

Whither the Rorschach? An Analysis of the Evidence

John Hunsley
University of Ottawa

J. Michael Bailey
Northwestern University

In the previous Special Section, the authors presented empirical evidence and logical analysis that were sufficient to demonstrate that the widespread use of the Rorschach in clinical, legal, forensic, and occupational settings is unwarranted on both scientific and ethical grounds (J. Hunsley & J. M. Bailey, 1999). To expand on their analysis and to respond to issues raised in the previous and current Special Sections, they begin their article by examining a number of conceptual issues that are at the heart of the disagreements about the Rorschach. The focus is then shifted to the central issue of clinical utility, with an emphasis on why current research is insufficient to demonstrate the utility of the Rorschach. Next, the psychometric issues raised by Weiner (2001) are addressed and an alternative perspective on the psychometric viability of the Rorschach is provided. Finally, the authors conclude with some suggestions for future directions that must be taken in research to address the substantive concerns raised by Rorschach critics.

The Special Section in the Special Series on the clinical utility of the Rorschach inkblot test, published in *Psychological Assessment* in 1999, was a major event in the ongoing debate about the scientific merits of the Rorschach. For the first time in a single forum, advocates and critics of the test marshaled their evidence and presented their cases for and against the Rorschach. Also for the first time, advocates and critics had an opportunity, in this second Special Section, to respond to the positions that were put forward and to evaluate the strength and consistency of the arguments for and against the clinical use of the Rorschach. Although the conclusions offered in both Special Sections are unlikely to resolve the Rorschach debate, we hope that these presentations will influence current perspectives on the test and will be instrumental in determining what subsequent generations of clinical psychologists are taught about this controversial test.

In the previous Special Section, we had an opportunity to present both empirical evidence and logical analysis that we believed were sufficient to demonstrate that the widespread use of the Rorschach in clinical, legal, forensic, and occupational settings is unwarranted on both scientific and ethical grounds (Hunsley & Bailey, 1999). Having reviewed the evidence presented by Hiller, Rosenthal, Bornstein, Berry, and Brunell-Neulieb (1999) and Dawes (1999) and the arguments proffered by Stricker and Gold (1999), Viglione (1999), and Weiner (2001), we have found little reason to alter this position. Indeed, if anything, Garb, Wood, Nezworski, Grove, and Stejskal's (2001) detailed review of the evidence presented in the first Special Section has served to amplify our concerns about the scientific status of the Rorschach.

We do believe, however, that the articles in these Special Sections have revealed a number of conceptual issues that are at the heart of the disagreements about the Rorschach, and it is these issues to which we turn first in our commentary. As Garb et al. (2001) have provided an exemplary critique of the empirical evidence presented in the first Special Section, we refrain from reiterating the concerns they raised except when they are directly germane to the themes we are discussing. We next focus on the central issue of clinical utility and illustrate why current research is insufficient to demonstrate the utility of the Rorschach. Following this, we address the important psychometric issues raised by Weiner (2001) and provide an alternative perspective on the psychometric viability of the Rorschach and Exner's Comprehensive System (CS). We conclude our presentation with some suggestions for future directions that must be taken in Rorschach research to address the substantive concerns raised by Rorschach critics.

A Test (or a Method?) in Search of Theory

Numerous theories have been proposed to account for how Rorschach responses provide valid data on both the inner life and observable behaviors of a person. To date, however, no one theory has the support of all users of the Rorschach. Exner (1989) rejected the projective hypothesis proposed by Frank (1948), which suggested that responses to the Rorschach are totally determined by internal emotional reactions and experiences. Instead, Exner proposed that some responses are informative about the manner in which an individual conceptualizes the world based on that individual's psychological makeup, whereas others are largely determined by the stimulus characteristics of the inkblots. However, there is little or no theoretical justification offered for most of the scales included in the CS or in other scoring systems (Costello, 1998; Kleiger, 1992)—why, for example, pairs and reflections should be related to narcissism, texture responses should be related to affective elements of interpersonal relationships, or color should be related to affective control. Exner's theoretical position, in turn,

John Hunsley, School of Psychology, University of Ottawa, Ottawa, Ontario, Canada; J. Michael Bailey, Department of Psychology, Northwestern University.

Correspondence concerning this article should be addressed to John Hunsley, School of Psychology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada. Electronic mail may be sent to hunch@uottawa.ca.

has been rejected by those who argue that the Rorschach is not a psychometric test at all, and various hermeneutic, experiential, and psychodynamic theories have been offered to account for the alleged clinical richness of the Rorschach (e.g., Aronow, Reznikoff, & Moreland, 1995; Costello, 1999; Te'eni, 1998).

From the perspective of scientific progress, a lack of agreement on the theoretical underpinnings of a phenomenon may not be problematic and, indeed, may even be an indication of the vitality and creativity of the research in an area.¹ In contrast to the position taken by Stricker and Gold (1999), Exner has argued persuasively that an overarching Rorschach theory is not necessary for the Rorschach test to demonstrate validity (Exner, 1997). Regardless of the epistemological stance taken by proponents, the lack of consensus among Rorschach scholars about whether the Rorschach is a test or not (e.g., Exner, 1997; Weiner, 2001) is of concern in clinical practice, for at least two reasons.

First, as there is no consensus among the community of Rorschach proponents about what the Rorschach actually is and why Rorschach responses are believed to provide data on psychological functioning, it is extraordinarily easy for criticisms of the Rorschach as a psychological test to be dismissed out of hand by various Rorschach proponents. Critiques of the psychometric properties of the test are viewed as either (a) irrelevant by those proponents who do not consider the Rorschach to be a test that can be judged by empirical evidence or (b) an indication of the need to eschew quantifying Rorschach test data (cf. Aronow et al., 1995; Costello, 1999; Te'eni, 1998). Most disconcertingly, for some proponents, evidence presented by those critical of the Rorschach's validity is simply ignored as the disrespectful and uninformed comments of those who do not value the Rorschach (e.g., Lerner, 2000). Thus, the lack of consensus on what exactly the Rorschach is and on a general theory of Rorschach responding may impede scientific progress in addressing the issues raised by either Rorschach advocates or critics.

The second reason to be concerned about the lack of consensus on the nature of the Rorschach has to do with the potential for erroneous presentations of the scientific validity of the Rorschach. Weiner (e.g., 1997) has previously argued that, based on accumulated empirical evidence, the CS for scoring the Rorschach and other specific scales not included in the CS should be considered scientifically sound. Weiner (2001), however, frequently moved between discussing the CS and the Rorschach Inkblot Method (RIM), which may lead to some confusion. By RIM, he meant the use of the Rorschach, whether or not it is administered and scored according to CS requirements. Thus, the psychometric adequacy of the RIM, as defined by Weiner, cannot rest on the qualities of the CS or other scales he believed to have empirical support. Rather, the psychometric adequacy of each and every summary score, scale, or qualitative interpretation based on the RIM must be demonstrated. Even a psychometrically sound CS cannot provide evidence for the RIM.

Unfortunately, this crucial distinction between the CS scales (and other scales that may have some empirical support) and the RIM may not always be made. One example will suffice to illustrate our point. After reviewing criteria for the admissibility of Rorschach data in court, McCann (1998) recently concluded that, consistent with Weiner's position, any forensic use of the Rorschach should rely primarily on the CS as, in his opinion, only the

CS had accumulated sufficient evidence of scientific validity. However, the extent to which such cautions are followed in forensic practice is simply unknown. In two recent surveys of forensic psychologists, questions were asked about the frequency of Rorschach use (Boccaccini & Brodsky, 1999) and about court admissibility of Rorschach evidence (Weiner, Exner, & Sciarra, 1996). However, in both surveys, it appears that the researchers only asked about Rorschach use and did not ask respondents to indicate the extent to which their forensic Rorschach use was based on the CS. The potential for category errors (i.e., equating the CS and the RIM) in the use of the Rorschach is not just semantic hair splitting—it has real-world implications. If court testimony is allowed on the basis of the scientific status of the CS, but the actual client data presented in court was not derived from the CS, then the contention that the data have scientific validity is unlikely to be tenable. For such reasons it is imperative that Rorschach scholars and those who use the Rorschach consistently and clearly differentiate between the CS, other empirically based scales, and idiographic interpretations of Rorschach data.

To return to the issues raised in this Special Series, the multifaceted nature of the Rorschach was clearly viewed as a strength of the test by Stricker and Gold (1999) and Weiner (2001), for, they contended, an enormous range of theories could be applied to the test data and almost any therapeutic orientation could be used to meaningfully interpret Rorschach data. In short, where we see the potential for confusion, bias, and inattention to scientific data, they see the potential for richness and complexity. This enormous difference aside, we fully concur with Stricker and Gold that future research must rely much more on theoretically informed construct validation. As they suggested, this requires a clear operationalization of the construct and its nomological net, informed selection of samples and criterion variables, and the collection of high-quality data. As a starting point for such an endeavor, researchers may wish to consider attempting a theoretically informed construct validation of "underlying depressive structures," one of the Rorschach constructs mentioned by Stricker and Gold (1999, p. 242). Stricker and Gold suggested that this concept, distinct from the overt manifestations of depression, is what is assessed by Rorschach data and that this accounts for the lack of correspondence between the Rorschach index of depression and actual depressive symptoms and diagnosis. Much would be gained from a well-conducted construct validation of underlying depression: The findings would be, of course, very important, but perhaps more importantly such a study could serve as a model for how to conduct theoretically informed Rorschach research on an untested construct.

What Constitutes an Appropriate Focus for Rorschach Validation?

