

FAILING THE BURDEN OF PROOF: THE SCIENCE AND ETHICS OF PROJECTIVE TESTS IN CUSTODY EVALUATIONS

Steven K. Erickson
Scott O. Lilienfeld
Michael J. Vitacco

Erard (2007/this issue) considers our review of psychological tests for family court evaluations to be inadequate and inaccurate. Because psychological evaluations play such a crucial role in many family court matters, it is imperative that we respond clearly and forcefully to his erroneous conclusions. Space does not allow us to respond to every charge put forth in the preceding article. Instead, we examine selected claims made by the author and argue that our conclusions regarding the use of psychological tests in family court matters are correct and harmonious with evidentiary rules and scientific principles.

Erard's criticism focuses on two central issues. First, he disavows our analysis of objective tests. Second, he believes that our treatment of projective tests (which he misleadingly refers to as "performance-based" tests, an extremely heterogeneous category that also subsumes all neuropsychological and laboratory tests) is unduly harsh. We strongly disagree on both scientific and ethical grounds. Before we delve into Erard's specific points, we review several fundamental precepts of science that provide a crucial backdrop for our discussion.

The scientific method is the backbone of all empirical knowledge. It separates speculation from fact, philosophy from science, and opinion from objectivity. Indeed, it was these principles to which the Supreme Court alluded in the watershed case of *Daubert v. Merrell Dow Pharmaceuticals* (1993):

The primary locus of this obligation is Rule 702, which clearly contemplates some degree of regulation of the subjects and theories about which experts may testify . . . The subject of an expert testimony *must be* "scientific . . . knowledge." . . . But, in order to qualify as "scientific knowledge," an inference or assertion must be [derived by the scientific method] (*Daubert* at 589, emphasis added).

As Erard points out, psychological tests are used in a process of hypothesis testing and theory building that allows clinical scientists to arrive at meaningful conclusions about the phenomena under study. Nevertheless, psychological tests must meet certain basic standards of scientific rigor, including adequate norms, reliability, and validity to provide accurate information. A test with little or no reliability and validity is not merely useless; it can be harmful because it can decrease the net validity of a psychological evaluation, often leading to erroneous clinical conclusions (Lilienfeld, Wood, & Garb, in press).

In science, as in law, the onus of demonstrating that a technique possesses adequate psychometric properties falls exclusively on those who propound its use. After all, the test's proponents are deriving clinical conclusions that can affect individuals' lives. Many of them are charging large sums of money for administering and interpreting this test. In some cases, they are offering expert testimony on the basis of this test's findings. Note that the

burden of proof does not fall on critics, who are merely calling on the proponents to provide supportive evidence. Erard neglects this vital rule of science, which leads him to draw profoundly misguided conclusions.

OBJECTIVE TESTS

Erard correctly notes that both the MMPI-2 and MCMI are imperfect measures of personality. In this respect, we are actually sympathetic to several of his criticisms of these measures, many of which we did not discuss because of space constraints. Indeed, there are continued discussions among professional psychologists regarding the limitations of these tests and ways to improve them. These concerns are more serious in terms of the MCMI, which is why we were equivocal regarding its use in forensic contexts. There are no serious calls, however, for the prohibition of the MMPI-2 in forensic cases. Contrary to Erard's assertion, several studies have examined the MMPI-2 in custodial disputes (Ollendick, 1984; Siegel, 1996; Bathurst & Gottfried, 1997; Posthuma & Harper, 1998; Siegel & Langford, 1998; Bagby, Nicholson, Buis, Radovanovic, & Fidler, 1999), including a study by a Rorschach inkblot test proponent, which we cited in our original article (Medoff, 1999).

Moreover, the overwhelming weight of evidence strongly supports the MMPI-2 as a reliable and valid measure of many personality traits. Legitimate concerns regarding some psychometric properties of the MMPI-2 (e.g., temporal stability of two-point codes) have not engendered calls for a moratorium on the test's use in clinical or forensic settings. In family court matters, the largest problem with the MMPI-2, as we noted, is arguably the appearance of defensiveness by the test taker, which some psychologists may interpret as "faking good" as an attempt to hide personality flaws. As we clearly stated, such interpretations are highly suspect and should only be made when additional evidence supports such conclusions.

