CHAPTER 3

Admissibility of the Rorschach

Joseph T. McCann
UHS Hospital, Binghamton, NY

F. Barton Evans
Private Practice, Bozeman, MT

Defending the Rorschach in Court: An Analysis of Admissibility Using Legal and Professional Standards

Joseph T. McCann

There are a number of legal and professional standards that guide admissibility of expert testimony and psychometric evidence. The most important legal tests for admissibility include the Federal Rules of Evidence (1992), the Frye test (United States v. Frye, 1923), and the Daubert standard (Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993). Within the profession, Heilbrun (1992) and Marlowe (1995) have outlined criteria for the selection of psychological tests in forensic settings. Using these legal and professional standards, the Rorschach is analyzed according to individual criteria. Although several issues require additional research, it is concluded that the status of the Rorschach Inkblot Method is such that it satisfies legal tests of admissibility and professional criteria that have been suggested. In those cases where the Rorschach is apt to be deemed inadmissible, it is likely due to how data from the instrument are utilized, rather than characteristics of the instrument itself.

Professional sentiments about use of the Rorschach Inkblot Method in forensic settings have ranged widely. There are those who suggest that the Rorschach is an instrument that has absolutely no validity and should not be used in any clinical or forensic settings. The strongest proponent of this position is Dawes (1994), who contended that the Rorschach is a “shoddy” instrument and “is not a valid test of anything” (p. vii). He went so far as to recommend to everyone that if they ever undergo a psychological evaluation and someone takes out the Rorschach, that the person should leave the room. There are, of course, others who contend that the Rorschach is a valid and reliable instrument that can yield


F. Barton Evans is the sole author of the update regarding McCann’s article.
very useful information when used properly. The clearest position on this approach has been outlined by Weiner (1996), who stated that those who are critical of the Rorschach have “not read the relevant literature of the last 20 years; or, having read it, they have not grasped its meaning” (p. 206). Given the need for empirically based methods when conducting forensic evaluations, the issues raised in this debate over use of the Rorschach Inkblot Method are of extreme importance to those who use the instrument in settings where their findings may be presented and challenged in court.

The Rorschach continues to hold a prominent place in training and practice settings. In surveys of training programs, teaching of the Rorschach Inkblot Method continues at a very high level, with 94% of programs accredited by the American Psychological Association teaching the instrument in 1974, 93% teaching it in 1984, and 85% teaching it in 1993 (Piotrowski & Zalewski, 1993; Ritzler & Alter, 1986). In addition, clinical use of the Rorschach has not diminished in the last several decades. Surveys of professional test usage document that the Rorschach is among the most widely used psychological assessment instruments in the United States. It has been shown consistently that the Rorschach is used by more than 80% of agencies or practitioners who engage in psychological assessment (Lubin, Larsen, & Matarazzo, 1984; Lubin, Wallis, & Paine, 1971; Piotrowski & Keller, 1989; Sundberg, 1961; Watkins, Campbell, Nieberding, & Hallmark, 1995).

Despite the favorable status that the Rorschach apparently has within the profession, there are still strong critics who state that the instrument lacks validity and reliability and should not be viewed as a measure of anything (Dawes, 1994; Ziskin & Faust, 1988). Moreover, as managed care and other economic pressures have impacted on the professional practice of psychology, mental health practitioners are finding themselves searching for alternative avenues of professional practice. Mental health professionals are engaging in forensic work with greater frequency, and this area of practice is often where comprehensive personality assessment, involving broad-based approaches to testing that include the Rorschach, has been relatively untouched by economic concerns (Acklin, 1996).

Given the increased frequency with which mental health professionals have become involved in evaluating individuals in legal settings, there has been a concomitant increase in commentary in the literature on the appropriate use of various psychological testing instruments in forensic practice. In particular, there are guides for using the MMPI (Pope, Butcher, & Seelen, 1993) and Millon inventories (McCann & Dyer, 1996) in forensic settings. However, there is limited literature on use of the Rorschach in forensic settings.

Some exceptions to this state of affairs are articles that document the psychometric characteristics of the Rorschach (Weiner, 1996, 1997) and the experiences of a large sample of professionals who use the Rorschach in forensic settings; who find that the instrument does not encounter significant challenge in the courtroom (Weiner, Exner, & Sciara, 1996). Meloy, Hansen, and Weiner (1997) also found that the Rorschach has been given legal weight by appellate courts throughout the United States. These articles document useful information on the reliability, validity, and clinical utility of the Rorschach as an instrument for describing personality, assisting in differential diagnosis, and identifying important psychological variables in forensic settings.
The purpose of this article is to reformulate the issues in this debate by analyzing its psychometric properties and its current scientific and clinical status within the context of established standards for admissibility of expert testimony. Legal standards established through case law and formal rules of evidence, and proposed guidelines found in the professional literature, constitute the two major sources from which criteria are drawn to analyze the appropriateness of using the Rorschach in forensic settings. Note that legal rules on admissibility of expert testimony vary across jurisdictions, and in many instances it is within the presiding judge’s discretion whether proffered testimony and the methodology on which it is based are ultimately admitted into court. The following analysis does not constitute legal or professional advice but is offered as a guideline that can be used in establishing the admissibility of one’s reliance on the Rorschach when answering psycholegal questions and performing psychological evaluations in forensic matters.

LEGAL STANDARDS FOR EXPERT TESTIMONY

It is the job of the trier of fact (i.e., judge or jury) in legal settings to decide ultimate issues such as the determination of guilt or innocence, appropriate monetary damages, and other matters critical to the outcome of a case. This process is done by weighing and evaluating evidence offered before a court. Mental health professionals frequently appear as expert witnesses to assist the trier of fact by providing information that is deemed important for deciding the legal issues. Because expert witnesses offer opinions that are based on knowledge that is assumed to be outside the range of familiarity for the typical juror, expert opinions typically serve to make information available to jurors that will assist their decision making.

Legal standards in the form of evidentiary rules serve to guide judges in deciding whether proffered expert testimony will be admissible. Evidentiary rules vary across jurisdictions, with Federal courts relying on the Federal Rules of Evidence (FRE; 1992) and state courts relying either on codified rules of evidence or extensive case law. Some state courts model their rules of evidence after the FRE; other state courts (e.g., New York) have a large body of case law governing the admissibility of evidence. Regardless of the specific jurisdiction in which an expert might testify, one of three major legal standards is likely to guide admissibility of testimony; these standards are the FRE, the Frye test (United States v. Frye, 1923), and the Daubert standard (Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993).

