PSYCHOLOGICAL TESTING AND CHILD CUSTODY EVALUATIONS IN FAMILY COURT: A DIALOGUE

A CRITICAL EXAMINATION OF THE SUITABILITY AND LIMITATIONS OF PSYCHOLOGICAL TESTS IN FAMILY COURT

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

Psychologists are frequently consulted by the courts to provide forensic evaluations in a variety of family court proceedings. As part of their evaluations, psychologists often use psychological tests to assess parents, guardians, and children. These tests can have profound effects on how psychologists arrive at their opinions and are often cited in their reports to the court. However, psychological tests vary substantially in their suitability for these purposes. Most projective tests in particular appear to possess little scientific merit for evaluations within family court proceedings. Despite these serious limitations, expert testimony derived from evaluations using both projective and objective tests is often admitted uncontested. This article reviews the psychometric properties of psychological tests that are widely used in family court proceedings, cautions against their unfettered use, and calls upon attorneys to inform themselves of the limitations of evaluations that incorporate these tests.

Keywords: *child custody*; *forensic evaluation*; *psychometrics*; *projective techniques*; *objective tests*

Perhaps nowhere is the use of mental health experts more contentious than in the realm of family court proceedings. This controversy is highlighted by numerous articles that have appeared in law reviews, bar journals, the popular press, and social science journals that have questioned the reliability and validity of testimony by mental health experts in family court (Eaton, 2004; Emery, Otto, & O'Donohue, 2006; Erickson, 2003; Jurow & Schaul, 2005; Tippins, 2005; Tippins & Wittmann, 2005). This criticism is based partially on the perception that many mental health professionals formulate their expert opinions based upon pseudoscientific assessment techniques that contribute to erroneous clinical conclusions (Emery, 2005; Otto, Edens, & Barcus, 2000). The recommendations that flow from such conclusions and subsequent opinion testimony can have persuasive authority before the court and often encroach on the ultimate issue (Waller & Daniel, 2004). Whether experts should opine on the ultimate issue is a hotly contested issue that many critics contend, and a recent New York trial court held, should be reserved solely for the trier of fact (*John A. v. Bridget M.*, 2005; Grisso, 2005; Tillbrook, Mumley, & Grisso, 2003; Tippins & Wittmann, 2005); however, other scholars have thoughtfully argued otherwise (Rogers & Ewing, 2003).

These criticisms have led some to call for a moratorium on ultimate opinion testimony by mental health experts, as well as aggressive cross-examination by attorneys when such testimony is presented (Tippins & Wittmann, 2005). Nonetheless, expert testimony

FAMILY COURT REVIEW, Vol. 45 No. 2, April 2007 157–174

by mental health experts in family court proceedings is routinely admitted and can exert profound effects by directly questioning the parental fitness of a party; the appropriateness of custodial and visitation orders; the veracity of physical, emotional, or sexual abuse allegations; and the likelihood of successful parental reunification (Emery et al., 2006; Erickson, 2003).

Mental health evaluations are often performed by licensed psychologists, who by virtue of their training are uniquely proficient in administering and interpreting psychological tests that measure a variety of phenomena, including personality traits, indices of psychopathology (e.g., depression, schizophrenia, psychopathy), intelligence, and cognitive impairment. These tests can provide useful information to psychologists that can assist them in determining the psychological strengths and weaknesses of the parents and child, the presence of a major mental illness, and personality characteristics that may hinder implementation of certain custody or visitation arrangements.

Despite their promise and potential clinical utility, no psychological test is infallible. Moreover, not all psychological tests are created equal in terms of their suitability for family court evaluations. In this article, we critically examine the most popular tests used in family court proceedings and examine whether these tests meet the admissibility thresholds for expert testimony under the lens of accepted scientific and legal standards of reliability and validity. We also probe the consequential arguments presented by many experts who rely on these tests to form the basis of their testimony. Finally, we suggest that attorneys carefully consider motions in limine, vigorous cross-examination, and other means to exclude expert testimony derived from psychological tests that have scientifically unacceptable reliability, validity, or both.

PSYCHOLOGICAL TESTS USED IN FAMILY COURT

Psychological tests are written, visual, or oral evaluations administered to assess the cognitive and emotional functioning of children and adults (American Psychological Association, 1985). Psychologists routinely use tests to assist them in formulating clinical and expert opinions because they often provide useful data that can corroborate or contradict diagnostic hypotheses, provide illumination into personality processes of the examinee, or detect the presence of a certain trait, characteristic, or illness. Psychological tests are typically divided into two types: objective (structured) and projective. This distinction underscores the fundamental differences in theory and administration that these tests use to arrive at their conclusions. Although both approaches have significant limitations related to family court proceedings, most projective tests in particular have substantial flaws that raise serious questions about their usefulness in these cases.

In addition, evidence regarding the use of psychological tests in family court is limited. Most studies have examined custodial evaluations (with some small exceptions examining human figure drawings and child sexual abuse); however, there is a paucity of research related to visitation, abuse and neglect, and juvenile delinquency proceedings. Nonetheless, because most psychological tests discussed in this article evaluate potentially relevant personality features, their results may be extrapolated to family court evaluations if: (a) their underlying psychometric properties are sound; (b) the interpretation made by the psychologist prudently considers factors that may influence the results (e.g., defensiveness, stress of litigation); and (c) the psychologist acknowledges the limitations of generalizing a test's results to parental abilities, child attachment, or other similar concepts.