Outside of this concern, there is no evidence that the MMPI-2 routinely classifies psychologically healthy individuals as maladjusted nor that its most widely interpreted indices (e.g., its eight major clinical scales) are almost entirely devoid of validity. As we will see, the same cannot be said of most projective techniques. Incidentally, Erard's assertion that the reported validity of the MMPI-2 scales may be inflated by "spurious monomethod variance" (p. 172) is a striking nonsequitor and betrays a lack of understanding of the distinction between construct and criterion validity. The MMPI-2 is used to assist in detecting psychopathology, so positive correlations with other measures of psychopathology (including interview-based and self-report measures) provide direct corroborations of its criterion validity.

PROJECTIVE TECHNIQUES

THE NUMEROUS PROBLEMS WITH THE RORSCHACH INKBLOT TEST

Erard claims that the Rorschach inkblot test is no less problematic than objective tests. This statement is so glaringly inconsistent with the mountain of research evidence to the contrary that it places the rest of his analysis in doubt. Indeed, the Rorschach¹ has a decades-long history of controversy that continues to grow as additional studies demonstrate its many shortcomings. These concerns are not merely academic pet peeves, but lasting and

substantial concerns from numerous respected social scientists regarding the Rorschach's validity and reliability. Erard does not refer to the litany of prominent assessment scholars over the past half century, including Lee J. Cronbach, Paul Meehl, Hans Eysenck, Joseph Zubin, Robyn Dawes, and Anne Anastasi, who have raised serious questions concerning the Rorschach's validity (Wood, Lilienfeld, & Garb, 2000). Paramount among contemporary concerns is the Rorschach's propensity to show psychologically healthy people—including children—as *severely* maladjusted.

Indeed, in 1999 and 2000, a group of well-known Rorschach experts reported several important studies on the Rorschach (Shaffer, Erdberg, & Haoian, 1999; Hamel, Shaffer, & Erdberg, 2000). Two of these studies were reported in the *Journal of Personality Assessment*, a journal well known to be sympathetic to Rorschach proponents. In the first of these studies, the researchers administered the Rorschach, the WAIS-III (an intelligence test), and the MMPI-2 to 123 adult nonpatients living in the community. The results from the MMPI-2 and WAIS-III for the group were average or even slightly above average compared with other Americans. In only one respect did these typically healthy Americans stand out: their Rorschach scores² indicated that a sizable proportion were seriously disturbed. For example, one in six of the participants scored in the pathological range of the Schizophrenia Index. Nearly a third gave at least a single reflection response, which according to Exner's (1991) scoring criteria (p. 173), is indicative of pathological narcissism. These same scholars subsequently published a study (Hamel et al., 2000) involving 100 preadolescent children with no history of psychological problems and above-average scores in psychological adjustment as determined by a well-validated measure, the Conner Parent Rating Scale-93 (Conners, 1989). Despite the weight of evidence that these children were emotionally healthy, the Rorschach indicated that 60% had pathological scores on the Schizophrenia Index. More than half had form scores that indicated thought disorder—a sign of psychosis. Nearly half had pathological scores on the Depression Index.

Similar studies abound, including a study by Mittman (1983), which found that, when psychologists trained by the Rorschach workshops classified patients based on Exner's norms, they misidentified more than 75% of the normal individuals as psychiatrically disturbed.