FRE

According to the FRE, evidence is admissible if it is “relevant” (Rule 401); more specifically, the evidence must make the existence of any fact that is of consequence to the outcome of the case more or less probable. However, not all relevant evidence is admissible, and Rules 402 and 403 of the FRE outline exceptions to the admissibility of relevant evidence, such as evidence that is not permitted by the U.S. Constitution (e.g., if a criminal defendant’s testimony violates the Fifth Amendment right against self-incrimination) or evidence that is prejudicial, confusing, or a waste of time.
Article VII of the **FRE** outlines the parameters for opinion and expert testimony. In particular, Rule 702 governs the testimony of experts:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill experience, training, or education, may testify thereto in the form of an opinion or otherwise.

A careful reading of this rule suggests that the standard for admissibility of expert testimony under the **FRE is helpfulness**—testimony must merely be helpful to the trier of fact in making his or her decision.

Under Rule 703 of the **FRE**, the bases of opinion testimony by experts are guided by the following:

The facts or data in the particular case on which an expert bases an opinion or inference may be those perceived by or made known to the expert at or before the hearing. If of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the participant, the facts or data need not be admissible in evidence.

This rule suggests that any method on which an expert relies may be admissible if it is of a type that is reasonably relied on by others in one’s professional community.

With respect to the use of the Rorschach, therefore, the issue under the **FRE** becomes whether reliance on this method is reasonable in light of what others in the professional community rely on and whether opinions derived, in part from Rorschach results, will be helpful to the jury. The **FRE** were designed to be fairly liberal in that they were intended to give judges latitude in deciding on the admissibility of evidence.

**The Frye Standard**

A second major legal standard governing admissibility of expert testimony is the **Frye** test that has been outlined in the original court decision in *United States v. Frye* (1923). Although this case represents a Federal opinion, it has been adopted by many state courts as the appropriate standard for determining admissibility of expert testimony. According to the **Frye** test, expert testimony is admissible if it is based on a methodology that has been sufficiently established to have gained *general acceptance* in the field to which it belongs. Therefore, the **Frye** test examines whether a particular theory, technique, or methodology is generally accepted in the field when determining if expert testimony is admissible.

Again with respect to the Rorschach, an important consideration will be whether the instrument is generally accepted in the field of clinical psychology and forensic assessment. This issue will be examined more carefully later in light of recent findings on use of the Rorschach in both clinical and forensic psychological evaluations.

**The Daubert Standard**

A third major legal standard for examining the admissibility of expert testimony is the U.S. Supreme Court decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993).
Although the case dealt with a civil action brought against the manufacturer of an anti-nausea drug for damages related to resultant birth defects, the *Daubert* decision addressed a critical issue on the admissibility of expert testimony in federal courts. Prior to the decision, federal courts had been essentially divided on whether the FRE or Frye test was the appropriate test for admissibility of proposed expert testimony. In a detailed decision, the *Daubert* court held that the appropriate standard of admissibility in federal courts was the FRE and that the general acceptance test under Frye was no longer appropriate as the sole determinative factor. In essence, the court stated that the FRE “assign to the trial judge the task of ensuring that an expert’s testimony both rests on a reliable foundation and is relevant to the task at hand” (p. 2799).

However, the *Daubert* decision outlined a variety of factors that judges might consider when making a determination as to whether expert testimony is admissible. The four major factors outlined in the decision include the following: (a) The theory or technique is scientific knowledge that can or has been tested, (b) the theory or technique has been subjected to peer review, (c) there is a known or potential rate of error, and (d) general acceptance in the field. These four factors were offered by the *Daubert* court as a guide for judges when deciding what expert testimony should be admitted.

The first of the four criteria suggests that courts give consideration to whether a specific technique or theory has been tested. In particular, it is important to determine the extent to which the methodology is falsifiable and capable of empirical testing. Although mental health professionals are not generally oriented toward the concept of falsifiability, this concept refers to the extent to which hypotheses have been generated about a theory or methodology and tested to determine their validity. The second of the four factors recommends that courts evaluate the extent to which the methodology has been subjected to peer review and publication. This factor is “relevant, but not dispositive” (p. 2797) in evaluating the extent to which flaws in the technique or methodology have been uncovered by the scientific community. The third factor refers not only to “known or potential” error rates, but also to the “existence and maintenance of standards controlling the technique’s operation” (p. 2797). Here the court appears to recommend that courts consider whether standards have been developed for using the methodology and for drawing conclusions from the data. Finally, the fourth criterion outlined in the *Daubert* decision is general acceptance, as it has been outlined earlier in the *Frye* opinion. That is, acceptance of the methodology is important reformation that should be considered when evaluating admissibility, but it is not the sole controlling factor as it is under the *Frye* test.

The *Daubert* opinion has generated a considerable amount of academic discussion and there is no consensus about implications of the opinion. There are some who have argued that the *Daubert* opinion will result in more liberal acceptance of expert testimony because of the FRE’s liberal “helpfulness” approach, whereas others argue that the *Daubert* criteria will result in more critical judicial analysis of proposed testimony and thus greater restrictions on expert opinion (Goodman-Delahunty, 1997). Despite this interesting issue, the major implication of the *Daubert* opinion is that judges continue to have broad discretion in deciding what expert testimony will be permitted in Federal courts.
Another important point to consider is that the Daubert decision does not spell the end of the Frye standard. Some states (e.g. Arizona, California, Florida, Nebraska, & New York) have explicitly rejected the “gatekeeping” role of the court outlined by Daubert and have instead retained the Frye general acceptance test. Therefore, the Frye standard will still apply in those state courts where the test remains in effect, whereas Daubert will control admissibility in all federal courts, or in those state courts that explicitly adopt that particular standard. Thus, regardless of the specific jurisdiction in which an expert testifies, the standard for admissibility will be either the FRE—Daubert or the Frye standard.