Table 1
General recommendations for objective tests

• Objective tests should be used to evaluate response-related styles and personality traits. Tests are not interpreted in a vacuum; instead, they should be used in conjunction with behavioral observations and collateral interviews to evaluate fully the parent and child.
• Psychologists should use objective testing to assist with diagnostic questions and use the results accordingly. As such, clinicians should tailor their test use to answer specific referral questions (e.g., Is depression hindering the father's parenting skills?).
• Contextual factors (e.g., demands of the evaluation) warrant consideration when interpreting test results. As an illustrative example, some evaluators "bump up" clinical scales to account for defensiveness in custody evaluations. This type of interpretation, when not empirically based, is scientifically and ethically questionable.
• Evaluators must be able to discuss their rationale for using objective inventories in their evaluations. It is not sufficient to state, "I give them to everyone." By focusing their evaluation methods on specific questions, clinicians can generate hypotheses relevant to child custody evaluations.
• Evaluators should only rely on well-validated instruments and scales when using objective tests. Many available scales (e.g., the Solomon Scale on the MMPI-2) are not adequately validated and should not be interpreted.

MULTISCALE INVENTORIES

Objective tests enjoy widespread use and acceptance among psychologists when used in general clinical practice. These tests generally rely upon the response of the examinee to a series of questions that are answered using a true/false or other nominal scale. Upon completion of the test, the examinee's answers are typically compared against the normative sample used when the test was constructed (i.e., subjects, ideally drawn from the general population, whose responses are used as a baseline) to generate scaled scores that measure various psychological traits. Most objective tests share two fundamental features. First, an examinee's score on the test requires no judgment on the part of the psychologist. Second, the respondent is asked to answer the questions in one of a very few fixed ways, such as by circling True or False. As a consequence, the number of potential responses to objective test items is limited.

COMMON MULTISCALE PERSONALITY INVENTORIES

Psychologists in family court proceedings are confronted with the difficult task of applying psychological test results with other clinical data to arrive at meaningful conclusions regarding the best interests of the child at issue. This task often requires assessing for mental health problems that could interfere with effective parenting or impede the successful implementation of custodial or visitation plans. Often complicating the issue of an effective psychological evaluation is the presence of defensiveness and minimization, the dominant response styles of parents undergoing child custody evaluations (Bagby, Nicholson, Buis, Radovanovic, & Fidler, 1999). That is, parents frequently present themselves as unrealistically virtuous while concealing their psychological deficits. Assessing underreporting is a challenge for evaluators and is one rationale for employing multiscale inventories in child custody evaluations.

Defensiveness can be defined as the minimization or denial of symptoms that often occurs in the presence of an external reward (Vitacco & Rogers, 2006). In some cases, however, defensiveness may reflect an ingrained psychological trait that relieves distressing

Table 2
Common psychological tests used in family court evaluations, their purported purpose, and whether they meet scientific and evidentiary standards

Test	*Purported purpose*	*Meet scientific and evidentiary standards?*	*Comments*
Minnesota Multiphasic Personality Inventory, 2nd Ed. (MMPI-2)	Assessment of general psychopathology	Yes	Solid reliability and validity data; known error rate; hundreds of peer-reviewed publications
Millon Multiaxial Personality Inventory, 3rd Ed. (MCMI-III)	Assessment of general psychopathology, particularly personality features	Equivocal	Problems with construct validity; known error rate; good publication record
Rorschach Inkblot Test	Assessment of general psychopathology	No	Many problems with reliability and validity; underlying methodology is suspect; peer-review record is questionable
Thematic Apprehension Test (TAT)	Assessment of general psychopathology	No	Substantial reliability and validity problems; unknown error rate; underlying methodology is suspect; uneven peer-review record
Human Figure Drawings	Unconscious aspects of personality	No	Unreliable and invalid; underlying methodology suspect; poor peer-review record
Anatomical Dolls	Nonverbal reporting of psychologically salient material	No	Unreliable and invalid; underlying methodology not accepted; few peer-reviewed publications
Bricklin Perceptual Scales (BPS)	Parental competence and ascertainment of the "Parent of Choice"	No	Unreliable and invalid; unknown error rate; underlying methodology not generally accepted; few peer-reviewed publications
Perception of Relationships Test (PORT)	"Gut-level" responses of the child; parental preference of child	No	Unreliable and invalid; unknown error rate; underlying methodology not generally accepted; few peer-reviewed publications
Parent Awareness of Skills Survey (PASS)	Parental understanding of childcare scenarios	No	Unreliable and invalid; unknown error rate; underlying methodology not generally accepted; few peer-reviewed publications
Parent Perception of Child Profile (PPCP)	Parental understanding of child's developmental stages; parental competence	No	Unreliable and invalid; unknown error rate; underlying methodology not generally accepted; few peer-reviewed publications

emotions. Whether indicating deceit or anxiety, the presence of defensiveness often interferes with the proper interpretation of psychological test results (Bagby et al., 1999; Bathurst, Gottfried, & Gottfried, 1997). Not surprisingly, defensive profiles occur frequently in family court evaluations and exceed those seen in other clinical samples such as hospitalized psychiatric patients (Posthuma & Harper, 1998; Siegel, 1996). In this section, we will examine the benefits and potential pitfalls of two common multiscale inventories used in family court evaluations: the Minnesota Multiphasic Personality Inventory-Second Edition (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and the Millon Clinical Multiaxial Inventory—Third Edition (MCMI-III; Millon, 1997).