Erard cites a well-known meta-analysis³ (Hiller, Bornstein, Brunell-Neuleib, Rosenthal, & Berry, 1999) as evidence that the Rorschach meets accepted psychometric standards. This meta-analysis indeed suggested that *some* of the Rorschach scores are valid, but Erard's claim that this study "conclusively demonstrated that the validity of the Rorschach is essentially equivalent to . . . the MMPI-2" (p. 175) is bewildering. This meta-analysis examined the results of only 30 randomly selected studies and examined only a small subset of important Rorschach variables. Inexplicably, Erard does not inform readers that approximately 160 of the 180 Rorschach Comprehensive System scores have not been adequately validated by independent investigators. This includes such pertinent and widely used scores as the Depression Index, Egocentricity Index, and indexes advocated for assisting with the diagnosis of posttraumatic stress disorder in forensic evaluations (Wood et al., 2000).

Erard sidesteps this point by asserting that the "validation of each score has been described" in Exner's (2000) book, but this cavalier reassurance is hardly satisfactory from a scientific or ethical standpoint. Erard does not inform readers that this book consists mostly of validation studies performed in Exner's own workshop, that many of these studies have not been subjected to peer review, and that most have not been replicated by independent investigators. Moreover, Erard's statement that "most important [Rorschach] variables are

subject to continuing validation” (p. 175) is remarkable. It implies that variables that are still being validated are appropriate for clinical use and for assisting with custody decisions. The ethical implications of this assertion are breathtaking.

Studies have also cast serious doubt on the Rorschach’s scoring and inter-rater reliability. Studies have shown that only 50% of Rorschach’s scores meet the gold standard of .90 (Acklin, McDowell, Verscell, & Chan, 2000; McGrath et al., 2005), which is often viewed as the minimum reliability needed for routine clinical applications. When this threshold is lowered to .80, about 75% of the Rorschach scores meet this burden. Erard’s claim that “the scoring reliability of the Rorschach has been firmly established” (p. 174) again misleads the reader, as it neglects to note the substantial proportion of Rorschach scores whose scoring reliability is doubtful. Studies of the Rorschach have been roundly criticized for their use of “percentage of agreement” in establishing inter-rater reliability because it tends to yield inflated estimates of true consistency (Wood, Nezworski, & Stejskal, 2006; Wood, Nezworski, & Stejskal, 1996; Erickson, 2003). Although better methods exist for measuring inter-rater reliability (i.e., correlation coefficients), several studies have revealed that the Rorschach fares even worse under these analyses (Acklin et al., 2000; Guarnaccia, Dill, Sabatino, & Southwick, 2001; Meyer et al., 2002).

In sum, so many studies have called into question the reliability and validity of the Rorschach that its use in forensic settings is wholly inconsistent with the tenets of *Daubert*. With the possible exception of the handful of indices that have received empirical support, such as its thought disorder indices,⁴ we can find no good reason to recommend the use of the Rorschach in family court evaluations.⁵

If the tenets of *Daubert* are insufficient to withhold the use of the Rorschach in most or all family court matters (as many jurisdictions relax the formal rules of evidence in family court), then the American Psychological Association’s Ethics Code satisfies this prohibition (American Psychological Association, 2002). Section 9.02(a) & (b) state:

- (a) Psychologists administer, adapt, score, interpret, or use assessment techniques, interviews, tests, or instruments in a manner and for purposes that are appropriate in light of the research on or evidence of the usefulness and proper application of the techniques.
- (b) Psychologists use assessment instruments whose validity and reliability have been established for use with members of the population tested. When such validity or reliability has not been established, psychologists describe the strengths and limitations of test results and interpretation.

Clearly, the Rorschach fails to meet these ethical thresholds, because the wealth of current research has demonstrated that the use of the Rorschach is not appropriate in forensic evaluations. Regrettably, it is virtually unheard of for a psychologist to provide a disclaimer in a forensic evaluation that reveals the limitations of the Rorschach and how he or she avoided bias in the interpretation as required in subsection 9.02(b).