PROFESSIONAL STANDARDS FOR FORENSIC PSYCHOLOGICAL USAGE

Although rules of evidence provide the framework within which the ultimate determination of admissibility of expert testimony is made, there have been guidelines offered in the professional literature in recent years to provide direction on the selection of psychological assessment instruments in forensic settings. It is important to recognize that these guidelines do not represent formal standards adopted by any particular professional organization or credential-issuing body. Nevertheless, they provide assistance to forensic mental health professionals who use psychological instruments to help answer psycholegal questions. The two major presentations on this issue are those by Heilbrun (1992) and Marlowe (1995).

Heilbrun’s Guidelines

According to Heilbrun (1992), the selection of a psychological assessment instrument for use in forensic evaluation must be guided by the relevance of the instrument to the legal standard that provides the framework for the forensic issues being evaluated. In particular, the relevance of any psychometric instrument can take one of two forms. The first type of relevancy pertains to those instruments that are direct measures of a legal construct that lie at the heart of a particular forensic assessment. Examples of such instruments include specific interviews and measures for assessing competency to stand trial. (Grisso, 1988), the Gudjonsson Suggestibility Scales as measures of interrogative suggestibility (Gudjonsson, 1997), and the Rogers Criminal Responsibility Assessment Scales (Rogers, 1984). A second form of relevancy described by Heilbrun is when a psychological assessment instrument measures a psychological construct (e.g., reality testing, impulsivity, etc.) that is a critical part of some issue (e.g., mental disease or defect) that must be addressed when answering a psycholegal question. In these situations, the relation between the psychological assessment instrument and forensic issues must be outlined in the professional’s written report or testimony.

The Rorschach is not a direct measure of any legal or forensic issue and instead falls in the second form of relevancy just described. That is, forensic use of the Rorschach typically involves the assessment of some personality variable that is a part of some underlying legal standard, and the clinician must make the connection between Rorschach results and the forensic issue being addressed clear in the report or through testimony. In more direct terms, the Rorschach is a personality assessment instrument and is likely to
have some relevance when any forensic issue can be connected in some way to the participant’s personality.

Although admissibility of expert testimony based in part on results from psychological testing is determined by legal rules of evidence, Heilbrun stated that mental health professionals should also formulate their own standards. Toward this end, he proposed seven guidelines that can serve as a framework for selecting psychological tests in forensic assessments. These seven guidelines fall under the three general categories of test selection, administration, and interpretation. The standards proposed by Heilbrun include the following:

1. The test should be commercially available and adequately documented in a manual and should be peer reviewed.
2. Reliability should be established, with a coefficient of .80 advisable or explicit justification for lower coefficients.
3. The test should be relevant to the legal issue or some psychological construct underlying the legal issue with available validation research.
4. The test should have a standard method of administration.
5. The test should be applicable to the population and purpose for which it is used.
6. Objective tests and actuarial data applications are preferred.
7. There should be some method for interpreting test results within the context of the individual’s response style.

These guidelines serve as a useful outline to assess the suitability of a particular psychological testing instrument for forensic assessment. Although they do not represent adopted standards by any format credential-issuing or professional organization, they do represent a useful outline to follow.

**Marlowe’s Hybrid Model**

In an attempt to operationalize an approach to evaluating admissibility of psychometric evidence, Marlowe (1995) developed a hybrid model (i.e., a blend of scientific and legal principles) for examining psychometric evidence. This model is comprised of a flowchart that formulates seven basic questions that are addressed in sequential order. According to Marlowe, the first question that must be addressed is whether the witness is recognized as an expert possessing scientific, technical, or specialized knowledge that will assist the trier of fact; this question is similar to the standard outlined in FRE Rule 702. The second question addresses whether the theory underlying the data collection is recognized, time tested, falsifiable, and committed to refinement. Once this question has been addressed, the third step in Marlowe’s model is to determine whether the psychometric instrument is recognized. In particular, the concern is that an instrument have adequate reliability, standardized norms, professional standards governing appropriate use, and that constructs measured by the instrument have some relevance to the forensic issue at hand. The fourth step in Marlowe’s hybrid model is that the instrument have appropriate data collection, reduction, and analysis procedures. The specific con-
cerns at this stage of analysis are that the instrument have justified norms, standardized administration procedures, and accepted methods for analyzing and scoring data.

The last three steps in Marlowe’s hybrid model are concerned less with the psychometric properties of the test and more with the manner in which the instrument and the data it yields are utilized. The fifth question addresses whether the data are irrelevant, prejudicial, or duplicative. These issues are more legally based and do not have particular scientific counterparts; they deal mainly with assuring that admissibility of psychometric data will not impede expediency in the judicial process. At the sixth stage in Marlowe’s model is whether the psychometric data should be excluded because they run afoul of prevailing social policies. The concern here is that the data do not contribute to the violation of constitutional or statutory rights, such as discrimination, racial bias, and the like.

The seventh step in Marlowe’s hybrid model is the most complex because it refers to the issue of validity of the expert’s reasoning when relying on particular psychometric data. The issue becomes whether the expert has developed a proper psycholegal formulation into which the data fit, the extent to which there are explicit empirical foundations (e.g. error rates, test validity, etc.) for using the test in a particular manner, and using sound reasoning and an acceptable level of expressed certainty in one’s opinion. In short, the seventh stage in this hybrid model constitutes a large portion of what would be covered during direct testimony and cross-examination.

**ADMISSIBILITY OF THE RORSCHACH: AN ANALYSIS**

A review of each legal and professional standard just presented will quickly reveal that there are many areas of overlap and several common themes run across the various standards. For instance, the Daubert opinion, Heilbrun’s guidelines, and Marlowe’s hybrid model are all concerned with the extent to which a particular psychometric instrument has been subjected to peer review. Likewise, general acceptance is the major consideration under Frye and is also one of many factors that the Daubert standard cites as a factor that judges should consider when deciding on the admissibility of expert testimony. Moreover, the extent to which there is an adequate scientific foundation on which an expert can rest his or her opinion is a major theme in each of the legal tests or professional guidelines presented.

One could go through each of the frameworks presented and analyze admissibility of the Rorschach; however, there would be considerable repetition in such an approach because of overlap that exists across the various approaches. Instead, major factors or issues presented across the various models are presented next and the Rorschach is examined in terms of how well it satisfies those major concerns.

