COMMENTARY

Problems of the Comprehensive System for the Rorschach in Forensic Settings: Recent Developments

James M. Wood, PhD
M. Teresa Nezworski, PhD
William J. Stejskal, PhD
R. K. McKinzey, PhD

James M. Wood is Clinical Psychologist and Associate Professor of psychology at the University of Texas at El Paso. His recent research focuses on the interrogative suggestibility of children and adults. He and his colleagues have published several critical reviews of the Comprehensive System for the Rorschach. M. Teresa Nezworski is Clinical Child Psychologist and Associate Professor at the University of Texas at Dallas. She has an active clinical practice in the area of behavioral medicine and trains clinical psychology interns in methods of assessment. Dr. William Stejskal has conducted psychological and neuropsychological evaluations in clinical, correctional, and human service settings since 1984. He has provided forensic evaluation and consultation services in criminal and civil proceedings in Federal and State courts in the Mid-Atlantic region. He is now at the Institute of Law, Psychiatry, and Public Policy at the University of Virginia. R. K. McKinzey is in private forensic practice on Oakland. He is a member of APA’s Div. 41 and NAN. He has an article on forensic neuropsychology at <www.cdaa.org/pubs/mckinzey.pdf>.

Address correspondence to: James M. Wood, PhD, Department of Psychology, University of Texas at El Paso, El Paso, TX 79968 (E-mail: jawood@miners.utep.edu).

© 2001 by The Haworth Press, Inc. All rights reserved. 89
ABSTRACT. The Comprehensive System for the Rorschach is currently the subject of heated controversy among psychologists. Much "common knowledge" about the test is either incorrect or in dispute. Psychologists who use the Rorschach in forensic settings can often be successfully challenged by well-informed attorneys and may risk becoming the subject of ethics complaints. This article identifies seven issues that are particularly relevant to use of the Comprehensive System for the Rorschach in forensic psychology.

KEYWORDS. Rorschach, comprehensive system, projective testing, forensic, expert witnesses

The Rorschach Inkblot test is one of the assessment instruments most commonly used by psychologists in clinical and forensic settings. Over the past five years, it has also become one of the most controversial. Recent research and critical reviews have shown that many widely promoted claims regarding the capabilities of the test are probably untrue. As an example, we begin by discussing the well-known Rorschach “reflection response.”

As is known to anyone familiar with the Comprehensive System for the Rorschach (CS; Exner, 1991, 1993), a subject’s response is scored as a reflection if it describes a mirror image or reflection ("trees reflected in a lake"). According to Exner (1991, p. 149), reflection responses “are not expected to appear in the records of older adolescents or adults.” Thus, the presence of even one reflection in a Rorschach protocol indicates that “a nuclear element in the subject’s self-image is a narcissistic-like feature that includes a marked tendency to overvalue personal worth” (Exner, 1991, p. 173). Among criminal offenders, reflection responses distinguish psychopaths from non-psychopaths, and “are consistent with pathological narcissism and omnipotence noted in Antisocial Personality Disorder” (Gacono & Meloy, 1992, p. 401; see also Gacono & Meloy, 1994).

All the foregoing information is widely known. Surprisingly, however, all of it is also either incorrect or in dispute. First, Rorschach reflection responses are actually quite common among older adolescents and adults. It is true that, according to the CS norms, only 7% of
nonpatient adult protocols contain a reflection response (Exner, 1993). However, in a recent normative study of 123 non-patient adults by Shaffer, Erdberg, and Haroian (1999), 29% of protocols contained a reflection. Wood, Nezworski, Garb, & Lilienfeld (2000) arrived at the same figure of 29% in a review of 8 additional non-patient studies with a total of 368 subjects. In the first published study of reflection responses, Exner himself (1969; see also Exner, 1993, p. 433) found reflection responses in 35% of protocols from a sample of nonpatient adults drawn mainly from a college population.