THE MMPI-2

Without question, the MMPI-2 is the most frequently administered psychological test in family court evaluations and is the most widely used personality measure in the United States (Otto & Collins, 1995). Its use in family court evaluations is ubiquitous. In a survey of 202 family court evaluators, Ackerman and Ackerman (1997) found that 92% of clinicians frequently used the MMPI-2 in their evaluations. As such, a review of the MMPI-2's use in family court evaluations is clearly warranted, especially as it relates to the evaluation of defensiveness and the assessment of maladaptive personality traits.

The MMPI-2 is a paper-and-pencil measure consisting of 567 true/false items that measures numerous areas of psychopathology in adults. Administration consists of asking the subject to answer all 567 questions (usually by filling in a "bubble sheet"), scoring of the measure (usually by computer) which generates numerous scales (together called a profile), and interpretation of the profile by the psychologist. In general, the MMPI-2 profile consists of two types of scales: those that measure validity and those that measure psychopathology. The validity scales are designed to assess whether the examinee's profile presents an accurate representation of his or her psychopathology and information regarding the examinee's approach to the test (i.e., open v. defensive, attentive v. haphazard answering). Some of the elements that these validity scales measure include exaggerated and unlikely psychopathology, inattentive answering by the examinee, and underreporting of psychopathology. It is the latter that is often an issue in family court evaluations.

Because the MMPI-2 is a global measure of abnormal personality, most examinees will produce some elevation on the clinical scales. This is because tests like the MMPI-2 inevitably detect sub-clinical presence of personality idiosyncrasies that are present in the general population and are only of clinical concern when they exceed a certain threshold. Extreme elevations on some validity scales suggest underreporting of psychopathology by the examinee. Interpretations of conscious or unconscious attempts by examinees to present themselves as unrealistically well-adjusted are common. Such defensive response styles have the capacity to invalidate an entire profile and may signify an examinee who wishes to conceal psychopathology of clinical concern. Moreover, even if the profile remains valid, defensive responses have the propensity to suppress clinical scales, rendering accurate interpretation of the profile difficult.

Most examinees in family court referrals produce defensive response styles. Thus, it is incumbent upon the psychologist to choose the most accurate validity scales when interpreting MMPI-2 profiles. In a meta-analysis of 14 studies that evaluated 12 separate MMPI-2 indices of underreporting, Baer and Miller (2002) found the Wiggins Social Desirability (WSD) and Positive Malingering (PM) scales to be most effective at detecting

minimization in child custody evaluations. One question that must be answered, however, is how much underreporting is necessary before the clinician should consider the MMPI-2 profile invalid. In discussing this point, Medoff (1999) differentiated between statistical and clinical significance. Medoff pointed out that, although defensiveness is often found in custody examinees in higher levels than in individuals from comparison groups not undergoing custody evaluations (i.e., statistical significance), this fact often provides little case-specific information (i.e., clinical significance). As such, evaluators should not over-interpret elevations on MMPI-2 defensiveness scales. Additionally, evaluators should rely more heavily on indices that have been validated in family court samples and that have demonstrated large effect sizes (i.e., WSD and PM) in measuring underreporting (Baer & Miller, 2002).

From counsel's perspective, expert testimony claiming that underreporting of psychopathology by a parent on the MMPI-2 suggests intentional deception by the parent should be challenged. Because research has shown that most parents do underreport psychopathology on the MMPI-2 during litigation, clinical impressions that a parent was intentionally deceptive during an expert's examination must be buttressed with other evidence independent of the MMPI-2 results. More crucially, however, clinical conclusions of deception should differentiate between common desires by parents to minimize their faults when their parental character is under scrutiny and defensiveness that suggests entrenched psychopathology.

An additional area of concern is interpretations of MMPI-2 clinical scales (i.e., scales used to assess psychopathology) when the validity scales suggest underreporting by the examinee. This practice is problematic and can lead to spurious clinical findings. Research has shown that experts often make clinical interpretations of MMPI-2 results when the validity scales clearly indicate an invalid profile (Hartman-Crouch, 2000). Such practices are troubling because once validity thresholds are exceeded, the veracity of any subsequent interpretations are highly suspect.