**Publication and Peer Review**

One major standard for evaluating admissibility of expert testimony is whether the methodology has been published and subjected to peer review. This issue is explicitly stated in the Daubert opinion and is also a standard in Heilbrun’s guidelines and Marlowe’s hybrid
model. Therefore, the operational criteria become whether the test is commercially available, has an adequate technical manual, and has research that has been published in peer-reviewed journals.

The Rorschach meets this criterion in several respects. In particular, the Rorschach plates are commercially available from most test publishers and they have been standardized perceptual stimuli for almost 80 years. Moreover, publication of the Comprehensive System method for scoring the Rorschach (Exner, 1974, 1978, 1986, 1991, 1993; Exner & Weiner, 1982, 1994) has become the clinical and forensic standard. Moreover, a considerable amount of methodologically sound research has been generated in the last 20 years on the Rorschach and published in peer-reviewed journals. According to Butcher and Rouse (1996), an average of 95.8 articles are published on the Rorschach each year, and this rate has held fairly constant over the past 20 years.

Some concerns about the Comprehensive System were raised by Wood, Nezworski, and Stejskal (1996a, 1996b) in that a large portion of the research on which Exner’s Comprehensive System is based consists of unpublished data that have not undergone peer review. However, Exner (1996) noted in response that many of these unpublished studies were integrated into several of his publications that have appeared in peer-reviewed journals. Moreover, the frequency of Rorschach-based research articles cited by Butcher and Rouse (1996) indicate that several independent researchers have investigated the instrument and published findings in a peer-reviewed journal. For example, Gacono and Meloy’s (1994) detailed research on antisocial disorders and the Rorschach produced over a dozen scientific studies, which appeared in four different peer-reviewed psychological and psychiatric journals.

None of the legal or professional standards make any specifications as to what constitutes adequate peer review. Moreover, the Daubert court seemed to be concerned more with the fact that the instrument be submitted to “the scrutiny of the scientific community…because it increases the likelihood that substantive flaws in methodology will be detected” (p. 2797). Perhaps no other psychological assessment instrument has been subjected to such intense scrutiny as the Rorschach, and major flaws that were uncovered in its scoring methodology were the impetus behind development of the Comprehensive System.

Standard Administration and Norms

The presence of a standard method for administering, scoring, and interpreting data from a psychometric instrument is one of the main criteria that has been set forth in professional guidelines for forensic use of psychological tests (Heilbrun, 1992; Marlowe, 1995); this criteria has not explicitly appeared in any of the legal tests for admissibility, but there are a number of scoring systems that have been developed for the Rorschach. The most widely used system is Exner’s Comprehensive System (Piotrowski, Sherry, & Keller, 1985). This system depends on a specific method of administration and scoring that is outlined in detail by Exner (1995). Moreover, there are clinical norms available for a variety of diagnostic groups as well as nonpatient adults, adolescents, and children (Exner, 1993, 1995). Based on the extensive normative data available for the Comprehen-
sive System, and the fact that most recent research is based on this particular scoring method, it can be concluded that forensic applications of the Rorschach satisfy the criterion of standardized administration and extensive normative data when the Comprehensive System is utilized (Meloy, 1991; Weiner, 1997). The use of an alternative system of scoring may not satisfy this particular professional guideline.

Reliability

The reliability of a psychometric instrument is explicitly identified by Heilbrun and Marlowe as a vital criterion to consider when selecting an instrument for use in forensic assessments. Moreover, the Daubert standard indirectly points to reliability as an important criterion for admissibility of expert testimony. With respect to Rorschach data, it is important to identify the type of reliability coefficients that can be calculated. McCann and Dyer (1996) outlined reasons why internal consistency is a preferred measure of reliability because it reflects the precision of measurement, whereas test—retest reliability is often affected by changes in the construct being assessed. However, internal consistency is a reliability measure that is well suited for multi-item psychometric scales, but it is not applicable to Rorschach data, which rely on discrete determinants and ratios. Instead, interrater and test—retest reliability are most applicable to Rorschach data.

Exner (1993) reported on the test-retest reliability of various determinants, ratios, and percentages from the Comprehensive System for both adults and children. The magnitude of the correlations for the ratios and percentages generally equal or exceed .80 and show good stability for both children and adults. One exception is es, which is composed of the sum of nonhuman movement and shading determinants. Data provided by Exner (1993) reveal that the test—retest reliability of inanimate movement (m) and diffuse shading (Y) are the lowest among all determinants at .26 and .31, respectively, over a 1-year period. These two determinants are used to compute es and account for the low stability. These variables are measures of situational stress and represent state affect; therefore, stability is expected to be low because of changes in state affect. These low reliability coefficients could instead be viewed as indicative of the validity of these measures.

Nearly all of the stability coefficients cited by Exner (1993) meet or exceed the criterion of .80 cited by Heilbrun, who stated that “the use of tests with a reliability coefficient of less than .80 is not advisable. The use of less reliable tests would require an explicit justification by the psychologist” (p. 265). A survey of Exner’s (1993) data reveals that those determinants that are reflective of trait aspects of personality have reliability coefficients that meet or exceed this criterion, whereas those Rorschach data with coefficients below this level represent state aspects of personality. Therefore, the explicit justification that could be offered for relying on Rorschach measures with lower reliability is that because they measure state versus trait aspects of personality, the lower stability is expected and supports the interpretive significance attributed to these data.

Another issue that is worthy of note concerning Rorschach reliability is the concern raised by Wood and his colleagues (Wood et al., 1996a, 1996b) that interrater reliability
of the Rorschach is unknown because the percentage of agreement figures cited by Exner (1993) are not reliability coefficients, for the reason that they do not take into consideration chance levels of agreement between raters as would a coefficient such as kappa. However, the stability coefficients cited by Exner (1993), which in fact are reliability measures, could not be achieved without adequate interrater reliability. If interrater reliability of the Rorschach were low, then there would be too much error variance and the high levels of test—retest reliability could not be achieved. Therefore, the most critical form of reliability for Rorschach data (i.e., test—retest reliability) offers a reliable measure that satisfies the criterion outlined in professional and legal standards for forensic psychological test use.

