Psychology, Public Policy, and Law
2002, Vol. 8, No. 2, 216–234
Copyright 2002 by the American Psychological Association, Inc.
1076-8971/02/$5.00 DOI: 10.1037//1076-8971.8.2.216

Failure of Rorschach-Comprehensive-System-Based Testimony to Be Admissible Under the *Daubert–Joiner–Kumho* Standard

William M. Grove
University of Minnesota, Twin Cities Campus

R. Christopher Barden
National Association for Consumer Protection in Mental Health Practices

Howard N. Garb
Pittsburgh Veterans Affairs Healthcare System and University of Pittsburgh

Scott O. Lilienfeld
Emory University

The Comprehensive System for the Rorschach (RCS) is currently the subject of intense scientific criticism. The normative data for many RCS scores are seriously in error and tend to make normal individuals appear maladjusted. Reliability is inadequate for many RCS scores, and validity for the great majority of RCS scores has not been adequately demonstrated. In addition, a substantial number of Rorschach Workshops studies, cited by B. Ritzler, R. Erard, and G. Pettigrew (2002) as supportive of the RCS, are unavailable for examination. Finally, B. Ritzler et al. misinterpret central issues of the relevant legal analysis, including crucial legal standards. The RCS clearly fails to meet the standards for admissibility set forth in the *Daubert*, *Joiner*, and *Kumho* decisions.

Should expert testimony based on the widely used Comprehensive System for the Rorschach (RCS) be admissible in court? In an earlier article that appeared in this journal (Grove & Barden, 1999), two of us discussed the admissibility of testimony from mental health experts. We concluded that some types of testimony, commonly offered by mental health professionals, do not meet the current standards that govern the admissibility of scientific evidence in federal and many state courts (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 1993; *General Electric v. Joiner*, 1997; *Kumho Tire Co., Ltd. v. Carmichael*, 1999). In 3 of the 19 pages of our article, we sketched a critique of the scientific and legal status of the RCS. We briefly outlined the reasons why the RCS does not meet the standards for admissibility set forth in the *Daubert*, *Joiner*, and *Kumho* decisions (which we hereafter refer to as *Daubert*).

Not surprisingly, our article has elicited a lengthy and impassioned defense of the RCS by three of its most noted proponents: Ritzler, Erard, and Pettigrew (2002). We welcome the opportunity to respond to them. During the past 5 years, the RCS has become the subject of intense scientific controversy. Ritzler et al. have presented one side of this debate. We present the other. As we show, Ritzler

William M. Grove, Department of Psychology, University of Minnesota, Twin Cities Campus; R. Christopher Barden, National Association for Consumer Protection in Mental Health Practices, West Salt Lake City, Utah; Howard N. Garb, Behavioral Health, Pittsburgh Veterans Affairs Healthcare System, and Department of Psychiatry, University of Pittsburgh; Scott O. Lilienfeld, Department of Psychology, Emory University.

Correspondence concerning this article should be addressed to William M. Grove, Department of Psychology, University of Minnesota, Twin Cities Campus, N218 Elliott Hall, 75 East River Road, Minneapolis, Minnesota 55455–0344. E-mail: william.m.grove-1@umn.edu

et al. have contradicted themselves, have made bold assertions without scientific support, and have neglected to discuss numerous scientific findings that contradict their conclusions.

We make five points that are critical to evaluating the legal admissibility of the RCS: (a) The RCS is currently the subject of intense scientific controversy, (b) the norms for many RCS variables are seriously in error and tend to make normal individuals appear pathological; (c) a substantial proportion of RCS variables probably cannot be scored at a level of reliability that is adequate for clinical and forensic use; (d) the validity of most RCS scores is questionable or undemonstrated; and (e) the data and studies of John Exner, which constitute the scientific foundation of the RCS, appear to be largely unavailable to the general scientific community for scrutiny. After discussing these five points, we delineate their relevance to the law and explain why, contrary to the assertions of Ritzler et al. (2002), the RCS fails to meet the standards for admissibility set forth in *Daubert* and subsequent cases.

The Controversy Surrounding the Comprehensive System

The most popular system for administering, scoring, and interpreting the Rorschach is the RCS (Exner, 1991, 1993, 2001b). However, in the past several years, the RCS has been engulfed in intense scientific controversy. Special sections on this controversy, and heated exchanges between advocates and critics, have appeared in just the past 2 years in the following professional journals: *Assessment*, *Clinical Psychology: Science and Practice*, *Journal of Clinical Psychology*, *Journal of Forensic Psychology Practice*, *Journal of Personality Assessment*, and *Psychological Science in the Public Interest*. In addition, this controversy has been covered by general news and popular science outlets, including the *New York Times* (Goode, 2001) and *Scientific American* magazine (Lilienfeld, Wood, & Garb, 2001). It seems warranted to conclude that a majority (indeed, in all likelihood, a substantial majority) of the relevant scientific community (comprising the predominantly academic university scientists who conduct and review the research in the field of personality assessment) does not view the RCS as a reliable system with either broad-based zero-order validity or forensically useful incremental validity. In sum, the RCS controversy is far from over, and the perception of the RCS in the eyes of a substantial proportion of the informed scientific community is decidedly negative.

Problems With the Comprehensive System Norms

An ostensible strength of the RCS is its provision of extensive normative data. Normative data describe how relatively normal individuals in the community perform on a measure. Ritzler et al. (2002) have referred to the "nationally representative norms" (p. 207) collected by Exner and his colleagues and have argued that they form a valuable basis of support for the RCS. By comparing a client's results with those normative data, one can make inferences about which scores are relatively high or low as compared with relatively normal individuals in the community. For example, if a client's score on a given RCS variable was 2 standard deviations above the mean score for individuals in the community, a

psychologist would presumably conclude that this person was likely to be extreme on that variable.

Serious problems exist with the RCS norms, although such problems were not mentioned by Ritzler et al. (2002). Perhaps most seriously, there is evidence that use of the RCS norms leads to the overperception of psychopathology; that is, when normal individuals are compared with the RCS norms, they too often tend to appear pathological.

Several recent studies have consistently found that American adults living in the community score in the maladjusted range when the RCS norms are used to interpret their results (Shaffer, Erdberg, & Haroian, 1999; Wood, Nezworski, Garb, & Lilienfeld, 2001a, 2001b; also see Aronow, 2001; Exner, 2001a; Hunsley & Di Giulio, 2001; Meyer, 2001; Widiger, 2001). For example, in a recent study of 123 nonpatient adults, Shaffer et al. (1999) found that about one-sixth of participants scored in the pathological range (>4) on the RCS Schizophrenia index (SCZI). Over one-fourth (29%) gave at least one Reflection response, a supposedly rare Rorschach indicator of narcissism (Exner, 1991). Furthermore, these presumably normal adults often appeared disturbed on other Rorschach measures of Perceptual Accuracy, Distorted Thinking, and Emotional Functioning. Similar findings were reported by Wood et al. (2001b), who reviewed the results for 14 RCS variables in 32 studies of nonpatient adults. They concluded that the RCS norms do not currently represent American nonpatient adults and probably never did.