One promising approach for understanding underreporting in custody evaluations is to consider competing hypotheses. Strong, Greene, Hoppe, Johnston, and Olesen (1999) conducted a taxometric analysis of 412 (206 pairs) parents who were given the MMPI-2 as part of custody evaluations. Taxometric analysis is useful in determining if data fit into identifiable categories that exist in nature, rather than merely in the minds of clinicians. Their taxometric analyses revealed one distinct category of underreporting in custody evaluations, which they labeled Impression Management (IM). IM is intentional self-distortion with the overall goal of creating a positive self-image. However, the researchers also found evidence of a dimensional style that they labeled Self-Deceptive Positivity (SDP). Strong et al. (1999) posited that SDP is unintentional and more reflective of a stable personality trait. Notably, this analysis of the MMPI-2 is consistent with evidence suggesting that defensiveness is intentional in some cases, whereas it reflects a more stable trait in others. Thus, when faced with elevated indicators of defensiveness on the MMPI-2, clinicians should not assume that their examinees are intentionally manipulating. Instead, they should first consider the possibility that a favorable trait (e.g., high degree of optimism) is the reason for high defensiveness scores. Making explicit assumptions about high underreporting scales places the examinee at considerable risk. For instance, if clinicians believe that they are being intentionally misled, that could adversely affect their recommendations.

Compared with the assessment of validity response styles, a paucity of research has focused on the utility of the MMPI-2 in evaluating psychopathology in family court examinees. Often these examinees exhibit elevations on clinical scales that measure antisocial personality features and paranoia, suggesting high levels of mistrust and anger (Bathurst

et al., 1997; Hartman-Crouch, 2000). The antisocial scale (Psychopathic deviate scale; Scale 4) is often elevated in individuals who are experiencing significant marital difficulties, which is consistent with undergoing a family court evaluation (Otto & Collins, 1995). The importance of contextual factors, including the demand characteristics of the evaluation (i.e., questions of parental unfitness) must be considered when making MMPI-2 interpretive statements. Clinicians relying too extensively on the MMPI-2 may neglect to account for factors, such as current life stressors (e.g., custody fight), that are responsible for clinical elevations.

In sum, the MMPI-2 is widely used by psychologists in family court evaluations. Although it justifiably has a solid reputation in the clinical field, care must be undertaken by the psychologist when using it in forensic contexts, particularly in family court matters. Defensive profiles are common and are often interpreted as indicative of deception on the part of the examinee. Such claims should only be made when other data support such a conclusion. Additionally, ad hoc adjustment of suppressed clinical scales or interpretation of invalid profiles are improper and should be challenged by counsel.

THE MCMI-III

The MCMI-III (Millon, 1997) is a measure comprising 175 true/false items that compose 28 scales intended for the overall assessment of psychopathology in adults. Like the MMPI-2, the MCMI-III has limited research supporting its use in assessing family court litigants. In a study of 259 parents, McCann et al. (2001) found the most common elevation on the MCMI-III in family court evaluations occurred on a scale measuring defensiveness. Other common, but subclinical elevations occurred on scales measuring Compulsive, Histrionic, and Narcissistic personality traits. In an earlier study conducted with 50 couples undergoing custody evaluations, elevations were found on the same four scales (Lampel, 1999).

There are several hypotheses to consider when examining these results. First, several scales on the MCMI-III have limited construct validity (i.e., empirical evidence that they measure what they purport to measure). Rogers, Salekin, and Sewell (1999) combined 33 separate MCMI studies in a meta-analysis and found adequate construct validity for only three (Avoidant, Schizotypal, and Borderline). Second, the MCMI-III manual reports that elevations on Histrionic, Compulsive, and Narcissistic scales are commonplace. As such, it is extremely difficult to make interpretative statements and develop subsequent recommendations on scale elevations considered normative. Lastly, as noted by Halon (2001), elevations on Defensive, Compulsive, Histrionic, and Narcissistic MCMI-III scales can be explained by a host of factors, unrelated to any type of psychopathology or motivation to mislead the evaluator. Again, contextual factors, including the nature of the evaluation and psychologically understandable response styles (e.g., the desire to create a good impression), must be considered when interpreting the MCMI-III.

FINAL THOUGHTS FOR THE USE OF MULTISCALE INVENTORIES IN FAMILY COURT

We have reviewed two commonly used objective, multiscale inventories used in family court. These tests are used frequently to evaluate adults involved in family court litigation and possess adequate reliability and validity for the assessment of general psychopathology. Although research regarding their use in family court is limited, they remain important

tools to help evaluators confirm or reject clinical hypotheses. Caution is warranted, however, in interpreting validity scales of these tests that may indicate defensiveness by the examinee, as such results are often attributed to contextual factors present in family court litigation where parental competency is often at issue. Furthermore, the ad hoc adjustment of clinical scales suppressed by such defensive responding for interpretive purposes has little empirical support and should be avoided.

COMMON PROJECTIVE TESTS USED IN FAMILY COURT

Unlike objective tests, projective tests rely upon the examinee's unrestricted responses to vague stimuli or situations to elucidate traits or characteristics of clinical importance. Many of these tests are premised largely on psychoanalytic theories of personality and rely upon the subjective interpretation of the psychologist to an examinee's responses to a series of pictures, images, or open-ended questions. Often the subject matter being assessed is unclear to examinees, requiring them to draw conclusions based upon contemplation. Theoretically, such internal contemplation and the subsequent response by the examinee reveal important aspects of the examinee's mental state. Thus, supporters of projective measures purport that the answers constructed by the examinee in response to ambiguous stimuli reveal salient psychological characteristics that can be measured and interpreted by the psychologist. Unlike most objective tests, projective measures have been enshrouded in long-standing controversy by clinicians and researchers alike since their inception. The genesis of the controversy surrounding projective tests centers on their lack of empirical support, outdated psychoanalytical theoretical underpinnings, and suspect peer-reviewed publication record.

