse groups of people and expected changes over time are two areas of limited data that the RIM shares with most personality assessment instruments. Closing such gaps in our knowledge in the years ahead, aided by enhancements in data processing technology and global communication, should continue to enhance the utility and luster of Rorschach assessment and other assessment methods as well.

References

- American Psychological Association (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.
- 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.
- Archer, R. P., & Krishnamurthy, R. (1997). MMPI-A and Rorschach indices related to depression and conduct disorder. *Journal of Person*ality Assessment, 69, 517-533.
- Atkinson, L. (1986). The comparative validities of the Rorschach and MMPI: A meta-analysis. *Canadian Psychologist*, 27, 238-247.
- Bornstein, R. F. (1999). Criterion validity of objective and projective dependency tests: A meta-analytic assessment of behavioral prediction. *Psychological Assessment*, 11, 48–57.
- Butcher, J. N., Nezami, E., & Exner, J. E., Jr. (1998). Psychological assessment in cross-cultural settings. In S. Kazarian & D. R. Evans (Eds.), *Cultural clinical psychology* (pp. 61–105). New York: Oxford.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum.
- Dawes, R. N. (1994). House of cards: Psychology and psychotherapy built on myth. New York: Free Press.
- Dawes, R. N. (1999). Two methods for studying the incremental validity of a Rorschach variable. *Psychological Assessment*, 11, 297-302.
- Erdberg, S. P., & Shaffer, T. W. (1999, July). An international symposium on Rorschach nonpatient data: Findings from around the world. Symposium conducted at the XVI International Congress of Rorschach and Projective Methods, Amsterdam.
- Exner, J. E., Jr. (1991). The Rorschach: A comprehensive system. Vol. 2. Interpretation (2nd ed.). New York: Wiley.
- Exner, J. E., Jr. (1993). The Rorschach: A comprehensive system. Vol. 1. Basic foundations (3rd ed.). New York: Wiley.
- Exner, J. E., Jr. (1995). A Rorschach workbook for the Comprehensive System (4th ed.). Asheville, NC: Rorschach Workshops.
- Exner, J. E., Jr., Thomas, E. A., & Mason, B. (1985). Children's Rorschachs: Description and prediction. *Journal of Personality Assessment*, 49, 12–20.
- Exner, J. E., Jr., & Weiner, I. B. (1995). The Rorschach: A comprehensive system. Vol. 3. Assessment of children and adolescents (2nd ed.). New York: Wiley.
- Finn, S. E. (1996). Assessment feedback integrating MMPI-2 and Rorschach findings. Journal of Personality Assessment, 67, 543–557.
- Gacono, C. R., & Meloy, J. R. (1994). The Rorschach assessment of aggressive and psychopathic personalities. Hillsdale, NJ: Erlbaum.
- Ganellen, R. J. (1996a). Exploring MMPI–Rorschach relationships. Journal of Personality Assessment, 67, 529–542.
- Ganellen, R. J. (1996b). Integrating the Rorschach and the MMPI in personality assessment. Mahwah, NJ: Erlbaum.