Validity

The validity of the Rorschach has been extensively discussed by a number of authors, and this topic is beyond the scope of this article. However, it is important to note that validity of a psychometric instrument is an explicit criterion outlined by Heilbrun and Marlowe. It is also mentioned in the Daubert standard. Despite unresolved issues about what is specifically being measured by a few Rorschach measures (cf. Exner, 1996; Wood et al., 1996a, 1996b), there are several articles documenting the validity of the Rorschach.

Weiner (1996, 1997) outlined how the Rorschach is capable of differentiating state and trait personality variables, measuring developmental change in children and adolescents, monitoring treatment improvement, and identifying distress in individuals with posttraumatic stress disorder. Other reviews have documented validity of the Rorschach for a variety of clinical symptoms and with various populations such as children and adolescents (Atkinson, 1986; Meyer, 1993; Ornberg & Zalewski, 1994; Parker, Hanson, & Hunsley, 1988). The validity of the Rorschach can therefore be supported by citing the research noted in these various reviews.

Rate of Error

The Daubert opinion states that when judges evaluate the admissibility of proffered expert testimony that the known or potential rate of error should be considered. However, the opinion does not explicitly state how rate of error should be defined. With respect to this criterion, the Rorschach data that have a rate of error established are those indexes that are utilized to assign a diagnostic classification to a particular participant. Although the Rorschach data have uses that extend beyond the assignment of a diagnostic classification (e.g., description of personality dynamics, assessment of personality structure, etc.), when use of the instrument is extended into the realm of diagnosis, the rate of error becomes an important consideration.

Several studies have examined the rate of error associated with use of the constellation scores from the Comprehensive System, such as the Schizophrenia Index (SCZI), Depression Index (DEPI), Coping Deficit Index, Hypervigilance Index, Suicide Constellation, and Obsessive Style Index. For example, Ganellen (1996) cited findings showing that the SCZI has an overall correct classification rate of 92% and 86% for psychotic dis-
orders and that the DEPI has correct classification rates of 79% and 88% for identifying depression. Meyer (1993) demonstrated that response frequency is a mediating factor that impacts on the ability of the Rorschach indexes to discriminate between various diagnostic groups, and when response frequency is average or high, the diagnostic validity of Rorschach indexes is better than in brief records.

Wood et al. (1996a, 1996b) pointed out that the DEPI has shown rather poor diagnostic power in cross-validation studies and falls prone to what is termed *shrinkage* during cross-validation. The results of independent studies have shown that the DEPI does not have a strong relation with self-report measures of depression (Bail, Archer, Gordon, & French, 1991; Meyer, 1993). Moreover, the Rorschach indexes need to be investigated further in independent research. However, the *Daubert* opinion does not require that a particular rate of error be zero; it is only necessary that expert opinion be based on a foundation where the potential rate of error can be established. In this respect, the Rorschach indexes have rates of error established, and the statistical power of the Rorschach as a research instrument, although relatively low in some respects, is similar to that found in other areas of behavioral science (Acklin, McDowell, & Ornduff, 1992). Therefore, the Rorschach’s rate of error for certain decisions can be determined and cited when using Rorschach data in forensic evaluations.

**General Acceptance**

Under the *Frye* test, general acceptance of a particular method or technique is the central factor that determines whether expert testimony is admissible. Under the *Daubert* standard, general acceptance is a major but not primary consideration. Therefore, whether a psychometric instrument has achieved wide acceptance is a major consideration when selecting psychological tests for forensic application.

The best measure of general acceptance is the pattern and frequency of psychological test use by practicing professionals. As noted earlier, the Rorschach has achieved wide acceptance as defined by frequency of its use. Surveys have consistently shown that the Rorschach is utilized by over 80% of mental health agencies (Lubin et al., 1971, 1984; Piotrowski & Keller, 1989). In a recent survey, the Rorschach was cited as the third most widely used psychological assessment instrument; it was utilized by 82% of a sample of 412 practicing clinical psychologists engaged in psychological assessment services (Watkins, Campbell, Nieberding, & Hallmark, 1995).

In a study of psychological test use in criminal forensic examinations, Borum and Grisso (1995) found that the Rorschach was used in 32% of criminal responsibility evaluations and 30% of evaluations of competency to stand trial. Another study by Ackerman and Ackerman (1997) on child custody evaluation practices showed that the Rorschach was the second most widely used instrument with adults (used by 48% of respondents) and the sixth most commonly used instrument for children (used by 27% of respondents). Although the relevance of Rorschach data to competency to stand trial and child custody evaluations is not as clear as it is for criminal responsibility evaluations, the data cited in surveys demonstrate that the Rorschach is a widely used instrument. Its general acceptance among those who use psychological tests can be demonstrated.
Relevance and Helpfulness

The FRE and the Daubert standard require that any proposed expert testimony be both relevant and helpful to the trier of fact. In addition, the criteria outlined by Heilbrun and Marlowe emphasize the need for psychometric evidence to be relevant in some way to the psycholegal issue being addressed by the expert witness. With respect to Rorschach testimony, admissibility under this standard will be determined not by any psychometric properties of the instrument, but rather by the manner in which it is used by the mental health professional. In other words, the forensic professional who relies on the Rorschach must demonstrate, through written reports or oral testimony, how Rorschach results are used in a particular case.

Rorschach results are likely to have some direct relation to the legal issues at hand when the test participant’s personality structure is directly relevant. For example, reality testing, the capacity to handle stress, and how one handles emotions may all be directly related to issues such as mental state at the time of offense or risk of violence to others. In other cases, the connection between Rorschach data may be less clear, and the relevancy will need to be more clearly stated. For instance, Rorschach data in child custody cases do not address directly parenting capacity or the quality of the parent-child relationship; however, the instrument can provide useful data on issues that have an impact on these factors, such as personality characteristics of the parents. Again, it is up to the mental health professional to establish the relevance of the Rorschach data to the issue being litigated.

Overall, admissibility or inadmissibility of Rorschach-based testimony, where relevance and helpfulness are primary issues, appears to be determined by how the test data are utilized, rather than by any specific characteristics of the instrument itself. This is supported by a survey of legal decisions on the Rorschach that documents when Rorschach testimony is deemed inadmissible, it is primarily due to invalid references being made by the expert witness (Meloy et al., 1997).