There is clear and consistent agreement among Rorschach advocates and critics alike in these two Special Sections that the Rorschach should be treated as a psychometric test that must meet the professional standards for such tests. Stricker and Gold (1999)

¹ We are indebted to the reviewer who brought this point to our attention.

and Weiner (2001) explicitly endorsed the need for appropriate norms and for evidence of test validity, although, as we have indicated, not all Rorschach proponents will agree with their perspectives on the Rorschach. The agreement among critics and at least some proponents can serve to focus discussions on what constitutes an appropriate focus for evaluating the validity of Rorschach scales. It is to this question to which we turn.

On the basis of the results of prior research, Viglione (1999) allegedly eschewed the use of self-report data in his review of Rorschach validity—it is puzzling, therefore, that he chose to include results of studies that use self-report data when they provided support for the validity of the Rorschach scales (see, e.g., his discussions of Cooper, Perry, & O'Connell, 1991; Hilsenroth, Handler, Toman, & Padawer, 1995; Hirschfeld, Klerman, & Gough, 1977; Perry & Viglione, 1991). Such inconsistencies aside, his position on the relevance of self-report data vis-à-vis the Rorschach is consistent with the stance taken by Stricker and Gold (1999) and Weiner (2001), who clearly suggested that self-report data should not be used in studying the convergent validity of the Rorschach (see also Ganellen, 1996, and Viglione, 1996). The consistency of this position, evidenced by all Rorschach proponents who contributed to the Special Series, would seem to suggest that the question of the relevance (or irrelevance) of self-report data has been clearly settled among Rorschach researchers. Such a conclusion would be unwarranted, however, as other Rorschach researchers have not been so reluctant to include self-report data as evidence for Rorschach scale validity. For example, in his review of the construct validity of the Rorschach Oral Dependency scale (ROD), Bornstein (1996) reported correlations between the ROD and self-report measures as part of the validity evidence available for the scale. Other researchers, such as Meyer (1999), continue to search for the conditions under which data from the Rorschach and Minnesota Multiphasic Personality Inventory (MMPI) can be expected to show a nontrivial association. Moreover, in the Hiller et al. (1999) meta-analysis, the weighted-mean effect size for the Rorschach and self-report data was $r = .28$, one of the larger effect sizes found for the Rorschach.² Perhaps the question of using self-report data in validating the Rorschach needs to be revisited by Rorschach proponents.

As the Rorschach is so commonly used by clinical psychologists and a number of the CS scores carry the labels of diagnostic categories, it seems reasonable to assume that psychiatric diagnosis would be an appropriate criterion for validation studies of at least some Rorschach scales. Indeed, there are many studies published in the past 30 years that have examined the link between psychopathology and Rorschach responses. Viglione (1999) concluded that the Rorschach has limitations in its use as a diagnostic instrument, which is a rather generous statement in light of a recent review of the validity of the Rorschach in examining psychiatric diagnoses, by Wood, Lilienfeld, Garb, and Nezworski (2000). Wood et al. (2000) examined studies that included diagnoses such as schizophrenia, depression, posttraumatic stress disorder, dissociative identity disorder, dependent personality disorder, narcissistic personality disorder, borderline personality disorder, antisocial personality disorder, and psychopathy. They found evidence that deviant verbalizations and bad form (i.e., providing responses that did not fit the nature of the inkblot) were associated with schizophrenia and, to some extent, with bipolar disorder and schizotypal

personality disorder. Similarly, they found evidence that patients with borderline personality disorder also tended to give more deviant verbalizations than did nondisordered individuals. As individuals with these conditions often make bizarre or unusual statements, it is hardly surprising that this also occurs when administered the Rorschach. Wood et al. found no strong evidence that the Rorschach could consistently detect depression, posttraumatic stress disorder, antisocial personality disorder, or any other psychiatric diagnoses. This is consistent with the results of the Hiller et al. (1999) meta-analysis that found a weighted-mean correlation of only .18 when psychiatric diagnoses were used as the criterion in Rorschach validity studies. Part of the reason for this low association may be that, in some instances, Rorschach researchers may have relied inappropriately on clinical diagnosis as a criterion against which to judge the Rorschach's validity.³ Many researchers use Rorschach data drawn from clinical practice files. Therefore, the ease of accessing diagnostic information may

² Throughout this article, we report only the weighted-mean effect sizes from Hiller et al. (1999), for four reasons. First, as Garb et al. (2001) noted, the use of weighted-mean effect sizes, not unweighted-mean effect size, is consistently recommended by meta-analysts (e.g., Hedges & Olkin, 1985). The strategy of weighting studies by their sample sizes correctly assigns proportionally more emphasis to studies that have larger samples and that, therefore, provide more precise estimates of the population effect size. Second, the assumption of homogeneity of variance is unlikely to be met when unweighted effect sizes are used and the resulting estimate of the population effect size is likely to be inflated (e.g., Weisz, Weiss, Han, Granger, & Morton, 1995). Third, the decision taken by Hiller et al. to pool the data from the studies they sampled suggests that the authors believed that the selected studies all provided data relevant to estimating the population effect sizes for the general validity of the MMPI and Rorschach. Thus, their suggestion that unweighted effect sizes were preferable to weighted effect sizes in their analyses is inherently inconsistent with their initial conceptualization of the focus of the meta-analysis and with their own sampling strategy. Fourth, Hiller et al.'s (1999) contention that unweighted values should be used because four of the selected MMPI studies with large samples were focused on "particularly easy" (p. 292) validation tasks is flawed. In the studies to which they refer, all of the validation tasks dealt with key issues of concurrent and criterion validity (i.e., can the validity scales detect dissimulation [Bagby, Buis, & Nicholson, 1995], does an MMPI scale correlate highly with informant behavioral ratings [Ben-Porath, Hostetler, Butcher, & Graham, 1989], can MMPI indices of maladjustment, psychotic symptoms, and neurotic symptoms distinguish among psychiatric inpatients, psychiatric outpatients, medical patients, and undergraduates [Friedman, Gleser, Smeltzer, Wakefield, & Schwartz, 1983], and does a MMPI scale correlate highly with a self-report measure of the same construct [Morton-Page & Wheeler, 1997]). The logic of purposefully underweighting the results of these studies precisely because the researchers addressed highly salient and clinically relevant hypotheses with strong research designs is extremely difficult to comprehend. Overall, though, despite our (relative) preference for reporting the weighting values, we are skeptical about the accuracy of both the weighted and the unweighted values. The primary reason for our skepticism is that we doubt that such a small sample of studies is likely to provide unbiased estimates of the Rorschach's or MMPI's validity vis-à-vis the validation criteria selected by Hiller et al. Given the hundreds of studies published on these instruments in the past 20 years, it is highly unlikely that a sample of 30 studies is sufficient to generate accurate population estimates.

³ This was suggested by a reviewer of this article.

have provided readily obtained but ultimately inappropriate criterion variables. As in other research domains, progress in Rorschach research is most likely to occur when careful consideration is given to the selection, measurement, and analysis of constructs.

Nevertheless, as the Rorschach has been touted for decades as a key test for use in assessing individuals with psychiatric disorders (e.g., Exner, 1993; Rapaport, Gill, & Shafer, 1946; Weiner, 1966), current attempts by some Rorschach proponents to dismiss this negative research evidence by arguing that the Rorschach was never intended to be a diagnostic test (e.g., Weiner, 1999, 2000) appear to be little more than post hoc rationalizations for consistently disappointing empirical results. Indeed, as recently as 1997, Weiner suggested that the Rorschach could play an important role in differentiating among diagnoses. If the Rorschach was never intended to be a test that is useful for diagnostic purposes, then it is difficult to grasp why dozens of Rorschach researchers have published articles in an attempt to validate the Rorschach for exactly these purposes, or why a recent assessment working group sponsored by the American Psychological Association concluded that Rorschach data can be used to differentiate among numerous Axis I and Axis II disorders (Kubiszyn et al., 2000).

In recent years, following positive reviews of the predictive validity of the Thematic Apperception Test (TAT; McClelland, Koestner, & Weinberger, 1989; Spangler, 1992), a view has been expressed among Rorschach proponents that the Rorschach also has the ability to predict long-term behavioral outcomes and choices (e.g., Masling, 1997; Stricker & Gold, 1999). Usually the basis for this claim is that both tests are "projective" tests, even though, as we have discussed, the CS does not treat the Rorschach as a "projective" test. Using analogic reasoning to draw a conclusion about the Rorschach's predictive validity is flawed for at least two other reasons. First, claims about similarities in the predictive validities of the two tests must be substantiated with evidence, not rhetoric. Although a specific comparison between Rorschach and TAT data was not made in Hiller et al.'s (1999) meta-analysis, their finding that the relation between the Rorschach and projective tests was essentially zero (weighted-mean $r = .03$) should give pause to those claiming the Rorschach possesses the alleged predictive virtues of the TAT. Second, McClelland et al. (1989) and Spangler's (1992) conclusion that apperceptive test data can predict such long-term outcomes as career success has erroneously been taken as an endorsement of the TAT itself. Across apperceptive test studies in the literature, there is little consistency in the cards selected for study by the researchers (Keiser & Prather, 1990), thus raising questions about the exact nature of the test stimuli that may have predictive validity. Furthermore, many of the studies reviewed by McClelland et al. and Spangler used apperceptive card stimuli that were not among the set of cards commonly known as the TAT that was developed by Murray (1943). Thus, even strong associations between Rorschach and true TAT variables would not necessarily indicate that the Rorschach has the predictive validity that some (non-TAT) apperceptive test variables may possess.