RORSCHACH INKBLOT TEST

Unquestionably, the most enduring projective test is the Rorschach inkblot test. It is widely used by psychologists in family court proceedings to assess personality traits and characteristics of both adults and children (Ackerman & Ackerman, 1997; Hagen & Castagna, 2001). The test consists of a series of ambiguous inkblots that are presented to the examinee. The psychologist queries examinees to elucidate their perceptions and interpretations of the inkblots. The examinee's responses are recorded and later scored by either the psychologist or computer software that derives numerous scores. These scores are then interpreted by the psychologist to arrive at conclusions regarding the examinee's psychological condition.

The Rorschach inkblot test has been criticized heavily on numerous grounds. There are serious questions regarding its ability to reliably and validly measure psychological constructs, such as psychopathology. Moreover, there is increasing evidence that the norms (population baselines) of the most commonly used system for Rorschach test scoring and interpretation, the Comprehensive System (CS; Exner, 2002), are inaccurate. These flawed norms often lead psychologists to overestimate examinees' psychopathology (Shaffer, Erdberg, & Haroian, 1999; Wood, Nezworski, Lilienfeld, & Garb, 2003; Wood, Nezworski, & Stejskal, 1996). For example, several studies have demonstrated that the Rorschach CS identifies healthy adults as maladjusted or even afflicted with severe mental illnesses such as schizophrenia (Grove, Barden, Garb, & Lilienfeld, 2002). Healthy examinees may appear to have profound mental abnormalities related to perceptions, logical thinking, and emotional functioning. Although some Rorschach test proponents (Meyer, 2001; Weiner, 2005) have

attempted to explain away these findings on methodological grounds, such as by claiming that the studies demonstrating flawed norms were based on unrepresentative samples, the evidence that the CS norms are flawed is extremely consistent across numerous studies. Moreover, there is little compelling evidence that the samples in these studies were unrepresentative of the general population (Wood, Nezworski, Garb, & Lilienfeld, in press).

More disturbing, the Rorschach CS can lead clinicians to classify psychologically healthy children as mentally ill or maladjusted. Research has demonstrated that above-average children with no history of mental illness can appear on the Rorschach inkblot test as psychotic, clinically depressed, cognitively impaired, or highly resistant to establishing and maintaining interpersonal relationships (Hamel, Shaffer, & Erdberg, 2000). The relevance of such a misdiagnosis in the realm of family court proceedings is obvious.

An additional concern regarding the Rorschach test is that many of its basic psychometric properties (i.e., underlying design elements) are highly problematic. The inter-rater reliability (i.e., how close two or more evaluators arrive at similar findings) of some commonly used CS scores is poor, with different evaluators arriving at widely differing scores *before any clinical interpretation is attempted* (Acklin, McDowell, Verschell, & Chan, 2000; Guarnaccia, Dill, Sabatino, & Southwick, 2001). More importantly, many of its scores bear little or no correlation with what they purport to measure and thus will lead to inaccurate clinical conclusions. For instance, the Depression Index has little association with a diagnosis of clinical depression (Wood, Nezworski, Garb, & Lilienfeld, 2001). Although the Rorschach test possesses moderate validity for detecting thought disorder and perceptual disturbance, its validity for detecting most personality traits (e.g., narcissism, aggression, impulsivity) and major psychiatric disorders appears to be negligible (Lilienfeld, Wood, & Garb, 2000).

On the positive side, global meta-analyses (i.e., mathematical syntheses combining all Rorschach test scores) indicate that the Rorschach test, *taken as a whole*, possesses modest validity that may even approach that of the MMPI-2 (e.g., Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999). Such evidence has frequently been invoked by Rorschach test proponents (e.g., Weiner, 2005) to justify the routine use of the Rorschach test in clinical practice. Such claims ignore the crucial point that the Rorschach CS consists of over 150 scores, the substantial majority of which have never been individually validated. That is, global meta-analyses tell us nothing about *which* CS scores are scientifically supported.

When more focused meta-analyses on specific Rorschach test scores have been conducted, the findings have been far less encouraging. Indeed, with the exception of Rorschach test scores designed to detect thinking disturbance and perhaps interpersonal dependency, there is scant support for the validity of the numerous scores derived from this instrument (Lilienfeld et al., 2000). It is perhaps telling that the most valid application of the Rorschach inkblot test, namely its use for detecting thought disorder, does not rest on psychoanalytic assumptions. Instead, the Rorschach test appears to be most valid when used as a *perceptual* rather than a projective test (Dawes, 1994).