The children's norms for the RCS appear to have similar problems. For example, Hamel, Shaffer, and Erdberg (2000) administered the RCS to a group of 100 children who had been screened for mental health problems. Overall, these children had better than average mental health, as measured by the Conners Parent Rating Scale (Conners, 1989). Nevertheless, the RCS scores for these children were strikingly discrepant from the RCS norms. As observed by Hamel et al. (2000),

> If we were writing a Rorschach-based, collective psychological evaluation for this sample, the clinical descriptors would command attention. In the main, these children may be described as grossly misperceiving and misinterpreting their surroundings and having unconventional ideation and significant cognitive impairment. Their distortion of reality and faulty reasoning approach psychosis. These children would also likely be described as having significant problems establishing and maintaining interpersonal relationships and coping within a social context. They apparently suffer from an affective disorder that includes many of the markers found in clinical depression. (p. 291)

In other words, for these essentially normal children, the use of the RCS would readily lead to gross overattributions of psychopathology, with potentially devastating consequences for fundamental legal rights.

Additional problems with the RCS norms have recently come to light. For example, of the 700 protocols constituting the 1991/1993 adult normative sample, 221 were duplicates (Exner, 2001b; J. E. Exner, personal communication, March 23, 2001). That is, the sample included only 479 distinct protocols, with 221 of the 479 protocols counted twice. This error was only recently revealed to the scientific community.

The 1991/1993 adult norms of the RCS have recently been revised, using the same pool of protocols originally collected in the late 1970s and early 1980s (Exner, 2001b). This revised adult normative sample also suffers from major problems. First, the use of these norms is likely to lead to the overperception of psychopathology because many of the protocols in the 2001 adult normative sample were also included in the 1993 sample (results for the 1991/1993 and 2001 normative samples are highly similar). Also, the protocols were apparently erroneously scored for Form Quality using outdated rules (Meyer & Richardson, 2001). That is, the protocols are actually 15–20 years old, and they were scored for Form Quality using rules that have subsequently been revised and thus are not comparable to currently scored protocols. It is sobering to realize that because of these scoring errors, the normative data for a number of important RCS variables (e.g., X + %, X - %, S - %, and the SCZI) have probably been seriously in error since the early 1980s.

In summary, the RCS norms are flawed, and their use can lead clinicians to overperceive psychopathology. It is striking that Ritzler et al. (2002) praised the RCS norms, yet they did not cite any evidence that would support their representativeness, and they never acknowledged that serious problems exist. The RCS norms are *not* "nationally representative" as Ritzler et al. (2002) claimed (p. 207) and, as previously stated, their use tends to misclassify normal people as pathological. Ritzler et al.'s failure to address this issue is a glaring omission: The problems with the RCS norms can lead to serious and systematic errors in conclusions about psychopathology when the RCS is used. These problems are directly relevant to the RCS's legal admissibility.

Inadequate Reliability of Comprehensive System Scores

According to Ritzler et al. (2002), the Rorschach is "valid and reliable" (p. 201). However, they did not cite, let alone discuss, so much as a single study on reliability, leading us to wonder whether Ritzler et al. propose proof by *ipse dixit*. From this evidentiary vacuum, they attempted to rescue the CS by noting that the RCS provides "specific instructions for administration and coding" (p. 210). They also argued that "some psychologists do not adhere to these standards when using the RCS, and as a result, their testimony would not qualify as expert under this guideline" (p. 210) Ritzler et al. (2002) clearly imply by contrast with such practices that interrater reliability will be adequate whenever psychologists *do* adhere to RCS standards for administration and scoring.

Such claims by Rorschach proponents regarding the interrater reliability of the RCS have often been seriously overstated (Wood & Lilienfeld, 1999). For instance, Ritzler (1995) contended that "every variable in the Comprehensive System has demonstrated substantial interrater reliability" (p. 230). This claim is clearly wrong and has been refuted by the results of several important studies (Acklin, McDowell, Verschell, & Chan, 2000; Gronnerod, 1999; Nakata, 2000; also see Meyer, 1997a, 1997b; Shaffer et al., 1999; Wood, Nezworski, & Stejskal, 1997). For example, in Acklin et al. (2000), Rorschach protocols were scored by two clinical psychology graduate students with advanced training and at least 3 years of experience in the use of the RCS. They scored protocols for clinical (n = 20) and nonclinical (n = 20) samples. When intraclass correlation coefficients

were calculated for 95 RCS scores, median reliability was in the low .80s for both samples. Although Exner (1993) has claimed for many years that scoring reliability is uniformly above a minimum acceptable threshold of .85 for RCS variables, Acklin et al. reported that about half of the RCS scores had reliability coefficients below Exner's own required minimum for RCS scores. In particular, reliability coefficients in the two samples were unacceptably low for the SCZI (.45 and .56), Adjusted D (.53 and .68), and X - % (.62 and .66). (Adjusted D is purported to be a measure of self-control under stress, and X - % is considered to be a measure of perceptual and mental distortion; Exner, 1991, 1993.)

What constitutes adequate reliability for a psychological test that is used in clinical and forensic settings? The classic text on psychological testing by Nunnally and Bernstein (1994) stated that reliability for tests used in clinical practice should be at least .90 and preferably will be at least .95 (also see Kaplan & Saccuzzo, 2001). Exner (1978) accepted a slightly lower standard for the Rorschach, selecting .85 as the minimum acceptable level of reliability for RCS scores; see also Groth-Marnat (1997).

Regarding the reliability of measures used in forensic settings, we would argue that the most relevant guidelines were those of Heilbrun (1992).[1] Heilbrun (1992) asserted that in forensic settings "the use of tests with a reliability coefficient of less than .80 is not advisable" (p. 265). He specifically noted that this guideline would apply to the interrater reliability of measures such as the Rorschach "in which clinical judgment plays a significant part in data combination or interpretation" (Heilbrun, 1992, p. 265).