- Garb, H. N. (1984). The incremental validity of information used in personality assessment. *Clinical Psychology Review*, 4, 641-655.
- Garb, H. N. (1998). Studying the clinician: Judgment research and psychological assessment. Washington, DC: American Psychological Association.
- Garb, H. N., Florio, C. M., & Grove, W. M. (1998). The validity of the Rorschach and the Minnesota Multiphasic Personality Inventory: Results from meta-analyses. *Psychological Science*, 9, 402-404.
- Garb, H. N., Florio, C. M., & Grove, W. M. (1999). The Rorschach controversy: Reply to Parker, Hunsley, and Hanson. *Psychological Science*, 10, 293–294.
- Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T. R., & Brunell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment*, 11, 278-296.
- Hilsenroth, M. J. (2000). Rorschach test. In A. E. Kazdin (Ed.), *Encyclopedia of psychology* (pp. 117–119). Washington, DC: American Psychological Association.
- Hilsenroth, M. J., Fowler, J. C., Padawer, J. R., & Handler, L. (1997). Narcissism in the Rorschach revisited: Some reflections on empirical data. *Psychological Assessment*, 9, 113–121.
- Hunsley, J., & Bailey, J. M. (1999). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment*, 11, 266-277.
- Masling, J. M. (1997). On the nature and utility of projective tests and objective tests. *Journal of Personality Assessment*, 69, 257-270.
- McCann, J. T. (1998). Defending the Rorschach in court: An analysis of admissibility using legal and professional standards. *Journal of Person*ality Assessment, 70, 135-144.
- McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review*, 96, 690-702.
- McDowell, C., & Acklin, M. W. (1996). Standardizing procedures for calculating Rorschach interrater reliability. *Journal of Personality As*sessment, 66, 308-320.
- Meloy, J. R., Acklin, M. W., Gacono, C. B., Murray, J. F., & Peterson, C. A. (Eds.). (1997). Contemporary Rorschach interpretation. Mahwah, NJ: Erlbaum.
- Meyer, G. J. (1997a) Assessing reliability: Critical corrections for a critical examination of the Rorschach Comprehensive System. *Psychological Assessment*, 9, 480-489.
- Meyer, G. J. (1997b). On the integration of personality assessment methods: The Rorschach and MMPI. *Journal of Personality Assessment*, 68, 297-330.
- Meyer, G. J. (1997c). Thinking clearly about reliability: More critical corrections regarding the Rorschach Comprehensive System. *Psychological Assessment*, 9, 495–498.
- Meyer, G. J. (1998). Error in research and assessment data with an erratum for Meyer (1993). Journal of Personality Assessment, 71, 195–211.
- Meyer, G. J. (1999a). Convergent validity of the MMPI and Rorschach scales: An extension using profile scores to define response and character styles on both methods and a reexamination of simple Rorschach response frequency. *Journal of Personality Assessment*, 72, 1–35.
- Meyer, G. J. (1999b). Introduction to the special series on the utility of the Rorschach for clinical assessment. *Psychological Assessment*, 11, 235– 239.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies,

R. R., Eisman, E. J., & Kubiszyn, T. W. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128-165.

- Meyer, G. J., & Handler, L. (1997). The ability of the Rorschach to predict subsequent outcome: Meta-analysis of the Rorschach Prognostic Rating Scale. Journal of Personality Assessment, 69, 1–38.
- Parker, K. C. H., Hanson, R., & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin*, 103, 367–373.
- Parker, K. C. H., Hunsley, J., & Hanson, R. K. (1999). Old wine from old skins sometimes tastes like vinegar: A response to Garb, Florio, and Grove. *Psychological Science*, 10, 291-292.
- Perry, W., Geyer, M. A., & Braff, D. L. (1999). Sensorimotor gating and thought disturbance measured in close temporal proximity in schizophrenic patients. Archives of General Psychiatry, 56, 277–281.
- Perry, W., Viglione, D. J., & Braff, D. (1992). The Ego Impairment Index and schizophrenia: A validation study. *Journal of Personality Assess*ment, 59, 165-175.
- Rorschach, H. (1942). Psychodiagnostics: A diagnostic test based on perception. Bern, Switzerland: Hans Huber. (Original work published in 1921)
- Smith, B. L. (1994). Object relations theory and the integration of empirical and psychoanalytic approaches to Rorschach interpretation. *Ror*schachiana, 19, 61–77.
- Stricker, G., & Gold, J. R. (1999). The Rorschach: Towards a nomothetically based, idiographically applicable, configurational model. *Psychological Assessment*, 11, 240-250.
- Viglione, D. J. (1999) A review of recent research addressing the utility of the Rorschach. Psychological Assessment, 11, 251-265.
- Weiner, I. B. (1993). Clinical considerations in the conjoint use of the Rorschach and the MMPI. Journal of Personality Assessment, 60, 142– 148.
- Weiner, I. B. (1995). Methodological considerations in Rorschach research. Psychological Assessment, 7, 330-337.
- Weiner, I. B. (1996). Some observations on the validity of the Rorschach Inkblot Method. *Psychological Assessment*, 8, 206–213.
- Weiner, I. B. (1998a). Principles of Rorschach interpretation. Mahwah, NJ: Erlbaum.
- Weiner, I. B. (1998b). Teaching the Rorschach Comprehensive System. In L. Handler & M. J. Hilsenroth (Eds.), *Teaching and learning personality* assessment (pp. 215-233). Mahwah, NJ: Erlbaum.
- Weiner, I. B. (1999). Contemporary perspectives on Rorschach assessment. European Journal of Psychological Assessment, 15, 78-96.
- Wenar, C., & Curtis, K. M. (1991). The validity of the Rorschach for assessing cognitive and affective changes. *Journal of Personality As*sessment, 57, 291–308.
- Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science*, 7, 3–10.
- Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1997). The reliability of the Comprehensive System: A comment on Meyer (1997). Psychological Assessment, 9, 490–494.

Received May 24, 1999

Revision received October 11, 1999

Accepted May 10, 2001