A final problem with the Rorschach test is its heavy reliance on unpublished or unavailable studies in its chief administration and scoring manual, *The Rorschach: A Comprehensive System* (Exner, 2002). Critics have been unopposed in their assertion that well over half of the studies cited in the manual were never formally peer reviewed and have not been made available for review to the larger scientific community (Grove et al., 2002). The discovery that 221 of the 700 samples used to establish the adult norms for the 1991/1993 edition of the manual were duplicates raises further questions concerning the veracity of the scoring method used to arrive at diagnostic conclusions. In sum, with the possible exception of detecting severe thought disorder in parents, there appears to be scant support for the use

of the Rorschach test in family court evaluations. Moreover, even for those relatively few Rorschach test scores that possess adequate validity, flawed norms render their clinical interpretation in the courtroom highly suspect. The continued pervasive use of the Rorschach test in family court evaluations is unwise at best and unethical at worst. Astute counsel should vigorously challenge expert opinions formulated with Rorschach test data.

THEMATIC APPERCEPTION TEST

The Thematic Apperception Test (TAT) is a commonly used projective test in which an examinee is shown a series of pictures and asked to construct a relevant story concerning the characters on each card (Murray, 1943). The TAT cards depict vague scenarios (e.g., a man walking away from a distressed woman) and thus lend themselves to a variety of responses. The test has a total of 31 cards, most of which portray equivocal social situations, although Card 16 is entirely blank. Like the Rorschach inkblot test, the TAT has been heavily criticized for its lack of standardization, common ad hoc administration, inadequate training by most clinicians, and clinically insignificant incremental validity (Ball, Archer, & Imhof, 1994; Groth-Marnat, 2003; Hunsley, Lee, & Wood, 2003; Rossini & Moretti, 1997).

Even supporters of the TAT admit that its administration varies widely among clinicians and that most examiners fail to use an adequate number of cards (Hibbard, 2003; Hibbard, Farmer, Wells, & Difillipo, 1994), thereby rendering proper interpretation difficult or impossible. TAT advocates have even suggested that interpretation of this measure can sometimes reflect the *opposite* personality trait than what is actually present in the examinee. Known as the "Walter Mitty" effect, examinees' response may reflect their desire or admiration of a characteristic rather than a true personality feature (Loevinger, 1987). Others have noted that the TAT may be subject to an "inhibition effect" whereby examinees successfully restrict the overall personality measure when attempting to suppress true personality features (Lilienfeld et al., 2000).

There is little persuasive evidence that most common clinical uses of the TAT exhibit acceptable scientific reliability or validity. Even when properly administered, which appears to be rare in actual practice, the TAT generally suffers from weak internal consistency, test-retest reliability, and limited construct validity depending on the scoring method used (Entwisle, 1972; Fineman, 1977; Lilienfeld et al., 2000; Winter & Stewart, 1977; but see Spangler, 1992 and Westen, 1991, for probable exceptions in the domain of achievement needs and perception of interpersonal relationships). There is no published evidence validating the use of the TAT for assessing a child's mental stability, parental preference, or perception of family cohesion within the forensic context. A handful of studies have examined the TAT among physically and sexually abused children (Freedenfeld, Ornduff, & Kelsey, 1995; Ornduff & Kelsey, 1996; Pistole & Ornduff, 1994), but none has validated the TAT during ongoing litigation. Only a single study has shown a modest ability of the TAT to detect childhood attention deficit and hyperactivity disorder (Costantino, Colon-Malgady, Malgady, & Perez, 1991) and that study called for further research before adoption of the TAT for this purpose.

Given the notable deficits of the TAT, its continued use with children is troubling. It remains the second most used personality measure of children among psychologists involved in child custody evaluations (Hagen & Castagna, 2001). A possible explanation is the TAT's pictorial quality, which children may find appealing.

Supporters of the TAT's use in family court evaluations often suggest that "it's just one more piece of data" implying that administration of this flawed test is essentially harmless.

Such explanations demonstrate an ignorance of both scientific and evidentiary principles and should be forcefully challenged by counsel. Specifically, a large body of evidence indicates that "more is not necessarily better" in the domain of psychological assessment. The addition of salient but invalid data to extant test data may frequently result in a *decrease* in the accuracy of clinical judgments, largely because clinicians often attend too heavily to the invalid information (Garb, Wood, Lilienfeld, & Nezworski, 2005).

HUMAN FIGURE DRAWINGS

Human figure drawings remain popular among psychologists who assess children in custodial evaluations (Ackerman & Ackerman, 1997; Hagen & Castagna, 2001). Proponents of these tests claim that they are able to quickly access unconscious aspects of personality (Harris, 1963; Riethmiller & Handler, 1997). These tests operate by having examinees draw various animate or inanimate objects, including people, buildings, or animals. Variations of this class of methods include the popular Draw-A-Person (DAP) test, as well as tests that require examinees to draw groups of families (often their own) or people involved in an activity ("kinetic" drawing methods). The assessment made by these tests rests entirely on the examinee's drawings, and in most cases, the psychologist's general impressionistic evaluation of the content and style of the drawings. These tests are typically administered in an ad hoc manner, with few experts utilizing formal scoring techniques, which themselves suffer from serious psychometric problems (Hunsel et al., 2003; Lally, 2001; Lilienfeld et al., 2000). Consequently, the reliability and validity of these tests is typically weak or nonexistent (Kahill, 1984; Lally, 2001).