Even more troubling, however, is the fact that the Society for Personality Assessment, a thinly-veiled advocacy group of Rorschach proponents, recently published a white paper titled “The Status of the Rorschach in Clinical and Forensic Practice: An Official Statement by the Board of Trustees of the Society for Personality Assessment” (a group whose flagship journal, the *Journal of Personality Assessment*, was initially called *Rorschach Research Exchange* and, later, the *Journal of Projective Techniques*) that advocates the use of the Rorschach in forensic evaluations (Society for Personality Assessment, 2005). This

blue-ribbon committee of Rorschach followers makes the absurd assertion that the Rorschach is as valid and reliable as electrocardiograms, mammography, magnetic resonance imaging, dental radiographs, Papanicolaou smears, and other mainstream medical tests (Society for Personality Assessment, 2005, p. 219). Remarkably, this paper neglects to mention the raging scientific controversy regarding the Rorschach's norms, the fact that most of the Rorschach's scores lack demonstrated evidence for validity, or the questionable reliability of many Rorschach scores. This document, which Erard cited without criticism or qualification, is much closer to propaganda than science. Like that of many other manuscripts by Rorschach proponents, it fails to straightforwardly acknowledge the many failings of the technique.

Likewise, Erard chastises us for the claim that the Rorschach is based largely on outmoded psychoanalytic theory. He claims, as did Exner (1992), that the Rorschach is "atheoretical" and a "problem-solving task" (Erard, 2007/this issue, p. 173). Such assertions are without merit and obscure the Rorschach's historical and theoretical origins. In fact, the Rorschach is deeply steeped in psychoanalytic ideas dating back to Hermann Rorschach, the founder of the infamous inkblots. As other Rorschach researchers have stated unequivocally in the *Journal of Personality Assessment*:

It appears then that *experienced* clinicians and the educated lay public are in agreement that the Rorschach is principally a projective technique (Aronow, Reznikoff, & Moreland, 1995, p. 218; emphasis added).

Psychoanalytical theory has played a central role in the spread of Rorschach's test and in training thousands of clinicians (Acklin, 1990, p. 384).

Using psychoanalytical theory to guide Rorschach interpretation is nothing new. Researchers, using psychoanalytic object-relations theories, have contributed significantly (Kleiger, 1992, p. 295).

More revealing, however, is the fact that many of the Rorschach scores unabashedly rely on psychoanalytic terminology such as "ego" (e.g., Ego Impairment Index, Conceptual Ego Strength Index). More fundamentally, in many cases Rorschach interpretation of a subject's response to ambiguous inkblots is firmly rooted in psychoanalytic concepts. Whether a participant has dependency issues because of food responses derives not so much from empirical observations regarding dependency and oral responses to ambiguous images, but from unproven psychoanalytic notions of psychosexual development. Ironically, the Rorschach indexes that appear to have the strongest evidence for validity, such as its thought disorder scores, are *nonprojective* in nature; they assume that the inkblots do resemble things and examine the extent to which people's perceptions deviate from them (Dawes, 1994). A short inventory of Rorschach studies published in the *Journal of Personality Assessment* reveals that many are imbued with psychoanalytic ideas, including the dubiously titled "A Neurotic Lawyer: AIDS or Oedipus?" (Peterson, 1994). Erard's reference to projective techniques as "performance-based" tests only obscures the largely psychoanalytic origins of the Rorschach and is manifestly misleading.

OTHER PROJECTIVE TECHNIQUES

Erard criticizes our evaluation of other projective techniques. We remain steadfastly behind our conclusions because of the substantial limitations of these tests in terms of psychometrics, administration, and peer review. As we noted, the clinical use of the Thematic Apprehension Test (TAT) is problematic because there is a paucity of established studies

on the TAT's reliability and validity. We agree with Erard that the TAT has promise for assessing object relations (Westen, 1991), but Erard neglects to mention that the TAT object relations indices lack adequate population-based norms and are therefore not appropriate for clinical or child custody use.