Falsifiability

The Daubert opinion cites falsifiability as a criterion to point out that expert testimony must be capable of being empirically testable for it to be admissible. In other words, the statements constituting a scientific explanation must be capable of being subjected to empirical test. In fact, the debate over the validity of the Rorschach is itself a reflection of the testability and falsifiability of the instrument; there are those who have for many years been trying to “falsify” the instrument. The use of Rorschach data for explaining and describing personality is particularly germane because some of the explanations derived from the instrument are fully capable of being tested as the research cited herein suggests. Problems arise when Rorschach data are utilized to provide explanations about unconscious processes and psychoanalytic hypotheses that are more difficult to subject to empirical testing. Of course, there are exceptions to these limitations, such as the work of Bornstein (1996) on oral dependency. The mental health professional who utilizes the Rorschach in forensic settings can establish
the testability, or falsifiability, of the method by using data that have been subjected to empirical analysis. Therefore, reliance on the Structural Summary from the Comprehensive System or other empirically established indexes will keep the expert’s utilization of the Rorschach within the permissible bounds of testability or falsifiability.

Response Style Interpretation

Although none of the legal tests for admissibility require an instrument to have some method for evaluating the individual’s response style to be deemed admissible. Heilbrun recommended that this be a criterion for test selection in forensic evaluations. Objective instruments such as the MMPI–2, MCMI-III, and PAI all have scales that measure response style, and one specialized measure, the Structured Interview of Reported Symptoms (Rogers, Bagby, & Dickens, 1992) has shown particularly good capacity for identifying malingering and dissimulation.

The Rorschach was once thought to be impervious to faking, malingering, and other forms of response bias (Stermac, 1988); however, the capacity of the instrument to identify various forms of biased responding has been studied. Participants who have been instructed to portray an impression of psychological health have been shown to provide significantly more popular responses ($P$), fewer stress related determinants ($es$), fewer inappropriate combinations ($FABCOM, INCOM$), and fewer dramatic or sensational contents (Seamons, Howell, Carlisle, & Roe, 1981). Exner (1991) reported on research that has shown individuals undergoing child custody evaluations, where the motivation to deny psychopathology is high, provide a high number of personalized ($PER$) and popular ($P$) responses and show high levels of intellectualization.

Malingering has also been examined on the Rorschach by a number of investigators. Essentially, the results suggest that there are trends in Rorschach data that may be indicative of malingering, but no definitive patterns emerge. For instance, informed malingerers could not be distinguished from patients with genuine psychosis (Albert, Fox, & Kahn, 1980), and a specific Rorschach index designed to detect malingering has not shown sensitivity (Basch & Alpert, 1980). Other studies have supported the finding that Rorschach protocols characterized by good form quality that include bizarre or inappropriate combinations are suggestive, but not conclusive, of malingering (Seamons et al., 1981). A restricted number of responses has also been a consistent finding among studies in which some form of dissimulation (i.e., denial, malingering, etc.) is suspected (Perry & Kinder, 1990).

In general, the studies conducted on dissimulation and the Rorschach have established trends in structural summary data that are suggestive of biased responding. However, no clear diagnostic cutoffs or operating characteristics have been established. Therefore, it is important to integrate other data sources when arriving at conclusions related to the possibility that a particular participant is dissimulating. Nevertheless, the Rorschach data can be reviewed for possible signs of biased responding because there are methods to evaluate the validity of the Rorschach data for a particular participant (e.g. number of responses, relative number of form-only responses, etc.).
CONCLUSIONS

An analysis of the current clinical and research status of the Rorschach reveals that it meets professional and legal standards for admissibility of psychometric evidence and expert testimony. This conclusion rests on the foundation of a large body of literature that exists for the Exner Comprehensive System because this method of administration and scoring is standardized, has documented psychometric characteristics, and has been the primary subject of Rorschach research over the past 20 years. Other Rorschach systems have not been as well documented and it is therefore recommended that any forensic application of the Rorschach rely primarily on the Comprehensive System. Moreover, primary use of structural summary data, as opposed to sequence and nonstandardized content analysis, is recommended because of the availability of research to support reliability and validity.

There are those who may continue to argue that the Rorschach is not an appropriate psychometric instrument for any use and that it should be banned not only from forensic but also clinical settings. However, as surveys of test usage continue to demonstrate, the Rorschach has maintained a very high level of acceptance and use. Moreover, debates such as the one between Wood et al. (1996a, 1996b) and Exner (1996) are fruitful because they contribute to refinements of the instrument’s use and also constitute “‘good science’ in part because it increases the likelihood that substantive flaws in methodology will be detected” (Daubert, 1993, p. 2797). One should also consider the fact that no legal or professional standard requires that a psychometric instrument be a perfect measure of psychological constructs. Moreover, legal admissibility does not require that all scientific issues be completely resolved to the satisfaction of all members of the professional community. Expert testimony must rest on methodology that is generally accepted, testable, standardized, relevant, and helpful. In all of these respects, the Rorschach is an appropriate methodology to utilize in forensic evaluation. However, satisfaction of these criteria does not guarantee that one’s testimony or reliance on the Rorschach will be admissible or given weight in any particular court of law.

As both Heilbrun and Marlowe noted, a major governing factor in use of psychometric instruments is the burden placed on the forensic professional to provide clear reasoning concerning how psychometric data relates to the specific psycholegal issue being evaluated. Meloy et al. (1997) found that although the Rorschach is cited favorably by courts; in those few instances where Rorschach testimony was limited or excluded it was due to invalid inferences being made by the psychologist, rather than the individual properties or characteristics of the instrument. Therefore, it is incumbent on the professional using the Rorschach in forensic settings to make clear inferences that have relevance to the issues being addressed. In some forensic evaluations, such as fitness to proceed or child custody evaluations, the connection may not be clear at first and the expert may need to provide more explicit reasoning that ties the data to a given issue. In other forensic evaluations, such as risk of violence or mental state at the time of offense, the connection may be more evident.