To summarize thus far, there seems little evidence that construct validation studies of the Rorschach can simply and uniformly rely on validation criteria such as psychiatric diagnosis or projective test data. Although the evidence for the association between Rorschach scales and self-report data is mixed, many Rorschach

advocates have argued that self-report data are not useful in validating the Rorschach. What then is left as possible validation criteria? The data from the Hiller et al. (1999) meta-analysis and from some of the studies reviewed by Viglione (1999) indicate that Rorschach researchers should strongly consider use of behavioral data as criteria for validating Rorschach scales. Although the reported value of a weighted-mean $r = .35$ is likely to be inflated by the inclusion of studies in which simple group membership was taken as examples of behavioral data (half of these studies simply compared data from norms or a normal control group with data from a group with some form of medical condition or psychological impairment), this result would suggest that greater attention should be paid to using observable behavioral indicators such as treatment attendance, hospitalization, arrest, and responses to behavioral tasks (such as delay of gratification) as validity criteria. Concrete, overt behaviors have frequently been used to validate the ROD (Bornstein, 1996) and should be used much more frequently in attempts to validate other Rorschach scales. Consistent, replicated evidence, indicating that a specific Rorschach scale can predict actual behavior as well as or better than other psychological assessment measures, would be sure to cause even the most ardent critics of the Rorschach to reconsider their position.

In conclusion, there appears to be a need for greater consensus among Rorschach researchers about what constitutes appropriate validation criteria for Rorschach scales. Given that the majority of extant Rorschach research has used either self-report or psychiatric diagnosis as data for convergent validity studies, it is difficult to conceive why such data should now be deemed as no longer relevant for validating any Rorschach scales. At the very least, some statements are necessary about the conditions under which such data constitute legitimate criteria for validation efforts. Given the limited evidence for convergence between Rorschach scales and projective measures found in the Hiller et al. (1999) meta-analysis, Rorschach proponents should not automatically assume that substantial associations exist among these tests and should not assume that claims about the putative value of other such tests can simply be generalized to Rorschach data. Finally, Rorschach researchers would be well advised to consider use of more overt behavioral indicators when testing the validity of Rorschach scales.

Toward the Next Generation of Rorschach Meta-Analyses

Notwithstanding the substantial methodological and statistical problems with the Hiller et al. (1999) meta-analysis outlined by Garb et al. (2001), the results appear to indicate that the general validity values obtained for Rorschach and MMPI variables are in the same range, with the weighted-mean effect sizes for the MMPI higher than that for the Rorschach ($r = .37$ and $.26$, respectively). These results are generally consistent with the range of results reported by Parker, Hanson, and Hunsley (1988) and the subsequent reanalysis of these data by Garb, Florio, and Grove (1998) to address the comparative validity of the Rorschach and MMPI.

In our opinion, nothing is to be gained by another global meta-analysis that addresses the question of whether the Rorschach can produce valid data. It is readily apparent to both Rorschach proponents and critics that some Rorschach variables can be valid in certain circumstances. As Weiner (2001) stated,

Rorschach scales will demonstrate greater validity for some purposes than for others. Thus, the next generation of Rorschach meta-analyses should address questions such as whether there are specific Rorschach scales that have consistent validity evidence and whether there are factors that act as moderators of the validity of the scales. To provide compelling evidence about the validity of the Rorschach to the scientific community, much more needs to be known about the range of populations and constructs for which Rorschach scales are meaningful and have demonstrated validity (cf. Aronow, 1999). We would strongly recommend that such meta-analyses include data from unpublished dissertations. In contrast to unpublished studies that may have never been submitted to peer review or that have been rejected for publication, unpublished dissertations are relatively easy to obtain and can provide data that can assist in determining whether there are biases in meta-analytic estimates based on published research.

Similarly, we believe there is little to be gained from additional global comparison of Rorschach data with data from other personality measures such as the MMPI. The goal of these Special Sections was to evaluate the clinical utility of the Rorschach, a goal not actually addressed by the Hiller et al. (2001) meta-analysis of scale validity. To begin to address the question of utility, future meta-analyses must be much more focused, using data in which Rorschach scales and other psychological measures or sociodemographic or historical data are directly compared with respect to their ability to predict specific, clinically relevant criterion variables. Knowing that the average validity of Rorschach variables is $r = .27$ reveals nothing about the comparative strengths and weaknesses of individual Rorschach scales, nor does it provide sufficient evidence to demonstrate that clinical decisions should be made on the basis of Rorschach data. What are needed are analyses that have the potential to inform us about which Rorschach variables have validities equal to or exceeding that found with other measures and, conversely, which Rorschach scales may be less valuable than other measures in predicting specific criteria. To our knowledge, no one has yet attempted to conduct such a meta-analysis. To illustrate our suggestion, given that critics and advocates agree that the Rorschach can detect disordered thinking (see our statements in the following section), a meta-analysis of the Thought Disorder Index and other similar scales might be in order. But, for such a meta-analysis to prove useful, it would be important to include studies that directly compare such Rorschach scales to self-report measures and interviews designed to tap disordered thinking.

Do Rorschach Data Meaningfully Inform Clinical Judgments?

Although, at one level, we can understand and appreciate Stricker and Gold's (1999) impassioned call for an idiographically and nomothetically informed use of Rorschach data, Rorschach proponents must still recognize that there is extremely limited evidence that the Rorschach can meaningfully add to the assessment and prediction of clinical or forensic variables. Viglione's (1999) review revealed two key features of the Rorschach research literature on the role of the Rorschach in clinical assessment. First, researchers have been incredibly creative in using the Rorschach with a staggering array of clinical problems, diagnoses, and clin-

ical services. Second, the smorgasbord of research findings he presented is remarkable for, with rare exceptions, its nonprogrammatic and noncumulative nature. Promising results are rarely replicated in the Rorschach literature, and intriguing interpretations of unexpected research findings are presented but often never subsequently pursued. Although similar conclusions could be drawn about many research topics, it is disconcerting that this is the case with a clinical test that has been used so frequently for the past several decades.

In their critique of Viglione's (1999) review, Garb et al. (2001) demonstrated that Viglione substantially overstated the case for Rorschach data contributing to the assessment of narcissism, depression, psychosis proneness, and biological predispositions to schizophrenia. Garb et al. also suggested that Viglione ignored negative findings in regard to the utility of the Rorschach in studies of clinical judgment. However, Viglione did cite five studies that he claimed demonstrated the value of Rorschach data in making clinical judgments (as noted by Garb et al., none of the studies used the CS to score the Rorschach data). As previous evaluations of this literature on Rorschach-informed clinical judgment have been so consistently negative (e.g., Garb 1984, 1998), we believe it is important to examine these five studies more closely to determine whether they do indeed provide evidence for the Rorschach's clinical utility.

In the study by Dudek and Marchand (1983), the Rorschach protocols of painters were compared with independent evaluations made about the style of their artistic work. Although such a study may be of interest for research on artistic creativity, we are at a loss to understand why Viglione (1999) chose to include this as an example of the clinical utility of the Rorschach. At the very least he should have indicated how such a study is relevant to the issue of clinical (not artistic) judgment.

Dana and Back (1983) conducted a study to examine the association between the Rorschach data of 59 children in preschool classes and evaluations of the children by a teacher and by the school director. Rorschach protocols were scored, according to the Klopfer scoring system, and interpretations were made for each protocol. Teachers and the school director were given an interpretive statement on each child that contained both interpretations from that child's Rorschach (50% of the interpretative statements) and interpretations from other children's Rorschachs. These raters were asked to indicate half of the statements with which they agreed and half with which they did not agree. In general, the content of these interpretive statements focused on the level of the child's intelligence, social skills, and emotional expressiveness. Phi coefficients were computed between the actual interpretative statements made of the child and the raters' endorsement of the statements. The average phi value reported by the authors was a statistically significant, but relatively small value of .28. However, this mean value is inflated because it included (a) the association among raters with (b) the association of the actual Rorschach interpretations and the teacher and director ratings. In fact, the level of association among raters (range from .40 to .60) was consistently higher than that between the actual Rorschach interpretation and the raters (range from .15 to .26). Moreover, even the level of association among raters is inflated because most of the interpretive statements that were true for one child were also true for many of the children (e.g., the statement "Gets along very

smoothly with peers" was made by the Rorschach evaluators for 14 of the 59 children), and no statistical adjustment was made for the fact that most of the interpretive statements would have applied to more than one child. Overall, therefore, the value of the Rorschach as demonstrated by this study is substantially less than what was described by Viglione (1999).

Bilett, Jones, and Whitaker (1982) examined the ability of expert clinicians to identify adolescents with schizophrenia on the basis of their unscored Rorschach responses records. Three aspects of this study are noteworthy. First, the Rorschach data were interpreted without any reference to norms or established scoring criteria—an activity inconsistent with Stricker and Gold's (1999) and Weiner's (2001) position on the appropriate use of the Rorschach. Second, in light of Wood et al.'s (2000) review of the literature on the assessment of schizophrenia with the Rorschach, it is hardly surprising that deviant verbalizations could be identified and used to discriminate between adolescents with and without schizophrenia even without the benefit of scoring. Third, no evidence was gathered in the original study to demonstrate whether the Rorschach data were unique in their ability to provide information on schizophrenia—one might speculate that a simple transcript of an interview with the adolescents would be as useful in identifying adolescents suffering from a severe mental illness.

Holzman et al. (1974) reported data from a study in which clinicians who used Rorschach data made diagnostic determinations that were moderately correlated with chart diagnoses, including diagnoses of psychotic conditions. Such a finding appears to be very relevant to the routine clinical use of the Rorschach. Because chart diagnoses were based on second edition *Diagnostic and Statistical Manual of Mental Disorders* (2nd ed.; *DSM-II*; American Psychiatric Association, 1968) criteria and because these criteria were notoriously unreliable, it would be important to attempt to replicate these results with current diagnostic criteria. On the other hand, as Weiner (1999, 2000) has argued that studies using diagnostic status as a validation criterion are inappropriate in the evaluation of the Rorschach, a failure to replicate the Holzman et al. (1974) results might be easily dismissed by Rorschach proponents.