Furthermore, contrary to common wisdom, the relationship between reflection responses and narcissism has never been established (see review by Nezworski & Wood, 1995). For example, empirical studies that have examined the relationship between Narcissistic Personality Disorder (NPD) and reflection responses have had serious methodological flaws and yielded ambiguous results (see review by Wood, Lilienfeld, Garb, & Nezworski, 2000a). In addition, reflections are generally unrelated to questionnaire measures of narcissism (Himelstein, 1983/1984; Jacques, 1990/1991; but see Hilsenroth, Fowler, Padawer, & Handler, 1997).

Finally, despite claims that the Rorschach is “ideally suited” for assessment of psychopathy (Meloy & Gacono, 1995, p. 414), reflection responses appear to bear little or no relationship to psychopathy or antisocial personality disorder. For example, eleven studies have examined the relationship of reflection response to scores on the Psychopathy Check List (Forth, Hart, & Hare, 1990; Forth, Kosson, & Hare, 1997; Hare, 1980, 1991). However, only two of these studies, with a total of 90 subjects, found a significant positive relationship between reflection responses and psychopathy (Gacono, Meloy, & Heaven, 1990; Loving, 1998). By contrast, nine of the studies, with a total of 592 subjects, did not find such a relationship (Darcangelo, 1996/1997; Egozi-Profeta, 1998/1999; Gacono, Meloy, & Berg, 1992; Muntz, 1998/1999; Murphy-Peaslee, 1993/1995; Ponder, 1998/1999; Siemsen, 1999; Smith, 1994/1995; Welsh, 1999; see also Smith, Gacono, & Kaufman, 1997, 1998).

Over the past 5 years, much “common knowledge” regarding the CS has been revealed as faulty or ill-founded. For example, until recently the scientific community accepted reports that the scoring reliability of CS variables was uniformly above a minimum acceptable threshold of \( r = .85 \) (Exner, 1993; Groth-Marnat, 1997). However, Acklin, McDowell, Verschell, & Chan (2000; see also Shaffer et al., 1999) have reported that only about half of CS variables meet this minimum standard. In fact,
Some CS scores appear to have scoring reliability below .30 (Acklin et al., 2000).

Other common claims regarding the CS have also been shown to be dubious. For example, it is now generally recognized that scores on the CS Depression Index bear little relationship to diagnoses of depression (for reviews, see Jorgensen, Andersen, & Dam, in press; Viglione, 1999; Wood, Lilienfeld et al., 2000a), and that the CS Suicide Constellation may be unrelated to suicidality (for reviews, see Garb, Wood, Nezworski, Grove, & Stejskal, in press; Wood, Nezworski, & Stejskal, 1996a; but see Viglione, 1999). Indeed, contrary to optimistic past claims, the Rorschach seems to bear little relationship to psychiatric diagnoses (Wood, Lilienfeld et al., 2000a, 2000b; but see Garfield, 2000; Kubiszyn et al., 2000; Lerner, 2000; Weiner, 2000). The problems with using the Rorschach as a diagnostic tool have been recognized even by well-known Rorschach advocates. For example, Irving Weiner (1999, pp. 336-337) has recently declared that the Rorschach “is not a diagnostic test, it was not designed as a diagnostic test, it is not intended to be a diagnostic test, and it does not in fact work very well as a diagnostic test, especially if what is meant by diagnosis is a DSM category.”

In the present article, we cannot hope to review the entire scientific debate that has recently engulfed the CS, although we strongly urge readers to review the relevant literature. Instead, we will focus on seven issues regarding the Rorschach that have potential implications for the practice of psychology in forensic settings.