Whether one accepts Nunnally and Bernstein's (1994) relatively high standard of .95 or Heilbrun's (1992) relatively low one of .80, a great many RCS scores do not possess the minimum reliability necessary for clinical and forensic applicability. Furthermore, overall reliability will only deteriorate further when clinicians are asked to both score *and* interpret protocols, as occurs in routine clinical practice. Indeed, there is little published evidence on the interrater reliability of RCS clinical interpretations, which may pose an even more formidable threat to the RCS's scientific status than does its interscorer reliability. For comparison purposes, it can be demonstrated that interrater scoring reliability is substantially higher for the Wechsler Adult Intelligence Scale, Third Edition (WAIS–III; Wechsler, 1997) and the Minnesota Multiphasic Personality Inventory–II (MMPI–2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). Intraclass correlation coefficients for the WAIS–III subtests have a minimum value of .90 and a median value of .95 (Psychological Corporation, 1997). Of course, aside from clerical errors, interrater scoring reliability is not a relevant concern for most computer-scored self-report measures like the MMPI–2.

In conclusion, although Ritzler et al. (2002) argued that the Rorschach is reliable, they did not describe results from a single study to support their claim. In contrast, after describing results from relevant studies, Lilienfeld, Wood, and Garb (2000) concluded that

> scoring reliability is problematic for a substantial number of RCS variables and the use of these variables to assess individual clients is inadvisable. For example, even

[1]This article was also cited by Ritzler et al. (2002).

> among psychologists who are highly experienced with the RCS or regarded as authorities, Rorschach scoring is not necessarily above challenge. Disagreements can have particularly serious implications if the test results are used to reach important clinical or legal recommendations. (p. 34)

Although it is not reviewed in this article, note that research on test–retest reliability also raises significant questions about the clinical and forensic use of the RCS (for reviews and detailed citations, see Garb, Wood, Nezworski, Grove, & Stejskal, 2001; Wood & Lilienfeld, 1999).

Lack of Satisfactory Evidence for the Validity of the Comprehensive System

Although they did not describe research on norms or reliability, Ritzler et al. (2002) did describe results from selected validity studies. They concluded that many RCS indices possess acceptably high levels of validity. Although we agree that a few Rorschach scores are nonnegligibly valid, the substantial majority of Rorschach scores have not demonstrated consistent relations to personality traits or psychological disorders. We discuss (a) results on how well the Rorschach can be used to assess personality traits and mental disorders, (b) results from global meta-analyses of validity studies, and (c) results of studies on incremental validity.

Validity of Individual Scores

To identify valid Rorschach scores, Wood, Nezworski, and Stejskal (1996b) set forth the following criteria. First, a score must demonstrate a consistent relation to a particular symptom, trait, or disorder. Second, results must be obtained in methodologically adequate studies. Third, positive findings must be replicated by independent investigators.

Attempts have been made to identify Rorschach scores that have been consistently related to a symptom or disorder in methodologically sound studies that were conducted by independent investigators (Lilienfeld et al., 2000; Wood, Lilienfeld, Garb, & Nezworski, 2000). With few exceptions, the great majority of Rorschach scores have not been supported by studies satisfying the aforementioned criteria. Only the following indexes meet these criteria:

> (a) Thought Disorder Index for the Rorschach (TDIR) in the assessment of thought disorder, (b) Rorschach Prognostic Rating Scale (RPRS) in the prediction of treatment outcome, (c) Rorschach Oral Dependency (ROD) Scale in the assessment of objective behaviors related to dependency, and (d) deviant verbalizations and poor form (as well as the CS, SCZI, and other indexes derived from these variables) in the assessment of schizophrenia . . . and borderline personality disorder. (Lilienfeld et al., 2000, p. 54)

Interestingly, although the RCS is composed of over 100 variables, only those measuring deviant verbalizations and poor form satisfied Wood et al.'s (1996b) criteria. The TDIR, RPRS, and ROD are not even part of the RCS. Hence, no matter how valid these indices may be, they cannot logically support the validity of the RCS. Moreover, one RCS variable that has been well validated, the SCZI,

is no longer part of the RCS (Exner, 2001b). It is ironic that one of the very few RCS scores that have been well validated was eliminated from the system.

According to Ritzler et al. (2002), "The RCS literature is replete with studies showing a positive correspondence between RCS variables and personality characteristics and behaviors" (p. 206). The reasons why Ritzler et al. and Lilienfeld et al. (2000) reached markedly different conclusions are quite evident. Lilienfeld et al. concluded that Rorschach variables were well supported by research only if the scores were consistently related to a symptom, trait, or disorder in methodologically sound studies conducted by independent investigators. Ritzler et al. did not use these criteria, nor did they state what criteria, if any, they adopted.

An example suffices to illustrate why these criteria are important. Ritzler et al. cited the work of Gacono and Meloy (1994) as demonstrating a positive correlation between RCS variables and aggressive and psychopathic personality characteristics. However, many of the studies conducted by Gacono, Meloy, and their colleagues are seriously flawed because they involved comparisons of results for patients with normative data. As discussed earlier, serious problems exist with the RCS norms. Erroneous norms can, when compared with scores from normal individuals, misclassify those individuals as disordered. "Pathological" Rorschach scores from a clinical group could likewise result from overpathologizing RCS norms, not from good construct validity. Hence, it is not surprising that individuals with aggressive and psychopathic personalities differed from the norms on a number of scores. Findings concerning the relation between the RCS and aggressive and psychopathic personality disorder (ASPD) characteristics were described in a comprehensive review by Wood et al. (2000):

> It appears that only four studies have used a comparison group, rather than normative data, when studying individuals with a diagnosis of ASPD. In addition, three studies have examined the relationship of Rorschach scores to the number of diagnostic criteria for ASPD in mixed patient groups. . . . It has been suggested that a large number of Rorschach variables are related to ASPD (Gacono & Meloy, 1994). However, a review of the seven empirical studies cited here indicates that most of these variables have never been replicated independently (e.g., Adjusted D, $X+\%$, Blood, Boundary Disturbance) or have been found in independent replications to be unrelated to ASPD. (p. 411)

Similarly, Wood et al. (2000) concluded that Rorschach scores are not significantly related to either conduct disorder or psychopathy.

Ritzler et al. (2002) singled out the following RCS variables as valid: RCS Schizophrenia Index (SCZI), Rorschach Oral Dependency Scale (ROD), Rorschach Prognostic Rating scale (RPRS), Human Experience Variable (HEV), Depression Index (DEPI), and Suicide Constellation (S-CON). Note that although there are over a hundred variables in the RCS, Ritzler et al. mentioned only these six by name. Three of the six variables identified by Ritzler et al. were also listed by Lilienfeld et al. (2000) as valid, so there is some agreement between the two sides. We will comment on these six variables.