The results of the study by Cerny (1984) verge on the astounding. In this study, seven psychologists were given information about the responses to the 10th Rorschach inkblot by 40 psychiatric patients with unspecified diagnoses. After reviewing the responses, they were asked to rate the patient's ability to use psychological understanding and to predict the course of treatment for the patient, both on 5-point scales (unfortunately the anchors for these scales were not reported in the original article). Using intraclass correlations, Cerny reported extremely high reliability among raters ($r = .85$ and $.92$, for the two items). Treatment staff who had worked with the patients were asked to make consensus ratings for patients on these two items. The correlations between the staff ratings and the psychologist ratings were $.68$ and $.67$, respectively. As any assessment researcher knows, these are truly remarkable findings which, if replicated, warrant serious consideration of the value of responses to the last Rorschach card, Card X. To our knowledge, however, these results, from what the author described as a pilot study, have never been replicated. We are left, therefore, to speculate on the generalizability of these results. One possibility is that these seven psychologists were truly impressive

diagnosticians who were able to use extremely limited data to produce ratings comparable to those made by treatment staff who had worked with the patients over an extended period of many years. A more likely possibility, to our minds, has to do with the nature of the psychiatric conditions experienced by the patients. As all patients were inpatients, it is reasonable to assume that they all had significant psychological impairment at the time of admission. It is also reasonable to assume that there were some patients with psychotic symptoms and some without psychotic symptoms. As stated previously, it may be reasonably simple to identify the Rorschach responses of those with psychotic spectrum illnesses given the increased likelihood of deviant verbalizations. Therefore, if our speculation is correct, the high concordance among raters and treatment staff may be due to (a) being able to identify those patients with psychotic symptoms on the basis of Card X responses and (b) giving these patients low ratings for their perceived level of psychological mindedness and treatment course. Of course, all of this is speculation, but clearly this study warrants an attempt at replication.

Taken together, therefore, we suggest that these five studies offer rather limited evidence to support Viglione's (1999) position that the use of Rorschach data enhances clinicians' judgments. As we stated, the results of some of the studies are potentially relevant to the clinical use of the Rorschach and therefore attempts to replicate the findings should be made. However, if, as we suspect, the interesting results from some of these studies boil down to the identification of bizarre and deviant verbalizations and the subsequent interpretation of the statements as evidence for thought disorder or a schizophrenic illness, researchers should examine whether such clinical interpretations and predictions could be made on the basis of less costly and time-consuming assessment data. Additionally, there is also a need to demonstrate that other Rorschach scales or qualitative interpretations of the Rorschach can routinely improve clinical judgments.

The Clinical Utility of the Rorschach (Revisited)

Suppose that a mental health hospital administrator is charged with deciding whether, or under what circumstances, the Rorschach should be used for psychological assessments in the hospital. Suppose further that the administrator's singular goal is to maximize benefit to patients, given typical budget constraints. The administrator knows that resources devoted to Rorschach assessment cannot be devoted to other activities, some of which (e.g., psychotherapy) provide substantial benefits to patients, on average. To address this issue, the administrator may well seek information on reliability, validity, and incremental validity, because very low values of these (especially validity and incremental validity) are sufficient to reject the Rorschach (or any instrument) without further consideration. However, even substantial validity and incremental validity cannot establish that the Rorschach is worth using in the hospital. It would be interesting, for example, if a Rorschach sign predicted artistic creativity well, but it is difficult to imagine that this would be sufficiently beneficial to justify the time and expense to administer the Rorschach. Thus, a conscientious administrator would not endorse the Rorschach on the basis of a report that reviewed only reliability and validity coefficients. Nor would it be sufficient to compare the validities of the Ror-

schach and some other instrument (e.g., the MMPI), unless the latter's clinical utility was well established. The administrator needs to know more. Are psychologists who use the Rorschach (in certain situations) better equipped to help their patients than psychologists who do not use it? Do patients who have been given the Rorschach have better outcomes than other patients? These questions—and not reliability, validity, or incremental validity—comprise the core considerations of clinical utility.

If the Rorschach is to be justified through its positive impact on clinical decision making, three questions need to be addressed:

1. What decision is a given Rorschach scale hypothesized to inform? For a clinical setting, this will typically involve offering or withholding an intervention.
2. What are the rates of true positives, true negatives, false positives, and false negatives attributable to the Rorschach scale (i.e., how do those quantities change because of the inclusion of the Rorschach data)?
3. What are the costs of using the Rorschach and the costs of false positives and false negatives, and what are the benefits of true positives and true negatives?

The administrator may not require precise estimates of the costs and benefits of the different outcomes, but the utility of the Rorschach (or any instrument) cannot be evaluated unless the above Questions 1–3 can be estimated. If the Rorschach is used in conjunction with other tests (as would be typical), the three criteria we have delineated still apply. Specifically, in this case “true positives, true negatives, false positives, and false negatives attributable to the Rorschach scale” mean the changes in the rate of true positives, true negatives, false positives, and false negatives with the inclusion of the Rorschach in the battery. The costs referred to in the third criterion are those added by the Rorschach's inclusion, above those incurred by the battery without the Rorschach.

To start to answer some of these questions, our hypothetical administrator would eagerly read the first Special Section on the Rorschach published in *Psychological Assessment*. Unfortunately, the only conclusion that the administrator could plausibly draw is that to date, after more than 50 years during which the Rorschach was heavily used and avidly researched, there has been no attempt to demonstrate its clinical utility (as opposed to its reliability, validity, and incremental validity). Hence at present, there is no evidence for the utility of any Rorschach scale.

In the first round of articles in the Special Series, we concluded that there is no evidence that the Rorschach has clinical utility (Hunsley & Bailey, 1999). In the same round, Viglione (1999) also ostensibly addressed the utility of the Rorschach. In his concluding section, he answered the question “Is the Rorschach useful?” with an unqualified “yes” (p. 260). Both articles included similar definitions of utility; therefore, their opposite conclusions are not attributable to differences in definition. What evidence did Viglione marshal in support of the Rorschach's clinical utility? The reader will search in vain for the requisite systematic consideration of decision accuracy, costs, and benefits that Viglione's own definition of utility requires (see Viglione, 1999, p. 251). Rather, Viglione's optimistic conclusions were possible only because he virtually ignored his own definition and instead reviewed evidence about validity. We have already noted our skepticism about the quality of these validity findings, but here we emphasize that even

if we accept every finding that Viglione reviewed, they are far from sufficient to establish utility.

The closest that Viglione came to consistency with his stated goal of evaluating the utility of the Rorschach was in his discussion of suicide risk. He concluded that the Suicide Constellation (S-Con) is associated with suicidal behavior, although the evidence he reported was mixed, and none of the studies assessed incremental validity. He then argued that because the cost of a false-negative error (i.e., falsely predicting that someone will not commit suicide) is so great, a high rate of false-positive errors should be tolerated, and hence, the S-Con is a useful indicator.

This line of reasoning is insufficient to establish utility. First, as Viglione surely knows, suicide risk prediction is the bane of psychological assessment. Because the risk of suicide is typically low in real clinical populations, false positives invariably outnumber true positives by orders of magnitude. For example, in a large study of a high-risk population (inpatients with affective disorders) a statistical model was developed using valid indicators, including previous suicide attempts and suicidal ideation (Goldstein, Black, Nasrallah, & Winokur, 1991). Of 1,906 patients, 46 eventually committed suicide. None of the completed suicides was predicted by the model. We doubt that the Rorschach would improve prediction over the easily assessed and obviously relevant variables included in the model, but that is an empirical question.

Another glaring omission in Viglione's argument is support for his assertion that the cumulative costs of false negatives outweigh the cumulative costs of false positives. Although the assertion has superficial appeal, as suicide is certainly tragic, it is true only under certain circumstances. Specifically, the cumulative benefit in terms of suicides prevented by intervening on the basis of the S-Con must exceed the cumulative costs of both giving the Rorschach and intervening unnecessarily on the basis of the S-Con. One simply cannot estimate these costs and benefits unless one specifies an intervention and knows its effectiveness. One common intervention to prevent suicide is involuntary civil commitment; how many nondangerous people is it permissible to hospitalize against their wishes in order to prevent (or postpone) one suicide? A less drastic intervention is therapy, but there are still considerable costs to providing therapy for numerous false positives and the efficacy of such treatment is unclear (Rudd, Joiner, Jobes, & King, 1999). In finite health care budgets, expenditures on therapeutic attempts to prevent suicide reduce expenditures available for other services. An analysis of the Rorschach's utility in predicting suicide must confront these issues directly rather than sidestepping them, as Viglione did.

Because of the low base-rate problem, we think that suicide prediction is unpromising territory for any assessment instrument, including the Rorschach. Are there more promising candidates than the S-Con, with respect to clinical utility? Viglione suggested that the Rorschach is useful in individualizing case conceptualizations and interventions and in preventing undesirable outcomes. Let us briefly consider these possibilities. The claim that the Rorschach aids in individualizing case conceptualizations is especially pertinent, because frequently no specific referral question motivates a psychological assessment, but rather, the goal is to “understand a case.” We are less impressed than Viglione with the likelihood that the many Rorschach indices meaningfully illuminate a patient's problems. Regardless, case conceptualization is not

a useful end in itself. Suppose, for example, that a Rorschach sign reliably increased the understanding or prediction of some patient characteristic, that this increased understanding or prediction was gratifying to clinicians, but that on average, clinicians' increased understanding or prediction provided no benefit to patients. It seems clear to us that the use of the Rorschach for this purpose could not be justified. At best, patients will waste their time completing the Rorschach for no benefit to them, and they may also have to pay for test administration and scoring. (For a relevant thought experiment, suppose that therapists found the information provided by the Rorschach extremely gratifying, but that research demonstrated that the Rorschach was slightly harmful to patients. Does anyone really believe that the Rorschach's use in this case could be justified?)