The prevailing popularity of these tests with children is probably due to their pictorial quality, ease of use, and unrestricted de facto interpretation methods. In the first author's experience, mental health experts frequently ask a child to draw a picture of his or her family with the expert suggesting to the court that, because the child neglected to include one parent or emphasized one parent over the other, the child "obviously" feels estranged from one parent and has a closer attachment to the other. There is no credible scientific support for this conclusion or for the broader claim that human figure drawings provide useful data regarding children involved in litigation. Even in nonlitigation settings, human figure drawings have been amply criticized for their failure to detect general psychopathology, lack of norms, poor inter-rater reliability, and validity that is only slightly better than chance (Burley, Handler, & Hammer, 1997; Hunsley et al., 2003; Kahill, 1984; Lally, 2001; Riethmiller & Handler, 1997; Scribner & Handler, 1987). The extremely weak psychometric properties of these tests and their frequent ad hoc administration make these tests ripe for evidentiary challenges.

ANATOMICAL DOLLS

Psychological assessment of children using anatomical dolls has endured throughout modern history (Sattler, 1998). Such dolls continue to enjoy widespread use among mental health experts despite a complete lack of standardization or agreed-upon scoring methods (Bartlett-Simpson, Kneeshaw, & Schaefer, 1993; Boat & Everson, 1988; Conte, Sorenson, Fogarty, & Rosa, 1991; Kendall-Tackett & Watson, 1992; Lie & Inman, 1991). The existence of over a dozen interview formats related to anatomical dolls renders research findings difficult or impossible to compare across investigations given that no identified method has gained prominence (Everson & Boat, 1990), except for the ad hoc administration that

plagues clinical practice. In fact, the American Psychological Association has determined that anatomical dolls have not met most of the minimal criteria for a valid psychological test (Koocher et al., 1996).

The purported concept behind anatomical dolls is that children can express material through doll play that they cannot report verbally (Hunsley et al., 2003). Anatomical dolls are often used in child sexual abuse investigations despite a dearth of research supporting the validity of this use (Everson & Boat, 1990; Hunsley et al., 2003; Kendall-Tackett & Watson, 1992; Lie & Inman, 1991). In one study, a majority of children without histories of sexual abuse engaged in touching, rubbing, poking, and pinching of various body parts on anatomical dolls, whereas another revealed that a quarter of 5-year-old boys without abuse histories placed the dolls in sexually suggestive positions when instructed to "show me what dolls can do together" (Boat & Everson, 1988; Everson & Boat, 1990). Other studies have failed to demonstrate acceptable reliability and validity of anatomical dolls and some have called for their prohibition in court-ordered evaluations (Skinner & Berry, 1993). Given these troubling findings, anatomical dolls should be avoided by mental health experts in family court evaluations.

SPECIALLY DESIGNED FORENSIC ASSESSMENT TESTS

The previously mentioned tests are general psychological tests that are used in a variety of clinical and forensic contexts, including family court forensic evaluations. A handful of tests are designed specifically to address family court issues. Although their tailoring for family court evaluations affords them a certain appeal, their uniform psychometric deficits severely undermine their value. Consequently, they should be avoided in clinical practice and expert opinions derived from such sources should be scrutinized by counsel.

BRICKLIN PERCEPTUAL SCALES

The Bricklin Perceptual Scales (BPS) is a projective test that purports to measure parental competence, supportiveness, follow-up consistency, and possession of admirable traits by having the child rate each parent on 32 different activities that supposedly assess these capacities (Bricklin, 1990; Emery et al., 2006). The child uses a stylus and indicates positive ratings for each parent on these 32 activities. The parent who receives the greater number of positives is designated the "Parent of Choice." The idea of the BPS is that its projective nature detects the child's unconscious preferences and therefore bypasses social desirability, distortion, and parental influence by obviating verbal responses from the child. However, there is no scientific evidence for this assumption. Moreover, the test (a) suffers from an absence of norms, (b) measures only a minority of potentially relevant parental capacities, and (c) lacks sufficient data regarding its validity (Heinze & Grisso, 1996; Melton, Petrila, Poythress, & Slobogin, 1997; Otto et al., 2000).

PERCEPTION OF RELATIONSHIPS TEST

The Perception of Relationships Test (PORT) is a series of seven projective drawing tasks completed by the child that hypothetically measure the "gut-level responses" that, when scored, reveal the caretaking parent of choice (Bricklin, 1993, p. 1). Serious questions regarding its reliability and validity properties, incomplete norms, vague administration, and scoring procedures, as well as other fundamental psychometric deficiencies preclude

the usefulness of this test (Heinze & Grisso, 1996; Melton et al., 1997; Otto, Buffingotn-Vollum, & Edens, 2003; Otto et al., 2000).

PARENT AWARENESS OF SKILLS SURVEY

The Parent Awareness of Skills Survey (PASS) is a projective test that utilizes 18 child-care scenarios and follow-up questions to assess parental strengths and weaknesses (Bricklin, 1990b). It, too, has serious shortcomings, including absence of reliability and validity data, norms, and defined scoring guidelines. Moreover, PASS has been criticized for the suggestion of its author that "the evaluator, by virtue of appropriate training in psychology and/or child development, can apply his or her own standards in assigning the suggested scores" (Bricklin, 1990b; Emery et al., 2006). In essence, the test's developer recommends that evaluators create their own scoring methods for the test, none of which can possess known reliability or validity criteria. This test falls so short of acceptable scientific principles that its use probably violates established ethical canons (American Psychological Association, 1994).