Ritzler et al. (2002) and Lilienfeld et al. (2000) agree that three of the variables have been reasonably well supported. Nevertheless, two of them (ROD, RPRS) were never part of the RCS, and the third (SCZI) no longer is. Ritzler et al.'s listing of these scores is especially troubling because they informed readers

they would describe research *only* on the RCS: "Because Grove and Barden (1999) directed their argument specifically toward the RCS, this approach is the sole method under consideration in this reply" (Ritzler et al., 2002, p. 201, footnote 1). Meta-analyses support the validity of the ROD and RPRS (Bornstein, 1997, 1999; Meyer & Handler, 1997; 2000), but no meta-analyses focusing on specific RCS variables have been published. Ritzler et al. may have included the ROD and RPRS in their list of well-validated variables because they have been widely studied, but it is erroneous to infer from the results for these non-RCS scores that the RCS is well validated.

We agree with Ritzler et al. (2002) that the RPRS, ROD, and SCZI are valid, although these variables are plainly irrelevant to the current debate, insofar as none are now part of the RCS. However, the additional claim of Ritzler et al. of adequate validity for the RCS scores of DEPI, HEV, and S-CON is not supported by the extant evidence. Results for these variables have been mixed or negative (Garb et al., in press; Jorgensen, Andersen, & Dam, 2000; Wood et al., 2000; Wood, Nezworski, & Stejskal, 1996a; Wood, Nezworski, Stejskal, Garven, & West, 1999). For example, Ritzler et al. scrupulously cited a study by Meyer (2000) that supports the use of the DEPI to diagnose depressive disorders, but their article overlooks entirely a series of studies that have failed to detect a statistically significant relation between the DEPI and diagnoses of depression (Archer & Gordon, 1988; Archer & Krishnamurthy, 1997; Ball, Archer, Gordon, & French, 1991; Caine, Frueh, & Kinder, 1995; Carlson, Kula, & St. Laurent, 1997; Carter & Dacey, 1996; Kadle, 1989; Lipkin, 1989; Lipovsky, Finch, & Belter, 1989; Meyer, 1993; Sells, 1990/1991; Silberg & Armstrong, 1992; Viglione, Brager, & Haller, 1988; but for mixed findings, see Ilonen et al., 1999, and for positive findings regarding the DEPI, see Jansak, 1996/1997, and Singer & Brabender, 1993). Although Ritzler et al. argued that Grove and Barden (1999) did not provide a balanced review of research on the Rorschach, they ironically omitted mention of every negative DEPI (or HEV or S-CON) study and reached the conclusion that the DEPI is valid in the face of a vast majority of negative studies.

In arguing that many RCS scores are valid, Ritzler et al. (2002) also referred to unnamed scores ostensibly related to aggressive and psychopathic personality traits (Gacono & Meloy, 1994), other unnamed variables purportedly related to treatment planning and outcome assessment (Weiner, 1999), "the many studies" (p. 207) conducted by Exner (no references given), the "large body of literature that was used originally by Exner (1993) to develop the RCS" (no references given; pp. 206–207), and Exner's "nationally representative norms" (p. 207). It is difficult to rebut Ritzler et al.'s assertions because bodies of research are alluded to, without clear claims about which scores have been validated, in which populations, and against which criteria. It is possible, however, to address Ritzler et al.'s (2002) claims about the norms; as already noted, serious problems exist with them and they are most certainly not nationally representative. In addition, as already discussed, RCS scores have not shown a clear or consistent relationship to antisocial personality disorder, conduct disorder, or psychopathy (Wood et al., 2000).

According to Ritzler et al. (2002), over 200 books and between 8,000 and 9,000 articles have been published on the RCS. It is troubling that despite this

large body of literature, validity has not yet been established for the substantial majority of RCS scores. Moreover, Ritzler et al.'s invocation of the large body of published literature on the RCS is a striking example of the logician's *ad populum* fallacy. There is no logical relationship between popularity and validity. At one time, for example, phrenology was in vogue.

Compounding the problem, Ritzler et al. (2002) mistakenly implied that this extensive body of published literature demonstrated that the Rorschach had been subjected to rigorous peer review. They criticized Grove and Barden's (1999) assertion that "pro-Rorschach studies often appear in a specialty journal, the *Journal of Personality Assessment*"(p. 228) on the grounds that this journal has a 60-year history of publication and high manuscript rejection rates. Nevertheless, Ritzler et al. fail to point out that this journal, which is undeniably the primary source of prominent publications on the Rorschach, has a lengthy history of close association with this technique. This journal was initially titled *Rorschach Research Exchange* and was later titled *The Journal of Projective Techniques* and the *Journal of Projective Techniques and Personality Assessment* before acquiring its present title. Moreover, all five past and present editors of this journal (Bruno Klopfer, Walter Klopfer, Irving Weiner, Bill Kinder, and now Gregory Meyer) are well-known proponents of the Rorschach and Rorschach researchers. It is clear that the primary repository of peer-reviewed published research on the Rorschach cannot be regarded as an impartial or neutral source of information.

Results From Global Meta-Analyses

Several meta-analyses have been conducted to estimate the average validity of Rorschach scores (e.g., Garb, Florio, & Grove, 1998; Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999; Parker, Hanson, & Hunsley, 1988). In these meta-analyses, results for many different Rorschach scores were pooled. For this reason, we refer to them as *global* meta-analyses.

The interpretation of these meta-analyses is complicated by several methodological problems. First, all of the results are based on relatively small sample sizes. For example, Ritzler et al. (2002) made a special point of noting that a validity coefficient of .41 was obtained in one meta-analysis, but this coefficient was based on only five studies (Parker et al., 1988, p. 370). Considering that Ritzler et al. (2002) reported that between 8,000 and 9,000 articles have been published on the Rorschach, finding five studies yielding (relatively) high validity for RCS indices is not especially encouraging, particularly given that not all five of these findings are based on the RCS. Second, unpublished studies were not included in any of the meta-analyses. This point is important because studies with positive findings are often more likely to be published than studies with negative findings (this is known as *publication bias* or the *file-drawer* effect; Rosenthal, 1979; see also Lilienfeld et al., 2000). A third problem is that inappropriate studies have sometimes been included in meta-analyses. For example, in some studies, a clinical group was compared with the RCS norms. Significant differences were found, and these positive findings were included in the meta-analysis. However, as we have already indicated, even results for groups of relatively normal individuals can be expected to differ from the RCS norms. Thus, positive findings

may have been obtained because there is a problem with the RCS norms, not because Rorschach indices are valid.

As noted by Grove and Barden (1999), average validity coefficients for the Rorschach have generally been approximately .30. These results provide some support for the Rorschach, but they do not demonstrate that all, most, or even many RCS scores are valid. In fact, global meta-analyses tell us little about what RCS scores are valid for what purpose. Results from global meta-analyses have not told us what scores are related to what symptoms or disorders, whether the studies are methodologically sound, or whether independent investigators have replicated positive findings. Thus, the results of global meta-analyses do not provide compelling support for the use of the Rorschach.