This alleged attribute of the Rorschach is useful only if improvement in case conceptualization attributable to the Rorschach benefits the patient. If use of the Rorschach is to be justified by its impact on "understanding the patient," then our hypothetical administrator would want good evidence that such understanding actually helps patients. At this point in time, there is no such evidence in the literature, although it would not be difficult to conduct a definitive study. One might randomly assign patients to be evaluated with or without the Rorschach, provide therapists with the resulting assessment data, and subsequently compare treatment outcome in the two groups. Until such a study is conducted, however, the claim that the Rorschach provides clinically useful case conceptualization remains undemonstrated.

Similarly, the goal of identifying which of several interventions will be most effective for a given patient is potentially useful. But finding a nonzero correlation between a Rorschach scale and treatment outcome is far from sufficient to demonstrate utility: Information must also be available on the cumulative costs and benefits. For example, Viglione cited as especially noteworthy a small ($N = 46$), unreplicated study in which a Rorschach scale, the Ego Impairment Index (EII), predicted outcome during a course of tricyclic antidepressant treatment over and above two depression symptom scales (Perry & Viglione, 1991). He did not even begin to suggest why this finding is clinically useful, but let us suppose that he believes that it should be used in the decision whether to prescribe such medication. Tricyclic antidepressants are sufficiently bothersome (side effects include weight gain, anxiety, and rapid heartbeat) and dangerous (overdose can cause death) that prescribing them on the basis of a small incremental association with the EII seems dubious. Rather, a more rational approach would be to try tricyclic medication as a last resort, after psychotherapy and less potentially harmful medications have failed. Would we want to deny patients at that stage access to this medication because their Rorschach data suggests that they are slightly less likely to respond favorably compared with other patients? Without specifying how a Rorschach scale is to be applied in practice and estimating costs and benefits, there is no basis for a claim of clinical utility.

Let us move on to Viglione's claim that the Rorschach aids in prevention efforts. We have already argued that Rorschach data are unlikely to be useful in preventing suicide. Another possibility vaguely suggested by his review is the prevention of psychosis by identifying individuals at risk. Again, the clinical utility of the Rorschach cannot be estimated without specifying what, precisely,

one wants to do with individuals so identified. Although prevention of psychosis is a laudable goal, there is no consensus that it is possible. Furthermore, the most likely intervention, antipsychotic medication, has many potentially severe side effects, and hence has substantial physical costs (especially for false-positive cases). Only by estimating the costs and benefits of intervening on the basis of Rorschach information can we know whether use of Rorschach data is clinically helpful or clinically harmful.

Viglione's stated intention was to evaluate clinical utility, yet his unambiguously positive conclusions were clearly at odds with ours. We suggest that the difference in our conclusions is due to differences in recognizing the nature of data that are necessary for determinations of clinical utility. Despite his intentions, Viglione did not evaluate clinical utility, but rather he compiled Rorschach validity findings that he believes are clinically important. However, as we have tried to make very clear, this leaves the clinical utility of the Rorschach an open question. On the basis of our review of the literature, we would recommend that our hypothetical administrator forego Rorschach assessment for now and spend the hospital's limited mental health budget on clinical activities of more demonstrated utility, such as empirically supported psychotherapies.

In closing this section, we note that some other popular instruments have also failed to generate necessary data in support of clinical utility. The most relevant example is the MMPI. We know of no set of findings that could be combined to show that the MMPI is a cost-effective tool for an important clinical goal. (We also note, however, that if the MMPI and Rorschach have equal validity—even if validity is near zero—then the MMPI will have greater cost-effectiveness because of its less costly administration and scoring.) Research necessary to demonstrate a test's clinical utility is not conceptually difficult, although it may be somewhat logistically onerous. We suspect that the primary reason why such research has not accumulated for the Rorschach, MMPI, or other tests is that these instruments have been so popular without it. We believe that this state of affairs should change and that cost-consciousness imposed by managed care will provide the impetus.

Response to Weiner (2001)

It is hardly surprising that Weiner (2001), one of the most knowledgeable and vocal proponents of the Rorschach, did not concur with our perspective on the empirical evidence for the Rorschach. Given his senior status in the Rorschach community of researchers and practitioners, it is, however, disappointing that he consistently chose to either ignore our arguments or to respond to our critiques with rhetoric rather than with empirical evidence. For example, our critique of previous Rorschach meta-analyses and of published statements that the Rorschach's validity has been established by these meta-analyses went essentially unchallenged by Weiner, as did our evaluation of the Meyer and Handler (1997) meta-analysis. Instead, Weiner primarily relied on broad dismissals of our position, based primarily on the number of empirical articles included in our reference list and in Viglione's (1999) reference list. This does little to further inform scientific debates about the merits of the Rorschach. Similarly, the presentation of a partial quotation from Parker, Hunsley, and Hanson (1999) does little to promote scholarship. The complete quotation referred to by

Weiner is as follows: "When presented with a well-designed study, psychologists should be prepared to accept that there may be domains in which the Rorschach has greater utility than the MMPI" (p. 292). The clinical utility of the Rorschach must be judged not on the volume or breadth of conducted research, but on the evidence of data stemming from well-designed, well-analyzed, and replicated research studies. The comment by Garb et al. (2001) and the detailed review by Wood et al. (2000) of the diagnostic validity of the Rorschach provide excellent models of the clinically and scientifically informed scrutiny that is necessary in reviewing the literature to determine the merits and limitations of the Rorschach.

Weiner proposed that a psychometrically sound instrument must be standardized in terms of its administration and scoring, must possess valid norms, and must demonstrate various forms of reliability and validity. We concur fully; however, as we will show, the evidence supporting the Rorschach and the CS do not meet the standards that Weiner himself described (cf. Hunsley, Lee, & Wood, in press).

With regard to standardization, the CS has certainly established a clear set of guidelines for the administration, inquiry, and scoring of the Rorschach responses, and the CS appears to be the dominant scoring system used by clinicians who use a standardized system (Piotrowski & Keller, 1989). Importantly, Weiner (2001) explicitly stated that the failure to follow standardized administration and scoring procedures is inappropriate in clinical settings. We applaud this stance and encourage the community of Rorschach researchers to take steps to uniformly denounce the tendency among many clinical users of the Rorschach to use the test inappropriately.

Standardized administration of a test is a starting point, but more is required for a test to be psychometrically sound. Contrary to Weiner's assertion that the question of CS reliability has been settled, Garb et al. (2001) reported that the test-retest reliability for most CS scales has yet to be established and that the limited data on interrater reliability provided mixed support, at best, for scale reliability. Addressing the question of scoring reliability, Weiner dismissed our concerns about the field reliability of the test (i.e., reliability as demonstrated in the routine clinical or forensic use of the test) as irrelevant to the question of the Rorschach's "intrinsic" reliability. As he suggested, one must distinguish between faulty administration of a test and a test that has poor reliability regardless of the quality of the administration. This is an important distinction for numerous reasons, including the requirements of current standards for the admissibility of psychological assessment data into court proceedings (i.e., the data must be obtained and interpreted appropriately; Marlowe, 1995). However, in practice, the boundary separating these two categories is likely to be blurred: If an instrument, because of the complexity or incompleteness of its scoring rules, cannot be appropriately scored when routinely administered, any research-based validity evidence is rendered meaningless when the instrument is used clinically. Contrary to the position attributed to us by Weiner, we do question whether the Rorschach can be coded reliably in routine clinical practice. There is clear evidence for scoring reliability in the context of carefully and expertly supervised research studies; our concern is that there is no extant evidence that demonstrates the level of scoring accuracy achieved when the CS is used in clinical practice. We fully agree with Weiner that when the scoring of a CS

protocol is clearly erroneous or incomplete, the fault rests with the clinician who conducted the administration and scoring. However, we do not have evidence that addresses the accuracy or reliability of scoring among those who are well-trained in the CS system (i.e., psychologists who received thorough initial training, are involved in regular advanced training workshops, and attempt to follow the scoring and interpretation directions provided in the various guides to scoring the CS) but who do not have ongoing access to the type of supervision or consultation available in research studies. Although one could frame this as a problem of scoring drift from a gold standard set by research studies, it is possible that the requirements for good CS scoring are such that it is difficult to achieve good scoring without frequent input from experts.

Weiner suggested that no one raises such questions about the field reliability of the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955) or similar tests; therefore, it is inappropriate to question the ability of clinicians to score the Rorschach. Actually, there is a sizeable literature on the accuracy with which the various Wechsler intelligence tests are scored in clinical settings. For example, Slate, Jones, Murray, and Coulter (1993) reported that WAIS-R protocols scored by practitioners contained numerous errors and that, cumulatively, the errors frequently resulted in up to a 5-point discrepancy with the true full scale IQ. Similar substantial scoring inaccuracies have been found when trained, experienced practitioners administer and score the revised Wechsler Intelligence Scale for Children (Slate, Jones, Coulter, & Covert, 1992) and the Wechsler Preschool and Primary Scales of Intelligence—Revised (Whitten, Slate, Jones, & Shine, 1994). In their study of scoring errors on the Wechsler Adult Intelligence Scale—Revised (WAIS-R) made by psychologists and graduate students, Ryan, Prifitera, and Powers (1983) found that summary scores varied as much as 28 IQ points and that the scorer's experience level had no effect on the accuracy of the scoring. Although these data suggest that there is something amiss in how practitioners score complex tests, we are loathe to simply blame the practitioners for this situation. We would suggest that it should not be assumed that a structured test with clear but complex scoring guidelines can be consistently and reliably scored in routine practice. As the CS scoring is even more complex than the scoring for these intelligence tests, it seems highly likely that there are substantial opportunities for scoring errors in the CS. Far from being a picayune issue, evidence that the Rorschach is accurately scored in real-world settings is sorely needed.