The Rorschach appears to be a useful instrument for addressing various diagnostic issues in forensic assessment. However, one must be aware of the various legal and profes-
sional standards guiding use of psychometric tests in this area. Although not all Rorschach issues have been resolved, this does not preclude its use in forensic settings. As noted in the *Daubert* opinion, “vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof” (p. 2798) are the most appropriate methods for attacking evidence. Disagreement among professionals does not constitute adequate grounds for the preclusion of evidence in legal settings; it is rules of evidence and case law that guide admissibility. The Rorschach can be defended in court adequately if the professional has a sound grasp of the professional literature and familiarity with legal standards for admissibility of expert testimony. The professional must also be prepared to articulate clearly his or her reasoning surrounding Rorschach data and its relevance to a particular psycholegal issue, and to know the research to address any attacks raised on cross-examination.

ACKNOWLEDGMENT

I thank J. Reid Meloy for his assistance during the completion of this article and two anonymous reviewers for their helpful comments and suggestions.

REFERENCES


3. RORSCHACH ADMISSIBILITY


McCann’s (1998) article on the admissibility of the Rorschach stands out in the literature on both the Rorschach and forensic psychological testing in general as an exemplary model for assessing the soundness of psychometric evidence for court and the expert testimony that flows from it. Because McCann’s overall structure and clarity of analysis remains fresh today, his original article is included in this book as an essential resource for forensic psychologists using the Rorschach. He reviewed the legal tests for admissibility, as well as professional standards for deciding what psychological tests to use in forensic settings. McCann noted that the Rorschach fared well in these analyses and that, when the Rorschach was deemed inadmissible, the judicial judgment arose from misuse of the data, rather than characteristics of Rorschach itself. The purpose of this section is to provide the reader with a brief update of relevant literature since McCann’s article.

An Update of McCann’s “Defending the Rorschach in Court”

F. Barton Evans
McCann (1998) noted a growing difference of professional opinion about the Rorschach and its use in forensic settings. Although the Rorschach has met with controversy since its beginning (see Viglione & Rivera, 2003), a series of criticisms began to surface claiming that the Rorschach lacked scientific reliability and validity (Dawes, 1994; Wood, Nezworski, & Stejskal, 1996) and the suggestion was made that it should not be used in court (Wood et al., 1996). Indeed, Garb (1999) made an unprecedented call for a moratorium on use of the Rorschach, and Grove and Barden (1999) argued for the inadmissibility of the Rorschach in court under Daubert/Kumho. This small, but vocal, faction of authors spawned a series of repetitious articles critical of the Rorschach (e.g., Grove, Barden, Garb, & Lilienfeld, 2002; Hunsley & Bailey, 1999, 2001; Lilienfeld, Wood, & Garb, 2000; Nezworski & Wood, 1995; Wood & Lilienfeld, 1999; Wood, Nezworski, Garb, & Lilienfeld, 2001a, 2001b; Wood, Nezworski, & Stejskal, 1996; Wood, Nezworski, Stejskal, & McKinsey, 2001), as well as those in support of excluding Rorschach testimony in the court room (Dawes, 1999; Grove & Barden, 1999; Grove et al., 2002; Lilienfeld et al., 2000). Forensic psychologists using the Rorschach are strongly encouraged to read these papers and be prepared to address the points raised in them.

Rorschach researchers have responded in strength and numbers to these critiques. For example, Weiner (1996) stated that those who are critical of the Rorschach have “not read the relevant literature of the past 20 years, or, having read it, have not grasped its meaning” (p. 206). Since Weiner’s article, well-designed meta-analyses (see Gronnerod, 2006, 2003; Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999; Rosenthal, Hiller, Bornstein, Berry, & Brunell-Neuleib, 2001) and other research have demonstrated that, when properly used, the Rorschach is a valid and reliable instrument that can yield valuable information. Rather than list the voluminous research support for the Rorschach, the reader is referred to Viglione and Meyer (chap. 2, this vol.) as a primary contemporaneous resource providing a thorough review of the scientific status of the Rorschach. In response to the growing discrepancy between the mounting evidence of scientific robustness of the Rorschach and the public’s perception of the Rorschach as “junk science” (fueled by the small faction disparaging the Rorschach in the media), the Board of Trustees of the Society for Personality Assessment (2005) published “The Status of the Rorschach in Clinical and Forensic Practice: An Official Statement by the Board of Trustees of the Society for Personality Assessment,” also known as the Rorschach White Paper (copy of this article can be found on the Web site of the Society for Personality Assessment at www.personality.org). The statement is written in a language suitable for judges, attorneys, administrators, and the public and also includes a substantial compilation of research references and tables to support its claims. Based on a careful review of the science underlying the test, the Rorschach White Paper concludes that “the Rorschach possesses documented reliability and validity similar to other generally accepted test instruments used in the assessment of personality and psychopathology and that its responsible use in personality assessment is appropriate and justified” (p. 221).

Furthermore, since McCann (1998), there are several important articles challenging the claims that the Rorschach is inadmissible as a scientifically valid instrument in court. Gacono, Evans, and Viglione (2002; chap. 1, this vol.) summarized the main issues for the forensic psychologist preparing for forensic testimony and contrasted the polemical
debate of Rorschach detractors with a review of the substantial literature supporting the use of the Rorschach in court. In an article on the Rorschach as exemplar of dealing with challenges to psychological assessment instruments in forensic settings, Hilsenroth and Striker (2004) provided essential reference materials and offered guidelines for proceeding with challenges to the Rorschach in court. They emphasized that debate or the lack of unanimous agreement among the scientific community about a psychological instrument does not automatically translate into inadmissibility. They also pointed out that forensic expertise resides primarily in how the psychological instrument is used, rather than the Rorschach itself. In addition, Hilsenroth and Striker (2004) cogently informed the reader that legal debate does not require the same neutrality or objectivity as scientific debate and the critiques of Rorschach detractors often fall into the category of legal debate. Next, Ritzler, Erard, and Pettigrew (2002a, 2002b) specifically challenged Grove and Barden’s (1999) and Grove et al.’s (2002) contentions for the inadmissibility of the Rorschach in court. Ritzler et al. (2002a) leveled a strongly worded criticism of Grove and Barden, stating that they “overlooked or minimized a substantial body of empirical data supporting the reliability and validity of the Rorschach Comprehensive System” and further “misinterpreted the language and intent of the Supreme Court decisions.” The articles by Ritzler et al. contribute to the growing chorus of concern about the selective use of the scientific literature by Rorschach detractors to bolster their arguments (see also Gacono, Evans, & Viglione, 2002). Erard and Evans (2006) took the Rorschach detractors to task for failing to follow the very principles of good science that they claim as the basis of their critiques, specifically factual accuracy (acceptance of the value of empirical data and analysis), thoroughness (objective review of all existing literature, including research contrary to one’s opinions), and humility (admission of potential limitations in conclusions). In conclusion, although there is debate about the use of the Rorschach in court, for the well-prepared Rorschach expert, little has changed substantively to contradict McCann’s (1998) conclusions regarding its acceptability in court other than an abundance of new meta-analyses and research supporting the underlying science of the Rorschach.