Finally, it can be pointed out that Ritzler et al. (2002) are inconsistent in their views concerning the usefulness of global meta-analyses. On the one hand, they cite results from global meta-analyses to buttress their claim that the RCS is valid. On the other hand, they also argue, "furthermore, it is neither possible nor reasonable to apply the concept of error rate to the RCS as a whole. Indeed, it is inappropriate to speak globally of the 'validity of the test'" (p. 208). Thus, Ritzler et al. (2002) clearly contradict themselves and invoke evidence from global meta-analyses in support of their claims, despite their admonition that such evidence is not meaningful.

Incremental Validity

An assessment instrument possesses incremental validity if validity increases when results for this instrument are added to other data. For example, one may ask psychologists to make judgments using history, interview, and objective test information. One could then ask them to make a second set of judgments using the history, interview, and objective test information plus Rorschach results. It is essential to examine incremental validity because the addition of a test to other information may not result in an increase in accuracy and may even result in a decrease in accuracy (Sechrest, 1963).

In clinical judgment studies, when psychologists have been given increasing amounts of information, the addition of the Rorschach has almost never led to an increase in the validity of judgments (for reviews, see Garb, 1984, 1998). Ritzler et al. (2002) appeared to disagree with this conclusion, but they did not cite a single study in which psychologists made more accurate judgments after being given Rorschach information.[2] It further weakens Ritzler et al.'s case that in several studies, validity actually decreased when the Rorschach was added to other information, even when psychologists used the RCS (Whitehead, 1985).

In other studies, multiple regression equations were used to make predictions. In these investigations, Rorschach scores were added to other variables to determine if the addition of the scores led to an increase in validity. Although of

[2]Ritzler et al. (2002) stated that they disagreed with the findings of Garb, but they cited his doctoral dissertation, not his two published reviews of the literature on incremental validity. Findings on the incremental validity of the Rorschach were not reviewed in Garb's dissertation. In fact, Garb (1985) did not even mention the Rorschach or any other projective technique in his dissertation.

interest, statistical incremental validity need not correspond to clinical incremental validity (i.e., the extent to which increased psychometric information leads clinicians to make more valid judgments) —and almost all use of the RCS appears to rely on clinical, not strictly statistical, score combination. Positive statistical results were found for a few RCS scores, but none of these results have been replicated by independent investigators (Lilienfeld et al., 2000). Perhaps the fairest conclusion that can be reached is that incremental validity has not been established for the vast majority of RCS scores (Lilienfeld et al., 2000). Notably, Ritzler et al. (2002) did not list a single RCS score demonstrated to possess good incremental validity above and beyond other assessment information (e.g., data derived from self-report measures).

The findings on incremental validity severely undermine the assertions made by Ritzler et al. (2002) regarding the multimethod use of the RCS. According to Ritzler et al. (2002), "the multimethod approach . . . lies at the heart of good personality assessment" (p. 210). They argued that the high error rates of many RCS variables are largely irrelevant because the Rorschach is almost never interpreted in isolation. Nevertheless, they failed to mention that a multimethod approach is no panacea and can in fact backfire in certain cases. Because more information is not always better in clinical assessment, we need evidence rather than plausible speculations to show that adding the RCS to other assessment data will usefully improve validity. Available data sometimes show declines in aggregate validity when Rorschach data are added to assessment batteries (Garb, 1984).

Questions Surrounding the Availability of Data and Manuscripts

Very few scores in the RCS possess well-demonstrated validity. Yet until recently, many psychologists assumed otherwise, because they accepted without question numerous claims published in the books of John Exner, creator of the RCS (Exner, 1974, 1978, 1986, 1991, 1993; Exner & Weiner, 1982, 1995). Exner's books cite over a hundred studies to support the reliability and validity of the RCS. These citations provide the foundation of the RCS's claims to scientific legitimacy.

Serious problems exist with the research cited in Exner's books (Wood et al., 1996a, 1996b; but also see Exner, 1995). First, most of the research consists of unpublished studies by Exner's Rorschach Workshops. These studies have never been peer reviewed. Second, these papers are apparently unavailable for examination by other researchers. Although the studies are cited using a format that suggests that they are written papers (Author[s], year, title, and study reference number), in many cases the papers were never written. Wood et al. (1996a) described their attempts to obtain copies of the Rorschach Workshops papers:

> Many readers of TRACS [The Rorschach: A Comprehensive System] are probably under the impression that the Workshops Studies are actual documents that can be examined by other scholars. However, this impression is often mistaken. In preparation for writing this article, we requested 23 of the Workshops Studies cited in TRACS. Letters from the Rorschach Workshops informed us that some of the Workshops Studies were not in their files. The methods and results of the remaining studies either had not been formally written up or could not be released.

> We were informed that the Rorschach Workshops could provide raw data related to specific questions, but that we might have to pay for computer costs. (p. 8)

Because it would not be helpful to obtain the raw data without having written descriptions of the Method and Results sections, Wood et al. (1996a) did not request the data for these studies. However, several years later, Wood and his colleagues requested the data for the RCS norms (J. Wood, personal communication, August 5, 2000). Exner refused to provide them (J. Exner, personal communication, December 8, 2000).

It is deeply troubling that most of the fundamental papers that Exner cites in support of the RCS have never been peer reviewed or published, are unavailable for scrutiny by other researchers, or in some cases have never even been written. Ritzler et al. (2002) did not address all of these issues, but they did claim that Exner's (1993) text contains "over 400" peer-reviewed studies that provide "empirical support for the RCS" and that "fewer than 10% of the supporting studies were cited as 'unpublished manuscripts'" (p. 207). However, these figures are seriously overstated because Ritzler et al. used an inappropriately broad definition of "supporting studies." It is true that there are more than 400 citations in Exner's book. However, most have no direct connection to the RCS. For instance, Exner included many citations to studies that did not use the Rorschach (e.g., studies on schizophrenia and other disorders), as well as numerous citations from the 1950s and 1960s that antedate the RCS. If all of these hundreds of citations in Exner's (1993) book are grouped together, including those that do not deal with the RCS, then it is true that "fewer than 10%" of the references are unpublished manuscripts. However, the relevant question is, How many of Exner's (1993) citations *regarding the RCS* are to his own work, and how much of his work has been peer reviewed? The answer to this question is clear to anyone who glances through the book: Nearly all of Exner's (1993) RCS citations are to his own works. Of his cited works, only 24% are peer-reviewed articles, and over 50% are unpublished studies from the Rorschach Workshops.