Weiner suggested that the CS norms are a significant strength of Exner's approach to the Rorschach. However, a recent study by Shaffer, Erdberg, and Haroian (1999) suggests there may be substantial problems with the CS norms. These researchers evaluated 123 nonpatient adults on the Rorschach, MMPI-2, and WAIS-R in order to compare the data from a current sample of normal adults with the norms for these widely used clinical tests. As expected, the MMPI-2 and WAIS-R data were very similar to the normative standardization data for these tests. Quite unexpectedly, though, the researchers found substantial differences between the Rorschach data and the CS norms. The general pattern that emerged with these Rorschach data was that the responses of these normal adults appeared very pathological when judged by using the CS data.

Although one might be tempted to treat these findings as simply anomalous, the discrepancy found between the Rorschach scores of a sample of normal participants and the CS norms has been found consistently in the literature. Using the Shaffer et al. (1999) study as a guide, Wood, Nezworski, Garb, and Lilienfeld (2001) examined studies in which nonpatient adults were administered the Rorschach and CS scales of distorted thinking or emotional problems were included (i.e., variables that showed substantial deviations from the CS norms in the Shaffer et al. study). In total, they examined 32 studies (including journal articles, book chapters, and unpublished dissertations) in which 14 different CS variables were included (i.e., *EB style*, *reflection responses*, $X \pm \%$, $X - \%$, *Afr*, *FC*, *Populars*, *Sum Y*, *Sum T*, *WsumC*, *MOR*, *Wsum6*, *Lambda*, and *Pure H*). The results were the same across all variables: In each case, the data from the sample of adults differed statistically and clinically from the CS norms. The smallest discrepancy between the aggregated means and CS norms (expressed as an effect size d) was $d = -0.18$, the largest discrepancy was $d = -1.67$, and the median absolute discrepancy was $d = 0.73$. The median ratio of the standard deviation of scores in the reviewed studies to the standard deviations in the CS norms was 1.7, suggesting substantial discrepancies between standard deviations. Although none of the 32 samples were chosen so as to be representative of the American adult population, it is hard to imagine that biased sampling could account for the consistent pattern of discrepancies from the CS norms. The net result of these deviations in the means and standard deviations is that many of the nonpatient adults in the studies would have been judged as exhibiting psychopathology based on the CS norms. Wood et al. suggested that the tendency for the CS to overpathologize may be at the heart of conviction held by many (including Weiner [2001]) that the Rorschach can detect pathology when other instruments fail to do so.

It also appears that these substantial problems with CS norms may not be restricted to the adult norms. Using data from 100 children screened for psychopathology and behavior problems (based on historical information and assessments of current functioning), Hamel, Shaffer, and Erdberg (2000) recently concluded that there were substantial and consistent discrepancies between the CS children's norms and the data from their sample of children. Moreover, considerable numbers of children scored in the clinical range on numerous indices of psychopathology (e.g., slightly under 50% obtained elevated scores on the Depression index). Clearly there is something amiss with the CS norms for both adults and children.

To recap, from our review of the scientific literature, there is little information on the extent to which standardized scoring systems for the Rorschach are used and scored reliably by clinicians, and there is substantial doubt regarding the accuracy of the CS norms for adults and children. The other realm that Weiner described as critical for a psychometrically sound test was that of validity. Garb et al.'s (2001) validity-related comments in response to the first round of articles in this Special Series demonstrate how little empirical support there is for Rorschach scales typically presented as being scientifically meritorious. As we indicated earlier in this article and in our contribution to the previous Special Section, there is mixed evidence in support of the validity of the Rorschach when compared with self-report data and little evidence in support of its validity when diagnostic information and projec-

tive test data are used as validity criteria. Thus, although there is some general validity evidence that supports some Rorschach scales, there is as yet no clear indication that most Rorschach scales consistently demonstrate validity with regard to the clinical purposes for which they are used (cf. Aronow, 1999). Finally, there is absolutely no evidence to support the contention made by Stricker and Gold (1999) and Weiner that individualized interpretations of the Rorschach data are valid and are a key asset of the test. The results of nomothetical approaches to the question of incremental validity suggest that Rorschach data add little beyond other commonly available clinical data, and there are no replicated studies using idiographic methods that have demonstrated the superiority of clinical judgments made on the basis of Rorschach data over those made without benefit of such data. Claims that individualized interpretations are a strength of the Rorschach must be supported by empirical evidence, not just the assertions by respected and knowledgeable authorities.

The Scientific Status of the Rorschach: Is There a Consensus?

So what has been learned from these two Special Sections? We suggest that, although it is evident that there are numerous disagreements about the extent to which the Rorschach or the CS has a firm scientific foundation, there were several critical principles on which both advocates and critics agreed. All contributors to these two Special Sections agreed either implicitly or explicitly that the data obtained from a Rorschach must be treated as psychological test data and that, therefore, the Rorschach must meet the professional standards expected of psychological tests. Accordingly, the Rorschach must be administered, scored, and interpreted according to established rules, such as those set out in the CS. As Rorschach scales must be scientifically evaluated on their psychometric properties, there needs to be strong evidence of reliability and validity across the populations for which the scales are used. For psychologists grounded in a scientific approach to psychological assessment, the need for agreement on these basic principles may seem self-evident. However, as we have noted, many Rorschach proponents whose views were not represented in these Special Sections continue to claim that Rorschach data should not be treated as test data.

The views presented in these Special Sections also revealed many of the challenges facing Rorschach proponents. There were substantial differences of opinion regarding whether Rorschach scales have demonstrated adequate reliability and validity and whether extant norms are appropriate for continued clinical use. As both advocates (e.g., Aronow, 1999) and critics (e.g., Wood, Nezworski, & Stejskal, 1996) have suggested elsewhere, a complete review of the empirical evidence for each Rorschach scale is required to evaluate the true scientific status of the Rorschach. There were also clear differences expressed among contributors in regard to the evidence base supporting the incremental validity of Rorschach scales or of interpretations based on Rorschach data. As we stated in our article in the previous Special Section, frequent calls for research on incremental validity have not been successful in encouraging such research. Perhaps Dawes's (1999) illustration of the analytic strategy for assessing incremental validity will provide the needed impetus for this sorely needed line of research.

Future Directions for Rorschach Research: "To Boldly Go . . ."

For Rorschach research to advance and fill the gaps in the literature, there are a number of steps that must be taken, which we outline below. However, there must first be a reasonable level of consensus among Rorschach proponents on what the test measures, how it should be used, and how it should be validated. Without such agreement, there is little chance that even high-quality Rorschach research will have a meaningful impact on the test's use in service settings.

Without question, the Rorschach test is a demanding one to score, and revisions to the CS require that Rorschach users maintain up-to-date knowledge on scoring criteria and interpretations. There is, therefore, much work to be done to evaluate whether clinicians can routinely score the Rorschach with the CS with a high degree of accuracy. Likewise, there is an urgent need for representative norms for the CS to be developed. Once appropriate norms have been established, extensive research will be required to determine whether they are relevant to the widespread international use of the CS.

More programmatic research is needed on Rorschach scales in order to meet the scientific requirement of replicated evidence for the validity of each scale. As the Rorschach is routinely used in a variety of service settings, a wide range of populations and assessment purposes will need to be represented in such programmatic research. Rorschach researchers also need to heed the various cautions that have been published about the methodological and statistical problems that are endemic in the literature (e.g., Vigliano, 1997). Even in recently published articles there are frequent examples of such errors, such as using inappropriate statistics to analyze Rorschach variables with skewed distributions and conducting factor analyses with insufficient sample sizes (e.g., Baity & Hilsenroth, 1999). As an added caution, researchers should avoid use of diagnosis as a validity criterion simply because diagnostic information is readily available in clinical files. Those who wish to use diagnosis as a criterion would be well advised to follow the checklist presented by Wood et al. (2000) in designing their studies. However, on the basis of the Hiller et al. (1999) meta-analysis, it appears that researchers should focus their validation efforts more on behavioral criteria than psychiatric diagnoses per se.

As Garb et al. (2001) discussed, there is a problem with some published Rorschach studies where there is selective reporting of positive research findings. Other examples of this pattern have been presented by Wood et al. (2000); therefore, we will provide only two examples to illustrate the problem. Aron (1982) reported data that a Rorschach anxiety scale was able to discriminate between participants who reported experiencing either a high or low number of stressful life events; however, Aron's (1980) dissertation also included analyses that found several CS variables were unable to discriminate between the two groups. In a rare instance of public disclosure of selective reporting, a recent erratum by Smith, Gacono, and Kaufman (1998) indicated that the published article (Smith, Gacono, & Kaufman, 1997) and the dissertation on which the article was based reached almost diametrically opposed conclusions about the validity of the Rorschach (see Wood et al., 2000, for more details). It is likely that some

selective reporting occurs in all research literatures, and it is obvious that all researchers need to exercise good judgment to ensure that selective reporting does not occur. However, as a guard against selective reporting, we believe it is essential that future meta-analyses of the Rorschach include data from dissertations, not just published articles.