Since McCann’s article, the Rorschach has maintained an important role and wide acceptance in professional practice settings. More recent surveys of professional test usage document that the Rorschach remains a commonly used psychological assessment instrument in the United States. Camara, Nathan, and Puente’s (2000) survey found that the Rorschach is the second most widely used personality assessment instrument among clinical psychologists and neuropsychologists. Archer and Newsome (2000) found it continued to be the second most widely used personality assessment instrument used in the assessment of adolescents. In forensic assessment, Quinnell and Bow (2001) and Boccaccini and Brodsky (1999) found that the Rorschach was the third most widely utilized personality assessment instrument used respectively in custody evaluations and personal injury evaluations. These surveys replicate the findings reported by McCann (1998), suggesting a noteworthy history of stable use of the Rorschach.

Despite the weight of evidence supporting the clinical and forensic use of the Rorschach, Lillienfeld et al. (2000) claimed that the Rorschach is “highly controversial” and not widely accepted in the relevant scientific community, citing “sustained and often
withering criticisms directed at projective testing during the past several decades” (p. 27). Schutz and Evans (chap. 12, this vol.) criticize Lilienfeld et al. for failing to mention or reference research in refereed journals contrary to their opinion and for essentially misinterpreting and misrepresenting the meaning of research by Piotrowski, Belter, and Keller (1998) on the impact of managed care on the decreased use of projective techniques. Lally’s survey (2003) of diplomats in forensic psychology assessed these experts’ opinion on the use of specific psychological tests in six areas: mental state at the time of offense, risk for future violence, risk for future sexual violence, competency to stand trial, competency to waive Miranda rights, and malingering. He categorized experts’ ratings of the Rorschach as equivocal-unacceptable for mental state at the time of offense and unacceptable to risk of violence assessment (53%), risk for future sexual violence (52%), competency to stand trial (60%), and to waive Miranda rights (63%), and malingering (55%). Whereas clearly this study suggests the possibility of a limited, if any, role for the Rorschach in these criminal matters, the survey leaves open the question of the basis for these judgments, especially background and training of the experts and their knowledge of the psychological literature on the use of the Rorschach in criminal matters. In light of the substantial research on the Rorschach with criminal populations beginning with Gacono and Meloy (1994) and of applications amply demonstrated in this book (chaps. 7–11, 16–20, 22, and parts of 23), the results of the Lally (2003) survey could be substantially misleading in the court room, if accepted uncritically.

McCann (1998) expressed his concern that there was an appearance of limited literature on use of the Rorschach in forensic settings. Unlike books for using the MMPI (Pope, Butcher, & Seelen, 2006) and the Millon inventories (McCann & Dyer, 1996) in forensic settings, Rorschach experts heretofore had no central resource for the Rorschach, but were limited to scouring the psychological literature for guidance in its use in court. The publication of The Handbook of Forensic Rorschach Assessment addresses McCann’s concern and demonstrates the broad use of the Rorschach in a variety of forensic settings.

Lastly, McCann noted that the Rorschach fared well in both in published legal decisions both in the courts in general (Weiner, Exner, & Sciara, 1996) and in appellate courts (Meloy, Hansen, & Weiner 1997). Updating the 10 years since Meloy, Hansen, and Weiner’s study, Meloy (chap. 4, this vol.) reported on the authority of the Rorschach in court, reviewing 150 published appellate citations from federal, state, and military courts between 1996 and 2005. He reported an average yearly citation rate nearly three times the rate of the previous study and found that, most frequently, the Rorschach did not receive special attention, but was discussed as one of a battery of psychological tests. Meloy found that, when the Rorschach was singled out, such legal cases could be classified in four categories: “the definitions and functions of the test in the court’s eyes; inappropriate use of the test; qualified use of the test; and d) criticism of the test” (p. 80). He noted that none of these cases included a Daubert challenge to the scientific adequacy of the test and that “There was not one case in which the Rorschach was ridiculed or disparaged by opposing counsel” (p. 85). Meloy concluded “it is clear that the test continues to have authority, or weight, in higher courts of appeal throughout the United States” (p. 86).

Additionally, Evans and Schutz (chap. 12, this vol.) conducted a Lexis Nexus search using “Rorschach and Child Custody” for legal citations from the last 50 years
(1956–2005), which yielded 43 hits. They found only five cases that touched on the admissibility or scientific status of the Rorschach, which fell into three of Meloy’s (2007) categories: the inappropriate use of the Rorschach, qualified use of the Rorschach, and criticism of the Rorschach. These legal cases showed that the court understood bad practice when they saw it, it recognized potential problems with the Rorschach were not sufficient grounds for appeal, and in the two cases, the court was split on whether or not the Rorschach was useful or unreliable. In the one case where the court found the Rorschach unreliable, a published dissent from one of the state appellate judges noted that the Rorschach had been admitted two previous times in that state, suggesting a further split in the opinion about the Rorschach.

In conclusion, McCann’s (1998) opinions about the legal admissibility of the Rorschach continue to be supported and confirmed. The Rorschach continues to demonstrate impressive scientific sturdiness that meets applicable guidelines for legal admissibility. It persists in its wide acceptance in both clinical and forensic professional practice settings and its legal authority within the courtroom has been affirmed. With the publication of The Handbook of Forensic Rorschach Assessment, the use of the Rorschach in a broad variety of forensic settings is now clearly established.

ACKNOWLEDGMENT

I wish to express my gratitude to Judy Maris for her support and invaluable editing help.

REFERENCES