Exner is not the only Rorschach researcher who has been unable or unwilling to share his data, although he is the only one (to our knowledge) who has consistently not provided copies of manuscripts for papers that he cites. For example, Ritzler et al. (2002) cited a study by Burns and Viglione (1996) as supporting the Rorschach Human Experience variable. However, when the data from this study were requested shortly after the article was published, they were reported to be lost and therefore unavailable for reanalysis (Wood, Nezworski, Stejskal, & Garven, 2001). In addition to being told that the data were lost, Wood and his colleagues were informed that all computer printouts were likewise unavailable (D. J. Viglione, personal communication, March 10, 1997). Such rapid and complete loss or destruction of published data is contrary to American Psychological Association journal policy and the custom in the field.

Ritzler et al. (2002) also cited a study by Meyer (1999) that ostensibly supported the validity of the Rorschach. When Meyer was asked by Howard N. Garb to share his data so that the appropriateness of his statistical analysis could be evaluated, he agreed to share some, but not all, of the data. In this case, there was a reasonable disagreement over what data were needed to evaluate the appropriateness of the statistical analysis, with the unfortunate consequence that

the data were not reanalyzed and questions about the statistical analysis remain unanswered.[3]

Implications for the Admissibility of Rorschach-Based Expert Testimony

The Rorschach is the subject of raging controversy and does not have a well-characterized error rate, hence failing *Frye* (*Frye v. United States*, 1923), *Daubert*, and relevance tests on these bases alone. Informed judges, the gatekeepers under *Daubert*, are the best defense against improper expert testimony. We join with Justice Stephen Breyer and others in encouraging courts to retain their own experts to analyze such scientific issues as the admissibility of Rorschach interpretations, thus avoiding forensic experts' misconduct. We strongly advise jurists to select such neutral experts by applying simple selection criteria: (a)The expert graduated from a major research university doctoral (PhD) or medical doctor (MD) residency program, fully accredited by the relevant professional associations; (b) the expert is employed as a tenured or tenure-track experienced faculty member at a major research university PhD or MD residency program, likewise fully accredited; (c) the expert has editorial board experience on journals published by the relevant major professional associations (e.g., American Psychological Association); (d) the expert is free of obvious financial conflicts of interest; and (e) (desideratum, but not necessary) the expert has received research grants as a principal investigator in substantial amounts from respected national granting agencies.

Unfortunately, typical court-appointed experts (especially those seen in family law and related matters) almost never meet these qualifications. If such qualified mental health experts cannot be found (or convinced to serve as a court-appointed expert), the court should rely on recommendations from local university psychology and medical school faculties (not professional schools or freestanding institutes) who conduct and publish relevant research. Teaching or clinical experience does not constitute scientific expertise, which is the basis of *Daubert* admissibility.

Ritzler et al. (2002) badly misinterpret central issues of the relevant legal analysis, including crucial legal standards. Legal professionals will note that Ritzler et al. appear deeply confused with regard to these standards. They mistakenly asserted that *Daubert* "requires" courts to rely on the "highest standards of the relevant professional community" (Ritzler et al., 2002, p. 211). In fact, however, the most cursory examination of *Daubert* demonstrates that the phrase "professional community" does not even appear in the decision. This error is especially troubling because Grove and Barden (1999) repeatedly and correctly quoted the key *Daubert* language: "the relevant scientific community."

Ritzler et al.'s (2002) error, which might appear to be merely semantic at first sight, is both crucial and tendentious. This is because clinicians, practitioners, and other nonscientist professionals have sought diligently to hijack the *Daubert* standards since they were first issued. The loser in many a *Daubert* motion attempts to argue that the "professional community" should decide what is scientifically reliable rather than the scientists. As it happens, the Rorschach (and

[3]Copies of this correspondence are available from Howard N. Garb.

many other controversial assessment and treatment methods in psychology) is popular among nonresearcher clinicians but much less well regarded among academic research scientists. Inasmuch as the U.S. Supreme Court has ruled that the relevant scientific community shall advise judges about admissibility of expert testimony, advocates of the Rorschach tend to operate at a marked disadvantage under *Daubert*. Just such a scientist–practitioner conflict was at the heart of the last several years' effort to banish repressed memory testimony from courtrooms (Barden, 2001a).

Another error of legal scholarship is Ritzler et al's (2002) claim that the Rorschach is "accepted by a relevant and substantial scientific community" (p. 201). Even if true, this would not constitute evidence that the RCS was "generally accepted" by "the [i.e., singular] relevant scientific community," which is what the Court requires. *Daubert* definitely does *not* propose that each party find a subpopulation of scientists (let alone nonresearcher clinicians) that has a favorable, if idiosyncratic, viewpoint. It was precisely this spectacle of "dueling experts," some or all relying on ill-proven hypotheses and shaky data, that *Daubert* tried to abolish.

Having reviewed Ritzler et al.'s (2002) idiosyncratic analysis of RCS credibility and the law, we have shown that judges and attorneys need not be confused by the vagaries of the raging Rorschach controversy. Clearly, the RCS is lacking in numerous areas required for *Daubert* admissibility, some of which we now list. First, the obviously strident ongoing controversy is evidence of the lack of general acceptance by the relevant scientific community. Second, the numerous examples of critical missing data are evidence of unavailability of key evidence to the relevant scientific community, which is a prerequisite to general acceptance by that community. Third, the questionable or inadequate reliability and construct validity of many important RCS variables is evidence of the lack of adequate, well-characterized error rates associated with RCS-based interpretations. Finally, the now-well-documented propensity of RCS interpretations to overpathologize even normal adults and children is evidence of excessive error rates and points also to an unacceptable asymmetry of the false-positive and false-negative errors. Hence, no responsible jurist should permit expert testimony involving the interpretation of Rorschach inkblots using the RCS. Of course, our central arguments apply a fortiori to Rorschach scoring systems other than the RCS, which rest on an even shakier scientific foundation.

One could even propound the thesis that attorneys, when faced with an expert witness offering Rorschach interpretations detrimental to a client, have an obligation to file motions to exclude such questionable evidence under *Frye* and *Daubert* decisions as well as relevance arguments. A competent attorney, armed with the information provided here, should be able to file successful motions *in limine* to exclude Rorschach-based testimony or, if necessary, to cross-examine Rorschach "experts" vigorously.