**THE RORSCHACH IS HIGHLY CONTROVERSIAL**

Perhaps the most important new development concerning the Rorschach is also the most obvious: After lying dormant for over two decades, the long-standing controversy among psychologists regarding the test has erupted again with unexpected force. For example, in 1999 and 2000, three journals (Psychological Assessment, Journal of Clinical Psychology, Assessment) published debates between Rorschach critics and proponents. One respected scholar has called for a moratorium on use of the Rorschach in clinical and forensic settings (Garb, 1999), and another article (by two of the present authors and their colleagues) has declared, “it seems particularly important that the Rorschach not be used to diagnose individuals in forensic contexts (Wood, Lilienfeld et al., 2000a, p. 417). As Robert Archer (1999, p. 309), editor of Assessment, comments, “the assumption that the Rorschach Comprehensive System rests solidly
and uniformly on an empirical foundation has been forced to undergo a significant re-examination.” The current Rorschach controversy is not limited to trivial points but touches on fundamental issues regarding the test’s scientific standing. Critics have pointed out that the unpublished reliability and validity studies of Exner (1991, 1993) that provide the empirical basis of the CS are often unavailable for scrutiny by independent scholars, and that many CS scores lack well-demonstrated validity (Garb et al., in press; Nezworski & Wood, 1995; Wood, Nezworski, & Stejskal, 1996a, 1996b; but see Exner, 1995, 1996). Additional debates have flared between Rorschach advocates and critics regarding such fundamental issues as scoring reliability, test-retest reliability, incremental validity, clinical utility, effects of method variance, cultural sensitivity, and research methodology (Acklin et al., 2000; Aronow, 1999; Aronow, Reznikoff, & Moreland, 1994, 1995; Costello, 1999; Dawes, 1994; Ganellen, 1996a, 1996b; Gann, 1995; Garb, 1998, 1999; Garb et al., in press; Garb, Wood, & Nezworski, 2000; Garfield, 2000; Hunsley & Bailey, 1999, in press; Meyer, 1997a, 1997b; Sechrest & McKnight, in press; Sechrest, Stickle, & Stewart, 1998; Stricker & Gold, 1999; Viglione, 1999; Weiner, 1996, 1999; Wood & Lilienfeld, 1999; Wood, Lilienfeld et al., 2000a, 2000b; Wood, Nezworski, Garb, & Lilienfeld, 2000; Wood et al., 1996a, 1996b, 1997; Wood, Nezworski, Stejskal, Garven, & West, 1999).

The forensic implications of such fierce scientific controversy seem clear: Even if the Rorschach is admissible into the courtroom under current legal rules (an issue that we discuss further below), expert witnesses’ credibility may be weakened if they are shown to have relied on a controversial technique such as the Rorschach. No longer can psychologists take the Rorschach into court and claim honestly that the test is widely accepted by the scientific community. In fact, we believe that psychologists who use the test, particularly in forensic settings, are under an ethical obligation to forthrightly describe the limitations of the test and the controversy that surrounds it (American Psychological Association, 1992, Standards 2.08a, 7.04b).

MANY PSYCHOLOGISTS ROUTINELY USE THE RORSCHACH FOR PURPOSES THAT ARE INVALID OR POORLY SUPPORTED

Four years ago, Wood et al. (1996b) challenged Rorschach proponents to identify CS scores that have been well validated according to
three simple criteria: (1) The score has consistently been found valid for a particular purpose in several studies, (2) the studies were methodologically sound, and, (3) the studies were carried out by independent groups of researchers. To date, only a handful of Rorschach variables have been shown to meet even these minimal criteria (Garb et al., in press). First, some Rorschach measures of deviant verbalizations or poor form (e.g., the CS Schizophrenia Index) have shown a consistent relationship to diagnoses of schizophrenia, psychosis, or Borderline Personality Disorder (Wood, Lilienfeld et al, 2000a). Second, some Rorschach measures (e.g., the total number of responses) may show low or moderate correlations with intelligence. Third, the Rorschach Oral Dependency scale (ROD; Bornstein, 1996, 1999) appears to have a valid relationship to dependent behavior (but see Garb et al., in press). Finally, the Rorschach Prognostic Rating Scale (RPRS) appears to be related to psychotherapy treatment outcome (Meyer & Handler, 1997; but see Garb et al., in press; Hunsley & Bailey, 1999, in press). However, neither the ROD nor the RPRS is part of the CS, and neither scale has current norms that would allow its use in forensic settings.

Any psychologist who ventures to rely on Rorschach scales other than the ones just listed is almost certain to be treading on shaky ground. But of course, virtually all psychologists who use the test also use its invalid or poorly validated scales. For example, the Rorschach is regularly held forth as a measure of post-traumatic stress disorder or sexual abuse victimization, even though the test is not well-validated for these purposes (Garb et al., 2000; Wood, Lilienfeld et al., 2000a; for an example of poorly validated Rorschach scores used in a forensic context, see Weiner, 1999).