As a final research suggestion, Rorschach proponents (and proponents of other psychological tests, for that matter) have for too long overlooked the importance of evaluating whether Rorschach data have utility for making common clinical decisions such as treatment assignment and planning. Clear evidence that using Rorschach-informed assessments led to better treatment outcomes or lowered attrition rates would provide a compelling indication of the clinical value of the Rorschach. Until such data are available, the promise that so many advocates believe the test has remains little more than an article of faith.

Final Conclusions From the Perspective of Rorschach Critics

Weiner (2001) ended his article with a challenge for critics of the Rorschach to present abundant and compelling evidence that the Rorschach does not work for specific purposes. Leaving aside for the moment that in the scientific enterprise it is the responsibility of the proponents of a theory or test to provide evidence for their position (Lett, 1990), we believe that the evidence presented in this article and, more generally, by the Rorschach critics in these Special Sections (Dawes, 1999; Garb et al., 2001; Hunsley & Bailey, 1999) and in other recent publications (e.g., Wood & Lilienfeld, 1999; Wood et al., 2000) meets and exceeds Weiner's challenge. The Rorschach critics in this Special Series have shown that significant problems currently exist with both the Rorschach and the CS. After many decades of research and hundreds of research studies, there remain huge gaps in our knowledge about the reliability, validity, and norms for Rorschach scales.⁴ As a result, there is only limited replicated evidence to support the validity of just a few Rorschach scales, and even with these scales there is absolutely no evidence to suggest that the scales have any utility in determining what clinical services should be provided to clients or what real-world outcomes are likely to accrue by virtue of the use of these assessment data.

As evident in this Special Series, Rorschach advocates continue to claim that the test has a special role to play in clinical assessment, a role that cannot be filled by other measures. This assertion is one of incremental validity and, as we (Hunsley & Bailey, 1999) and Garb et al. (2001) have demonstrated, there is simply insufficient evidence to support this enormous faith in the incremental validity of the Rorschach.

⁴ Of course, there are questions about norms, reliability evidence, and validity evidence for many psychological tests. As the Rorschach is one of the tests most commonly used by psychologists (Watkins, Campbell, Nieberding, & Hallmark, 1995), it should be expected to have psychometric evidence comparable to other commonly used tests such as the Wechsler intelligence tests and the MMPI and MMPI-2. We would suggest that, on a comparative basis, there is far less compelling psychometric evidence for the Rorschach CS system than is the case for these other tests (cf. Hunsley & Di Giulio, 2001).

In conclusion, there is so little replicated evidence stemming from high-quality studies that we believe there is no scientific basis to support the continued widespread use of the Rorschach in clinical, legal, forensic, and occupational settings (cf. Grove & Barden, 1999). Although we are confident that the quality of Rorschach research will continue to improve, based on the history of Rorschach research and use, we doubt that the present lack of requisite scientific evidence will influence most clinicians who use the Rorschach. This being said, it does appear that there may be significant changes in the offing for Rorschach usage. Where the pressures of science have been ineffective in altering the assessment practices of many clinicians, the pressures of managed care seem to be exerting a more powerful effect. Several commentators have suggested that the assessment practices of clinical psychologists must undergo changes in order to meet the demands of managed care (e.g., Groth-Marnat, 1999, 2000; Piotrowski, 1999), and recent surveys have found that, due to managed care requirements, (a) many psychologists who previously used the Rorschach and other time-consuming tests now do so less frequently (Piotrowski, Belter, & Keller, 1998) and (b) many APA-accredited internships have decreased their emphasis on the teaching and use of tests such as the Rorschach and TAT (Piotrowski & Belter, 1999). Just as forces external to organized psychology have been responsible for influencing clinical psychology's current emphasis on the use of empirical evidence to guide treatment selection (Hunsley & Johnston, 2000), so too may it be that the fate of the Rorschach is ultimately determined not by science but by pressures for accountability and cost-effectiveness in health care services.

References

- American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, DC: Author.
- Aron, L. (1980). The Rorschach: Stressful life events and locus of control. (Doctoral dissertation, St. John's University). *Dissertation Abstracts International*, 41-04B.
- Aron, L. (1982). Stressful life events and Rorschach content. *Journal of Personality Assessment*, 46, 582-585.
- Aronow, E. (1999). The Rorschach: An integrative approach. *Contemporary Psychology*, 40, 546-547.
- Aronow, E., Reznikoff, M., & Moreland, K. L. (1995). The Rorschach: Projective technique or psychometric test? *Journal of Personality Assessment*, 64, 213-228.
- Bagby, R. M., Buis, T., & Nicholson, R. A. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: Replication and extension. *Psychological Assessment*, 7, 84-92.
- Baity, M. R., & Hilsenroth, M. J. (1999). Rorschach aggression variables: A study of reliability and validity. *Journal of Personality Assessment*, 72, 93-110.
- Ben-Porath, Y. S., Hostetler, K., Butcher, J. N., & Graham, J. R. (1989). New subscales for the MMPI-2 Social Introversion (Si) scale. *Psychological Assessment*, 1, 169-174.
- Bilett, J. L., Jones, N. F., & Whitaker, L. C. (1982). Exploring schizophrenic thinking in older adolescents with the WAIS, Rorschach, and WIST. *Journal of Clinical Psychology*, 38, 232-243.
- Boccaccini, M. T., & Brodsky, S. L. (1999). Diagnostic test use by forensic psychologists in emotional injury cases. *Professional Psychology: Research and Practice*, 30, 253-259.
- Bornstein, R. F. (1996). Construct validity of the Rorschach Oral Dependency Scale: 1967-1995. *Psychological Assessment*, 8, 200-205.
- Cerny, M. S. (1984). One last response to the Rorschach test: A second chance to reveal oneself. *Journal of Personality Assessment*, 48, 338-344.
- Cooper, S. H., Perry, J. C., & O'Connell, M. (1991). The Rorschach Defense Scales: II. Longitudinal perspectives. *Journal of Personality Assessment*, 56, 191-201.
- Costello, R. M. (1998). Psychometric definition of Rorschach determinant component structure. *European Journal of Psychological Assessment*, 14, 116-123.
- Costello, R. M. (1999). Two foundations of Rorschach assessment revisited. *European Journal of Psychological Assessment*, 15, 64-77.
- Dana, R. H., & Back, P. R. (1983). The concurrent validity of child Rorschach interpretations. *Journal of Personality Assessment*, 47, 3-6.
- Dawes, R. M. (1999). Two methods for studying the incremental validity of a Rorschach variable. *Psychological Assessment*, 11, 297-302.
- Dudek, S. Z., & Marchand, P. (1983). Artistic style and personality in creative painters. *Journal of Personality Assessment*, 47, 139-142.
- Exner, J. E. (1989). Searching for projection in the Rorschach. *Journal of Personality Assessment*, 53, 520-536.
- Exner, J. E. (1993). *The Rorschach: A comprehensive system. Vol. 1. Basic foundations* (3rd ed.). New York: Wiley.
- Exner, J. E. (1997). The future of Rorschach in personality assessment. *Journal of Personality Assessment*, 68, 37-46.
- Frank, L. K. (1948). *Projective methods*. Springfield, IL: Charles C. Thomas.
- Friedman, A. F., Gleser, G. C., Smeltzer, D. J., Wakefield, J. A., & Schwartz, M. S. (1983). MMPI overlap item scales for differentiating psychotics, neurotics, and nonpsychiatric groups. *Journal of Consulting and Clinical Psychology*, 51, 629-631.
- Ganellen, R. J. (1996). Exploring MMPI-Rorschach relationships. *Journal of Personality Assessment*, 67, 529-542.
- 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., Wood, J. M., Nezworski, M. T., Grove, W. M., & Stejskal, W. J. (2001). Toward a resolution of the Rorschach controversy. *Psychological Assessment*, 13, 433-448.
- Goldstein, R. B., Black, D. W., Nasrallah, A., & Winokur, G. (1991). The prediction of suicide: Sensitivity, specificity, and predictive value of a multivariate model applied to suicide among 1906 patients with affective disorders. *Archives of General Psychiatry*, 48, 418-422.
- Groth-Marnat, G. (1999). Financial efficacy of clinical assessment: Rational guidelines and issues for future research. *Journal of Clinical Psychology*, 55, 813-824.
- Groth-Marnat, G. (2000). Visions of clinical assessment: Then, now, and a brief history of the future. *Journal of Clinical Psychology*, 56, 349-365.
- Grove, W. M., & Barden, R. C. (1999). Protecting the integrity of the legal system: The admissibility of testimony from mental health experts under Daubert/Kumho analyses. *Psychology, Public Policy, and Law*, 5, 224-242.
- Hamel, M., Shaffer, T. W., & Erdberg, P. (2000). A study of nonpatient preadolescent Rorschach protocols. *Journal of Personality Assessment*, 75, 280-294.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T. R., & Brunell-Neulieb, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment*, 11, 278-296.