One last important legal point was alluded to by Ritzler et al. (2002). Legal professionals should not assume that Ritzler et al.'s portrayal of careful and reliable RCS scoring, methodical reference of these scores to representative norms, and basing interpretations only on well-validated scores is an accurate description of the typical courtroom behavior of expert witnesses who rely on the Rorschach. R. Christopher Barden has observed that experts claiming to use the

RCS sometimes use a wholly idiosyncratic amalgam of RCS scores (not scored according to the manual and essentially never explicitly referred to norms), illusory (or at any rate unvalidated) correlations, and unquantified intuition. Still other clinicians either do not use the RCS or rely almost exclusively on subjective interpretations of inkblot responses. These are admittedly anecdotal observations and, hence, are only suggestive, but they are buttressed by findings indicating that many clinicians who use the Rorschach either do not regularly use the RCS or have not received formal training in this method. For example, a substantial proportion, and perhaps even a majority of psychologists in juvenile court clinics who administer the Rorschach, do not use the RCS (Pinkerman, Haynes, & Keiser, 1993), and approximately one third of graduate courses on the Rorschach do not include training in the RCS (Hilsenroth & Handler, 1995).

References

Acklin, M. W., McDowell, C. J., Verschell, M. S., & Chan, D. (2000). Interobserver agreement, intraobserver reliability, and the Rorschach Comprehensive System. *Journal of Personality Assessment, 74,* 15–47.

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist, 47,* 1597–1611.

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

Archer, R. P., & Krishnamurthy, R. (1997). MMPI–A and Rorschach indices related to depression and conduct disorder: An evaluation of the incremental validity hypothesis. *Journal of Personality Assessment, 69,* 517–533.

Aronow, E. (2001). CS norms, psychometrics, and possibilities for the Rorschach technique. *Clinical Psychology: Science and Practice, 8,* 383–385.

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

Barden, R. C. (2001a). Informed consent in psychotherapy: A multidisciplinary perspective. *The Journal of the American Academy of Psychiatry and the Law, 29,* 160–166.

Barden, R. C. (2001b, May 11). *Science intensive litigation: Protecting the integrity of the legal system from pseudoscientific "expert" testimony*. Invited address to the national meeting of the American Bar Association Litigation Section, Phoenix, AZ.

Bornstein, R. F. (1999). Criterion validity of objective and projective dependency tests: A meta-analytic assessment of behavioral prediction. *Psychological Assessment, 11,* 48–57.

Burns, B., & Viglione, D. J. (1996). The Rorschach Human Experience variable, interpersonal relatedness, and object representation in nonpatients. *Psychological Assessment, 8,* 92–99.

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

Caine, S. L., Frueh, B. C., & Kinder, B. N. (1995). Rorschach susceptibility to malingered depressive disorders in adult females. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 10, pp. 165–174). Hillsdale, NJ: Erlbaum.

Carlson, C. F., Kula, M. L., & St. Laurent, C. M. (1997). Rorschach revised DEPI and CDI with inpatient major depressives and borderline personality disorder with major depression: Validity issues. *Journal of Clinical Psychology, 53,* 51–58.

Carter, C. L., & Dacey, C. M. (1996). Validity of the Beck Depression Inventory, MMPI, and Rorschach in assessing adolescent depression. *Journal of Adolescence, 19*, 223–231.

Conners, K. (1989). *Manual for Conners' rating scales*. North Tonawanda, NY: Multi-Health Systems.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U. S., 113 S. Ct. 2786 (1993).

Exner, J. E. (1974). *The Rorschach: A comprehensive system. Vol. 1: Basic foundations*. New York: Wiley.

Exner, J. E. (1978). *The Rorschach: A comprehensive system. Vol. 2: Interpretation*. New York: Wiley.

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

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

Exner, J. E. (1993). *The Rorschach: A comprehensive system. Vol. 1: Basic foundations* (3rd ed.). New York: Wiley.

Exner, J. E. (1995). Comment on "Narcissism in the Comprehensive System for the Rorschach." *Clinical Psychology: Science and Practice, 2*, 200–206.

Exner, J. E. (2001a). A comment on "The misperception of psychopathology: Problems with the norms of the Comprehensive System for the Rorschach." *Clinical Psychology: Science and Practice, 8*, 386–388.

Exner, J. E. (2001b). *A Rorschach workbook for the Comprehensive System* (5th ed.). Asheville, NC: Rorschach Workshops.

Exner, J. E., & Weiner, I. B. (1982). *The Rorschach: A comprehensive system. Vol. 3: Assessment of children and adolescents*. New York: Wiley.

Exner, J. E., & Weiner, I. B. (1995). *The Rorschach: A comprehensive system. Vol. 3: Assessment of children and adolescents* (2nd ed.). New York: Wiley.

Frye v. United States, 293 F. 1013, D. C. Cir. (1923).

Gacono, C. B., & Meloy, J. R. (1994). *The Rorschach assessment of aggressive and psychopathic personalities*. Hillsdale, NJ: Erlbaum.

Garb, H. N. (1984). The incremental validity of information used in personality assessment. *Clinical Psychology Review, 4*, 641–655.

Garb, H. N. (1985). Linear and configural statistical models in the diagnosis of mental disorders. (Doctoral dissertation, University of Illinois at Chicago, 1985). *Dissertation Abstracts International, 45* 7B, 2307.

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.

General Electric Co. v. Joiner, 118 S. Ct. 512 (1997).

Gittelman, R. (1980). The role of psychological tests for differential diagnosis in child psychiatry. *Journal of the American Academy of Child Psychiatry, 19*, 413–434.

Goode, E. (2001, February 20). What's in an inkblot? Some say, not much. *The New York Times*, pp. D1, D4.

Gronnerod, C. (1999). Rorschach interrater agreement estimates: An empirical evaluation. *Scandinavian Journal of Psychology, 40*, 115–120.

Groth-Marnat, G. (1997). *Handbook of psychological assessment* (3rd ed.). New York: Wiley.

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.

Hagen, M. (1997). *Whores of the court: The fraud of psychiatric testimony and the rape of American justice*. New York: HarperCollins.

Hamel, M., Shaffer, T. W., & Erdberg, P. (2000). A study of nonpatient preadolescent Rorschach protocols. *Journal of Personality Assessment, 75*, 280–294.

Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior, 16*, 257–272.

Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T. R., & Brunell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment, 11*, 278–296.

Hilsenroth, M. J., & Handler, L. (1995). A survey of graduate students' experiences, interests, and attitudes about learning the Rorschach. *Journal of Personality Assessment, 64*, 243–257.

Hunsley, J., & Di Giulio, G. (2001). Norms, norming, and clinical assessment. *Clinical Psychology: Science and Practice, 8*, 378–382.

Ilonen, T., Taiminen, T., Karlsson, H., Lauerma, H., Leinonen, K., Wallenius, E.,et al. (1999). Diagnostic efficiency of the Rorschach schizophrenia and depression indices in identifying first-episode schizophrenia and severe depression. *Psychiatry Research, 87*, 183–192.