In common practice, psychologists who use the Rorschach interpret routinely scores for which good validity data are lacking. This practice is particularly problematic in forensic settings because it can now easily be challenged by opposing attorneys. Such challenges have been very rare in the past (Weiner, Exner, & Sciara, 1996), probably for three reasons. First, until recently, few attorneys or their experts have realized how vulnerable the Rorschach is to serious challenge. Second, attorneys usually lack the necessary psychological knowledge themselves to effectively challenge an expert witness regarding the validity of the Rorschach. Third, in legal cases where the Rorschach is most likely to be used (e.g., custody cases), considerations of time or money may prevent the parties from mounting such a challenge or hiring appropriate experts.
In the future, however, challenges to forensic psychologists who use the Rorschach are likely to become more common, as attorneys and the psychologists who assist them become more widely informed regarding the weak validity of most Rorschach scores.

**RORSCHACH SCORING IS SUBSTANTIALLY LESS RELIABLE THAN PREVIOUSLY ASSUMED**

As already noted, the scoring reliability of the CS is much lower than has long been assumed (Acklin et al., 2000; Gronnerod, 1999; Shaffer et al., 1999), and some important CS variables exhibit a level of reliability that is highly problematic. For example, the interrater reliabilities of the Schizophrenia Index in two samples studied by Acklin et al. (2000) were .452 and .560. Similar weak reliability was exhibited for Adjusted $D$ (.533 and .678), $X-%$ (.621 and .656), and $FC:CF + C$ (.543 and .165).

These unfavorable numbers indicate that scoring accuracy can be a major concern when the Rorschach is used in forensic contexts. Even if a particular psychologist is highly experienced with the Rorschach or regarded as an authority, his or her Rorschach scoring is not necessarily above challenge. Because many Rorschach scores have moderate or even low reliability, two experts with the highest qualifications may score a protocol much differently if they work independently. For this reason, in legal cases involving the Rorschach it is often advisable to have the protocol re-scored by a second expert who does not know the first expert’s scores. In our experience, such a re-scoring procedure often reveals errors in the original scoring or important discrepancies between the two sets of scores, which may later be used to challenge an expert witness who has scored and interpreted the test.

**THE NORMS OF THE COMPREHENSIVE SYSTEM ARE PROBABLY IN ERROR**

The extensive norms of the CS for both children and adults are often held forth as a great scientific and clinical achievement (e.g., Weiner, 1998). Yet evidence has gradually accumulated that the norms for important CS variables are seriously in error. For example, in a study of 123 nonpatient adults in California, Shaffer et al. (2000) found that the means and standard deviations of CS variables often differed substan-
tially from CS norms. In a follow-up study, Wood, Nezworski et al. (2000) examined 14 CS variables in 32 additional studies of non-patient adults. The participants in these studies exhibited statistically and clinically significant differences from the CS nonpatient norms for all 14 variables. Overall, the discrepancies had the effect of making non-patients appear “pathological” in comparison with the CS norms. Wood, Nezworski et al. recommended that psychologists refrain from using the CS norms in clinical or forensic contexts with either children or adults, and cited relevant ethical principles that bear on the use of inappropriate norms (American Psychological Association, 1992, Standards 1.14, 2.02a, 2.07b, 2.08a, 7.04b; see also American Psychological Association, 1999). In forensic contexts, any psychologist who uses the current CS norms to guide interpretations, particularly for Rorschach indices lacking good validity, may be exposed to strong legal challenges, or risk becoming the subject of an ethics complaint.