Other projective techniques used in family court evaluations suffer from similar shortcomings. The forensic assessment tests such as the Bricklin Perceptual Scales, Perception of Relations Test, Parent Awareness of Skills Survey, and Parent Perception of Child Profile test have *few or no* peer-reviewed articles supporting their use. Thus, Erard and other supporters of these unsubstantiated tests apparently expect the courts to rely on their *ipse dixit* in deciding whether testimony based on these inadequately validated tests should be admitted. Although Erard claims that we misunderstand "applicable evidentiary standards," we boldly surmise that courts frown upon such practices, as most assuredly does the American Psychological Association's Ethical Code.

CONCLUSION

Psychological testimony can be valuable in family court matters where complicated and delicate issues of emotional health, parental bonding, and parental strife are frequently present. Such testimony is strengthened when validated and reliable tests are utilized to help evaluators arrive at accurate conclusions regarding the parties involved in litigation. Testimony based on faulty psychological tests presents a direct threat to the integrity of clinical conclusions and should be vigorously challenged by counsel. Explanations that such tests are just "one piece of data" are unacceptable and grossly at variance with fundamental doctrines of science. They also demonstrate ignorance of the literature on dilution effects, which indicates that adding low validity tests to a battery of valid tests can decrease the net validity of clinical judgments and predictions (Ruscio, 2000; Lilienfeld et al., in press).

Finally, Erard's response is especially remarkable for the information that it does *not* present. Astute readers will note that Erard does not cite a single study demonstrating that the Rorschach or other projective techniques possess validity in custody decisions. Yet he confidently assures readers that these techniques should be used in family court, asking them to accept his word that they "offer complementary contributions in developing rich, three-dimensional evaluations in family law matters" (p. 176). Those seeking to include scientifically problematic instruments in forensic evaluations must bear in mind that the onus lies on them to demonstrate to the scientific community that such tests possess sufficient psychometric properties to properly assess target behaviors. This burden is justifiably high, because expert test interpretations can contribute to profoundly important clinical and legal decisions. In family court matters, the stakes are even higher, given the potentially life-altering issues at hand.

NOTES

1. For the ease of nonexpert readers, we use the term "Rorschach" to mean John Exner's *The Rorschach: A Comprehensive System*, the dominant scoring system for the Rorschach inkblot test.

2. Contrary to Erard's claim, software programs exist to assist in the scoring and interpretation of the Rorschach, including one authored by John Exner himself (see Exner, 1985). There are others, as an Internet search with the terms "Rorschach" and "software" demonstrates.

3. Erard claims that we fundamentally misunderstand meta-analysis. The definition in our original article, albeit simplified (by the request of reviewers), is correct.
4. While the thought disorder indices of the Rorschach have demonstrated good construct validity, their norms remain problematic and, thus, of questionable utility.
5. Contrary to Erard's assertion, there are indeed *Daubert* and *Frye* jurisdictions. As mentioned in our original article, however, most *Frye* jurisdictions have incorporated the spirit of the *Daubert* holding through case law.