Jansak, D. M. (1997). The Rorschach Comprehensive System Depression Index, depression heterogeneity, and the role of self-schema (Doctoral dissertation, California School of Professional Psychology, San Diego, 1996). *Dissertation Abstracts International, 57*, 6576B.

Jorgensen, K., Andersen, T. J., & Dam, H. (2000). The diagnostic efficiency of the Rorschach Depression Index and the Schizophrenia Index: A review. *Assessment, 7*, 259–280.

Kadle, J. C. (1989). The suitability of standard Rorschach scoring categories for nonpatient and clinically depressed elderly (Doctoral dissertation, Case Western Reserve University, 1989). *Dissertation Abstracts International, 50*, 2154B.

Kaplan, R. M., & Saccuzzo, D. P. (2001). *Psychological testing: Principles, applications, and issues* (5th ed.). Belmont, CA: Wadsworth.

Kumho Tire Co., Ltd. v. Carmichael, 119 S. Ct. 1167 (1999).

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1*, 27–66.

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2001, May). What's wrong with this picture? *Scientific American, 284*, 80–87.

Lipkin, L. R. (1989). Rorschach performance as a test of a theory of late-onset depression in women (Doctoral dissertation, Temple University, 1988). *Dissertation Abstracts International , 49*, 2864B.

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

Meyer, G. J. (1993). The impact of response frequency on the Rorschach constellation indices and on their validity with diagnostic and MMPI–2 criteria. *Journal of Personality Assessment, 60*, 153–180.

Meyer, G. J. (1997a). Assessing reliability: Critical corrections for a critical examination of the Rorschach Comprehensive System. *Psychological Assessment, 9*, 480–489.

Meyer, G. J. (1997b). Thinking clearly about reliability: More critical corrections regarding the Rorschach Comprehensive System. *Psychological Assessment, 9*, 495–498.

Meyer, G. J. (1999). The convergent validity of MMPI and Rorschach scales: An

extension using profile scores to define response and character styles on both methods and a reexamination of simple Rorschach response frequency. *Journal of Personality Assessment, 72*, 1–35.

Meyer, G. J. (2000). On the science of Rorschach research. *Journal of Personality Assessment, 75*, 46–81.

Meyer, G. J. (2001). Evidence to correct misperceptions about Rorschach norms. *Clinical Psychology: Science and Practice, 8*, 389–397.

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.

Meyer, G. J., & Handler, L. (2000). Incremental validity of the Rorschach Prognostic Rating scale over the MMPI Ego Strength scale and IQ. *Journal of Personality Assessment, 74*, 356–370.

Meyer, G. J., & Richardson, C. (2001, August). *An examination of changes in Form Quality codes in the Rorschach Comprehensive System from 1974 to 1995*. Paper presented at the Midwinter Meeting of the Society for Personality Assessment, Philadelphia.

Nakata, L. M. (2000). Interrater reliability and the Comprehensive System for the Rorschach: Clinical and non-clinical protocols (Doctoral dissertation, Pacific Graduate School of Psychology, 1999). *Dissertation Abstracts International, 60* (8B), 4296.

Nunnally, J. C., & Bernstein, I. C. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

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.

Pinkerman, J. E., Haynes, J. P., & Keiser, T. (1993). Characteristics of psychological practice in juvenile court clinics. *American Journal of Forensic Psychology, 11*, 3–12.

Psychological Corporation (1997). *Wechsler Adult Intelligence Scale, Third Edition. Wechsler Memory Scale, Third Edition, technical manual*. San Antonio, TX: Author.

Ritzler, B. (1995). Putting your eggs in the content analysis basket: A response to Aronow, Reznikov, and Moreland. *Journal of Personality Assessment, 64*, 229–234.

Ritzler, B., Erard, R., & Pettigrew, G. (2002). Protecting the integrity of Rorschach Expert Witnesses: A Reply to Grove and Barden (1999) Re: The Admissibility of Testimony Under *Daubert/Kumho* Analyses. *Psychology, Public Policy, and Law, 8*, 201–215.

Rosenthal, R. (1978). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638–641.

Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement, 12*, 153–158.

Sells, J. E. (1991). A validity study of the DEPI index: The Rorschach Comprehensive System (Doctoral dissertation, University of Utah, 1990). *Dissertation Abstracts International, 51*, 5590B.

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.

Silberg, J. L., & Armstrong, J. G. (1992). The Rorschach test for predicting suicide among depressed adolescent inpatients. *Journal of Personality Assessment, 59*, 290–303.

Singer, H. K., & Brabender, V. (1993). The use of the Rorschach to differentiate unipolar and bipolar disorders. *Journal of Personality Assessment, 60*, 333–345.

Viglione, D. J., Brager, R. C., & Haller, N. (1988). Usefulness of structural Rorschach data in identifying inpatients with depressive symptoms: A preliminary study. *Journal of Personality Assessment, 52*, 524–529.

Wechsler, D. (1997). *WAIS–III administration and scoring manual*. San Antonio, TX: Psychological Corporation.

Weiner, I. B. (1999). Contemporary perspectives on Rorschach assessment. *European Journal of Psychological Assessment, 15*, 78–86.

Whitehead, W. C. (1985). *Clinical decision making on the basis of Rorschach, MMPI, and automated MMPI report data*. Unpublished doctoral dissertation, University of Texas Health Science Center at Dallas.

Widiger, T. A. (2001). The best and the worst of us? *Clinical Psychology: Science and Practice, 8*, 374–377.

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. (2001a). 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, M. T., Garb, H. N., & Lilienfeld, S. O. (2001b). Problems with the norms of the Comprehensive System for the Rorschach: Methodological and conceptual considerations. *Clinical Psychology: Science and Practice, 8*, 397–402.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996a). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science, 7*, 3–10.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996b). Thinking critically about the Comprehensive System for the Rorschach: A reply to Exner. *Psychological Science, 7*, 14–17.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1997). The reliability of the Comprehensive System for the Rorschach: A comment on Meyer (1997). *Psychological Assessment, 9*, 490–494.

Wood, J. M., Nezworski, M. T., Stejskal, W. J., & Garven, S. (2001). Advancing scientific discourse in the controversy surrounding the Comprehensive System for the Rorschach: A rejoinder to Meyer (2000). *Journal of Personality Assessment, 76*, 369–378.

Wood, J. M., Nezworski, M. T., Stejskal, W. J., Garven, S., & West, S. G. (1999). Methodological issues in evaluating Rorschach validity: A comment on Burns and Viglione (1996), Weiner, (1996), and Ganellen (1996). *Assessment, 6*, 115–129.