**USE OF THE CS WITH AMERICAN MINORITY GROUPS OR NON-AMERICANS IS PROBLEMATIC**

Although Rorschach proponents often suggest that the Rorschach is well-suited for use with minorities or non-Americans (Butcher, Nezami, & Exner, 1998; Ritzler, 1996; Viglione, 1999), there is substantial evidence to the contrary. Studies have reported that Blacks, Hispanics, Native Americans, and non-American groups often score differently on important variables for both the CS and other Rorschach approaches. Furthermore, there have been a substantial number of critiques regarding cross-cultural use of the Rorschach and particularly the lack of appropriate normative data (for summaries of this literature, see Garb et al., in press; Wood & Lilienfeld, 1999). For example, Krall et al. (1983) found that inner-city black children differed from then-current CS norms on 5 (50%) of 10 Rorschach variables. Glass, Bieber, & Tkachuk (1996) found that incarcerated Alaskan Native Americans differed from the CS norms for two-thirds of Rorschach scores. Boscán (1999/2000) found that Rorschach scores of 101 Mexican college students differed significantly in many respects from the CS norms. In addition, Boscán discussed several studies of Central and South American groups that had reported similar results. In light of such findings, the use of the Rorschach with American minorities or non-Americans is open to serious challenge in forensic contexts.
“AUTHORITATIVE” RORSCHACH BOOKS ARE OFTEN UNBALANCED OR OUT-OF-DATE

It is not uncommon for expert witnesses to cite books by Exner (1991, 1993) or Weiner (1998) as highly reliable and authoritative sources on the CS. However, the literature reviews in these books are often out of date or unbalanced, and their conclusions are often inconsistent with the scientific evidence (Costello, 1999; Nezworski & Wood, 1995; Wood et al., 1996a; Wood, Lilienfeld et al., 2000a). For example, these books make a number of ill-founded claims of the type already discussed in this article (e.g., scoring reliability for Rorschach scores is excellent; the CS norms are accurate; reflection responses are related to anti-social behavior or psychopathy; the Egocentricity Index is related to self-concern or self-esteem; the Depression Index is related to depression). Furthermore, these books often fail to mention negative research findings regarding the CS, or the controversy that currently surrounds the test. In addition, Exner’s (1991, 1993) books on the Rorschach are particularly problematic because they rely heavily on the unpublished studies of his Rorschach Workshops, which are generally unavailable for scrutiny by other scholars (Wood et al., 1996a; Wood & Lilienfeld, 1999; Garb et al., in press). In a forensic context, any expert who uses these books to support opinions in court is potentially vulnerable to challenge by an attorney familiar with the broader Rorschach literature.

THE RORSCHACH AND THE DAUBERT CRITERIA

At present, scholars disagree whether the Rorschach meets the U.S. Supreme Court’s “Daubert criteria” for admissibility of scientific evidence in court (Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993). Although McCann (1998) has argued that the CS (but perhaps not other Rorschach systems) meets the Daubert criteria, Grove and Barden (1999) have reached the opposite conclusion. In part, such disagreements may simply represent a time lag. McCann’s article was published in 1998, and appears to have been written when the current Rorschach controversy was still taking shape. For example, McCann’s analysis seemed to rest on the then-common assumptions that (a) the CS norms are psychometrically sound, (b) the scoring reliability of most CS scores is very good, (c) many CS scores have shown a well-demonstrated relationship to psychiatric diagnoses and psychopathy, and
the Rorschach has general acceptance in the psychological community. However, in light of recent developments in the Rorschach controversy, such assumptions appear questionable. New legal analyses, based on more recent information, may lead to different conclusions (e.g., Grove & Barden, 1999).

In closing, we suggest that the fate of the Rorschach in forensic contexts may hinge on factors other than its legal admissibility. No matter what the Daubert criteria seem to say, many judges will probably continue to admit dubious psychological assessment techniques such as the Rorschach into court (McKinsey & Ziegler, 1999). However, even if the Rorschach is admitted into court, it may well prove a liability to the side that uses it. As we have indicated, the Rorschach is subject to challenge on numerous points. The expert who uses it may be subjected to considerable embarrassment if cross-examined by a well-informed attorney. In addition, the risk of ethics complaints, based on inappropriate use of the test, will probably grow greater with each passing year. In our opinion, factors such as these may lead psychologists to re-evaluate whether their use of the Rorschach is consistent with widely accepted standards for assessment procedures in forensic settings (Heilbrun, 1992). Because the test seems to be valid for only a few very narrow purposes, most psychologists may eventually abandon its use in forensic settings.

REFERENCES


RECEIVED: 10/06/00
REVISED: 10/20/00
ACCEPTED: 10/30/00