REFERENCES

- Acklin, M. W. (1990). Psychoanalytic theory in Rorschach testing. *Journal of Personality Assessment, 55*, 384–386.
- Acklin, M. W., McDowell, C. J., II, Verscell, M. S., & Chan, D. (2000). Interobserver agreement, intraobserver reliability, and the Rorschach comprehensive system. *Journal of Personality Assessment, 74*, 15–47.
- Aronow, E., Reznikoff, M., & Moreland, K. L. (1995). The Rorschach: Projective technique or psychometric test. *Journal of Personality Assessment, 64*, 213–228.
- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*, 1060–1073.
- Bagby, R., Nicholson, R. A., Buis, T., Radovanovic, H., & Fidler, B. J. (1999). Defensive responding on the MMPI-2 in family custody and access evaluations. *Psychological Assessment, 3*, 24–28.
- Bathurst, K., Gottfried, A. W., & Gottfried, A. E. (1997). Normative data for the MMPI-2 in child custody litigation. *Psychological Assessment, 9*, 205–211.
- Conners, K. (1989). *Manual for Conners Rating Scales*. North Tonawanda, NY: Multi Health Systems.
- Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 579–599 (1993).
- Erickson, S. (2003). Psychological testimony on trial: Questions arise about the validity of popular testing methods. *New York State Bar Journal, 75*, 19–26.
- Exner, J. (1985). *Rorschach interpretation assistance program*. San Antonio, TX: Psychological Corporation.
- Exner, J. E. (1992). Some comments on “a conceptual critique of the EA:es comparison in the Comprehensive Rorschach System.” *Journal of Personality Assessment, 4*, 297–300.
- Guarnaccia, V., Dill, C. A., Sabatino, A., & Southwick, S. (2001). Scoring accuracy using the Comprehensive System for the Rorschach. *Journal of Personality Assessment, 77*, 464–474.
- Hamel, M., Shaffer, T. A., & Erdberg, P. (2000). A study of nonpatient preadolescent Rorschach protocols. *Journal of Personality Assessment, 75*, 280–294.
- Hiller, J., Bornstein, R., Brunell-Neuleib, S., Rosenthal, R., & Berry, D. (1999). A comparative meta-analysis of Rorschach and MMPI validity: Special section I. The utility of the Rorschach in clinical assessment. *Psychological Assessment, 11*, 278–296.
- Kleiger, J. H. (1992). A conceptual critique of the EA: escomparison in the Comprehensive Rorschach System. *Journal of Personality Assessment, 4*, 288–296.
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (in press). Why questionable psychological tests remain popular. *Scientific Review of Alternative Medicine*.
- McGrath, R. E., Pogge, D. L., Stokes, J. M., Cragolino, A., Zaccario, M., Hayman, J., et al. (2005). Field reliability of comprehensive system scoring in an adolescent inpatient sample. *Assessment, 12*, 199–209.
- Medoff, D. (1999). MMPI-2 validity scales in child custody evaluations: Clinical versus statistical significance. *Behavioral Sciences & the Law, 17*, 409–411.
- Meyer, G. J., Hilsenroth, M. J., Baxter, D., Exner, J. E., Fowler, J. C., Piers, C. C., et al. (2002). An examination of interrater reliability for scoring the Rorschach comprehensive system in eight data sets. *Journal of Personality Assessment, 78*, 219–274.
- Ollendick, D. G. (1984). Scores on three MMPI alcohol scales of parents who receive child custody. *Psychological Reports, 55*, 337–338.
- Peterson, C. A. (1994). A neurotic lawyer: AIDS or Oedipus? *Journal of Personality Assessment, 63*, 10–26.
- Posthuma, A. B., & Harper, J. F. (1998). Comparison of MMPI-2 responses of child custody and personal injury litigants. *Professional Psychology: Research and Practice, 29*, 437–443.
- Ruscio, J. (2000). The role of complex thought in clinical prediction: Social accountability and the need for cognition. *Journal of Consulting and Clinical Psychology, 68*(1), 145–154.
- Shaffer, T. W., Erdberg, P., & Haoian, J. (1999). Current nonpatient data for the Rorschach, WAIS–R, and MMPI-2. *Journal of Personality Assessment, 73*, 305–316.

- Siegel, J. C. (1996). Traditional MMPI-2 validity indicators and initial presentation in custody evaluations. *American Journal of Forensic Psychology, 14*, 55–63.
- Siegel, J. C., & Langford, J. S. (1998). MMPI-2 validity scales and suspected parental alienation syndrome. *American Journal of Forensic Psychology, 16*, 5–14.
- Society for Personality Assessment. (2005). The status of the Rorschach in clinical and forensic practice: An official statement by the Board of Trustees of the Society for Personality Assessment. *Journal of Personality Assessment, 85*, 219–237.
- 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., & Stejskal, W. J. (1996). The comprehensive system for the Rorschach: A critical examination. *Psychological Science, 7*, 3–10.
- Wood, J. M., Nezworski, M., & Stejskal, W. J. (1996). Thinking critically about the comprehensive system for the Rorschach: A reply to Exner. *Psychological Science, 7*, 14–17.