- Hilsenroth, M. J., Handler, L., Toman, K. M., & Padawer, J. R. (1995). Rorschach and MMPI-2 indices of early psychotherapy termination. *Journal of Consulting and Clinical Psychology, 63*, 956-965.
- Hirschfeld, R. M. A., Klerman, G. L., & Gough, H. G. (1977). A measure of interpersonal dependency. *Journal of Personality Assessment, 41*, 610-618.
- Holzman, P. S., Proctor, L. R., Levy, D. L., Yasillo, N. J., Meltzer, H. Y., & Hurt, S. W. (1974). Eye-tracking dysfunctions in schizophrenic patients and their relatives. *Archives of General Psychiatry, 31*, 143-151.
- Hunsley, J., & Bailey, J. M. (1999). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment, 11*, 266-277.
- Hunsley, J., & Di Giulio, G. (2001). Norms, norming, and clinical assessment. *Clinical Psychology: Science and Practice, 8*, 378-382.
- Hunsley, J., & Johnston, C. (2000). The role of empirically supported treatments in evidence-based psychological practice: A Canadian perspective. *Clinical Psychology: Science and Practice, 7*, 269-272.
- Hunsley, J., Lee, C. M., & Wood, J. M. (in press). Controversial and questionable assessment techniques. In S. O. Lilienfeld, J. M. Lohr, & S. J. Lynn (Eds.), *Science and pseudoscience in contemporary clinical psychology*. New York: Guilford Press.
- Keiser, R. E., & Prather, E. N. (1990). What is the TAT? A review of ten years of research. *Journal of Personality Assessment, 55*, 800-803.
- Kleiger, J. H. (1992). A conceptual critique of the EA:es comparison in the Comprehensive Rorschach System. *Psychological Assessment, 4*, 288-296.
- Kubiszyn, T. W., Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., & Eisman, E. J. (2000). Empirical support for psychological assessment in clinical health care settings. *Professional Psychology: Research and Practice, 31*, 119-130.
- Lerner, P. M. (2000). A nonreviewer's comment: On the Rorschach and baseball. *Journal of Clinical Psychology, 56*, 439.
- Lett, J. (1990). A field guide to critical thinking. *Skeptical Inquirer, 14*(2), 153-160.
- Marlowe, D. B. (1995). A hybrid decision framework for evaluating psychometric evidence. *Behavioral Sciences & the Law, 13*, 207-228.
- 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 Personality Assessment, 70*, 125-144.
- McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Bulletin, 96*, 690-702.
- Meyer, G. J. (1999). The convergent validity of the MMPI and Rorschach scales: An extension using profile scores to define response and character styles on both methods and a re-examination of simple Rorschach response frequency. *Journal of Personality Assessment, 72*, 1-35.
- Meyer, G. J., & Handler, L. (1997). The ability of the Rorschach to predict subsequent outcome: A meta-analysis of the Rorschach Prognostic Rating Scale. *Journal of Personality Assessment, 69*, 1-38.
- Morton-Page, D., & Wheeler, M. S. (1997). Early loss, depression, and lifestyle themes. *Individual Psychology, 53*, 58-66.
- Murray, H. A. (1943). *Thematic Apperception Test manual*. Cambridge, MA: Harvard University Press.
- Parker, K. C. H., Hanson, R. K., & 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 reply to Garb, Florio, and Grove. *Psychological Science, 10*, 291-292.
- Perry, W., & Viglione, D. J. (1991). The Ego Impairment Index as a predictor of outcome in melancholic depressed patients treated with tricyclic antidepressants. *Journal of Personality Assessment, 56*, 487-501.
- Piotrowski, C. (1999). Assessment practices in the era of managed care: Current status and future directions. *Journal of Clinical Psychology, 55*, 787-796.
- Piotrowski, C., & Belter, R. W. (1999). Internship training in psychological assessment: Has managed care had an impact? *Assessment, 6*, 307-315.
- Piotrowski, C., Belter, R. W., & Keller, J. W. (1998). The impact of "managed care" on the practice of psychological testing: Preliminary findings. *Journal of Personality Assessment, 70*, 441-447.
- Piotrowski, C., & Keller, J. W. (1989). Use of assessment in mental health clinics and services. *Psychological Reports, 64*, 1298.
- Rapaport, D., Gill, M., & Shafer, R. (1946). *Diagnostic psychological testing*. Chicago, IL: Year Book.
- Rudd, M. D., Joiner, T. E., Jobes, D. A., & King, C. A. (1999). The outpatient treatment of suicidality: An integration of science and recognition of its limitations. *Professional Psychology: Research and Practice, 30*, 437-446.
- Ryan, J. J., Prifitera, A., & Powers, L. (1983). Scoring reliability on the WAIS-R. *Journal of Consulting and Clinical Psychology, 51*, 149-150.
- Shaffer, T. W., Erdberg, P., & Haroian, J. (1999). Current nonpatient data for the Rorschach, WAIS-R, and MMPI-2. *Journal of Personality Assessment, 73*, 305-316.
- Slate, J. R., Jones, C. R., Coulter, C., & Covert, T. L. (1992). Practitioners' administration and scoring of the WISC-R: Evidence that we do err. *Journal of School Psychology, 30*, 77-82.
- Slate, J. R., Jones, C. R., Murray, R. A., & Coulter, C. (1993). Evidence that practitioners err in administering and scoring the WAIS-R. *Measurement and Evaluation in Counseling and Development, 25*(4), 156-161.
- Smith, A. M., Gacono, C. B., & Kaufman, L. (1997). A Rorschach comparison of nonpsychopathic conduct disordered adolescents. *Journal of Clinical Psychology, 53*, 289-300.
- Smith, A. M., Gacono, C. B., & Kaufman, L. (1998). Erratum: A Rorschach comparison of psychopathic and nonpsychopathic conduct disordered adolescents. *Journal of Clinical Psychology, 54*, 1151.
- Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin, 112*, 140-154.
- Stricker, G., & Gold, J. R. (1999). The Rorschach: Toward a nomothetically based, idiographically applicable configurational model. *Psychological Assessment, 11*, 240-250.
- Te'eni, D. R. (1998). Nomothetics and idiographics as antonyms: Two mutually exclusive purposes for using the Rorschach. *Journal of Personality Assessment, 70*, 232-247.
- Viglione, D. J. (1996). Data and issues to consider in reconciling self-report and the Rorschach. *Journal of Personality Assessment, 67*, 579-587.
- Viglione, D. J. (1997). Problems in Rorschach research and what to do about them. *Journal of Personality Assessment, 68*, 590-599.
- Viglione, D. J. (1999). A review of recent research addressing the utility of the Rorschach. *Psychological Assessment, 11*, 251-265.
- Watkins, C. E., Jr., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 26*, 54-60.
- Wechsler, D. (1955). *WAIS manual: Wechsler Adult Intelligence Scale*. New York: Psychological Corporation.
- Weiner, I. B. (1966). *Psychodiagnosis in schizophrenia*. New York: Wiley.
- Weiner, I. B. (1997). Current status of the Rorschach Inkblot Method. *Journal of Personality Assessment, 68*, 5-19.
- Weiner, I. B. (1999). What the Rorschach can do for you: Incremental validity in clinical applications. *Assessment, 6*, 327-339.

- Weiner, I. B. (2000). Using the Rorschach properly in practice and research. *Journal of Clinical Psychology, 56*, 435-438.
- Weiner, I. B. (2001). Advancing the science of psychological assessment: The Rorschach Inkblot Method as exemplar. *Psychological Assessment, 13*, 423-432.
- Weiner, I. B., Exner, J. E., & Sciara, A. (1996). Is the Rorschach welcome in the courtroom? *Journal of Personality Assessment, 67*, 422-424.
- Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117*, 450-468.
- Whitten, J., Slate, J. R., Jones, C. H., & Shine, A. E. (1994). Examiner errors in administering and scoring the WPPSI-R. *Journal of Psychoeducational Assessment, 12*, 49-54.
- Wood, J. M., & Lilienfeld, S. O. (1999). The Rorschach inkblot test: A case of overstatement? *Assessment, 6*, 341-349.
- Wood, J. M., Lilienfeld, S. O., Garb, H. N., & Nezworski, M. T. (2000). The Rorschach test in clinical diagnosis: A critical review, with a backward look at Garfield (1947). *Journal of Clinical Psychology, 56*, 395-430.
- Wood, J. M., Nezworski, M. T., Garb, H. N., & Lilienfeld, S. O. (2001). The misperception of psychopathology: Problems with the norms of the Comprehensive System for the Rorschach. *Clinical Psychology: Science and Practice, 8*, 350-373.
- Wood, J. M., Nezworski, T., & Stejskal, W. J. (1996). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science, 7*, 3-10.

Received May 11, 2000

Revision received November 14, 2000

Accepted April 23, 2001 ■

Call for Nominations

The Publications and Communications Board has opened nominations for the editorships of *Journal of Experimental Psychology: Animal Behavior Processes*, *Journal of Personality and Social Psychology: Personality Processes and Individual Differences*, *Journal of Family Psychology*, *Psychological Assessment*, and *Psychology and Aging* for the years 2004-2009. Mark E. Bouton, PhD, Ed Diener, PhD, Ross D. Parke, PhD, Stephen N. Haynes, PhD, and Leah L. Light, PhD, respectively, are the incumbent editors.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2003 to prepare for issues published in 2004. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations are also encouraged.

Search chairs have been appointed as follows:

- Lucia A. Gilbert, PhD, and Linda P. Spear, PhD, for *JEP: Animal*
- Sara Kiesler, PhD, for *JPSP: PPID*
- Susan H. McDaniel, PhD, and Mark I. Appelbaum, PhD, for the *Journal of Family Psychology*
- Lenore W. Harmon, PhD, for *Psychological Assessment*
- Randi C. Martin, PhD, and Joseph J. Campos, PhD, for *Psychology and Aging*

To nominate candidates, prepare a statement of one page or less in support of each candidate. Address all nominations to the appropriate search committee at the following address:

Karen Sellman, P&C Board Search Liaison
 Room 2004
 American Psychological Association
 750 First Street, NE
 Washington, DC 20002-4242

The first review of nominations will begin December 14, 2001.