PARENT PERCEPTION OF CHILD PROFILE

The Parent Perception of Child Profile (PPCP) ostensibly measures a parent's understanding of child development across multiple domains (Bricklin & Elliot, 1991). The test relies on an undocumented supposition that parents who more accurately assess these criteria related to their own children make better parents (Emery et al., 2006). Moreover, the veracity of the parent's statements is left to the evaluator's judgment and investigation, rendering the test's inter-rater reliability doubtful. Nevertheless, this criticism is difficult to evaluate given the absence of any published reliability or validity data (Otto et al., 2000).

BURDEN OF PROOF

Experts are utilized in legal proceedings to provide testimony that helps prove or disprove the existence of an alleged fact. Litigation in family court frequently involves accusations regarding a party's psychological unfitness as a parent that engenders fervent recriminations that leave the court with the unenviable task of ascertaining which party will best provide the child with a healthy, nurturing environment. Additional proceedings in family court, including allegations of abuse or neglect, juvenile delinquency, and disputes over visitation are equally difficult, with statutory and case law providing only minimal guidance as to how the court should best decide these complex matters (Emery et al., 2006). Given the intimacy of family relationships, most allegations made about a party's mental unfitness are unsupported by impartial witnesses or other direct testimony. Moreover, they involve delicate questions of judgment as to what constitutes parental fitness and imprecise predictions of the stability of this phenomenon when both parties are usually within major life transitions.

Psychologists are relied upon by family courts to provide expert opinions regarding mental health issues involving litigants and how adult behavior in the past, present, and foreseeable future may impact a child. These experts are useful in providing such testimony given their advanced training in human behavior and their clinical experience. Although some critics contend that there exists little, if any, empirical support for forensic mental

health evaluations within the family court context (Tippins & Wittmann, 2005), these criticisms neglect to account for the expertise that psychologists possess regarding psychological disorders, child development, and assessment of personality strengths and weaknesses that are often directly relevant in family court litigation. Nonetheless, it is incumbent upon experts to utilize scientifically sound methods to arrive at their conclusions that form the basis of their expert opinion. The use of psychometrically poor or unsubstantiated assessment tests places an expert's opinion squarely in question because the data used to derive that opinion are inherently flawed. From both a legal and scientific perspective, the burden is on the proponent of the evidence to demonstrate that the method used is acceptable. If the proponent is unable to do so, the method should be recognized as nothing more than a source of hypothetical possibilities awash in a sea of uncertainties.

LEGAL PERSPECTIVE

The U.S. Supreme Court has grown increasingly wary of expert testimony and has signaled its concern by holding that trial courts should actively gauge whether proffered expert evidence meets certain thresholds before being admitted. These bedrock principles were established by the Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993) wherein the Court held that trial courts should consider whether (1) the underlying methodology and principles of the testimony have been or can be tested, (2) the principles and methodology have been subjected to peer review and publication, (3) the known or potential error rate is acceptable, and (4) the underlying principles have obtained general acceptance in the scientific community. Although many states do not explicitly follow the *Daubert* criteria and remain *Frye* jurisdictions (*Frye v. United States*, 1923), wherein expert testimony only needs to be "generally accepted" by the scientific community to be admitted, many of these jurisdictions have adopted the *Daubert* principles through case law (*John A. v. Bridget M.*, 2005). Consequently, some psychological tests critically viewed through the *Daubert* lens fail to meet these minimal standards that the courts have imbued with importance. Expert testimony derived through them should be challenged and excluded. Motions in limine that challenge the propriety of expert opinions based on these faulty tests can place the issue directly before the court prior to testimony by the expert. In harmony with *Daubert*, such motions should seek to preclude all testimony proffered by experts who use tests that have serious psychometric defects because the foundation of such opinions are flawed.

SCIENTIFIC PERSPECTIVE

Two basic maxims of science are that the proponent of a theory bears the onus of demonstrating its veracity and that the scientific method is fundamental for the advancement of testable truth claims. The scientific method holds that scientific discoveries are based upon observations and reasoning that propose tentative explanations—known as hypotheses—that are testable and replicable. A tenet of modern science is that scientific theories must be falsifiable (i.e., they must be framed in such a way to be potentially proven wrong); otherwise they are mere opinions or armchair conjectures. In psychometric assessment, fidelity to fundamental concepts of reliability, validity, and normative comparisons is crucial in fulfilling the scientific method.

Many psychological tests used in family court evaluations are marked by inadequate reliability and validity, and hence should not be used to form expert opinions. Some defenders

of these methods claim that such tests are "just one more piece of data," implying that even if faulty, their use is at worst harmless. This position demonstrates a disregard or unawareness of data, discussed earlier, on the diluting effects of invalid assessment information on existing information (Garb et al., 2005). Implicit in this argument is the idea that such problematic tests *might* yield incremental validity (i.e., validity above and beyond other assessment information) and that assessment conclusions can only benefit from the addition of more psychometric information. This widely invoked contention is scientifically indefensible, because a large body of evidence indicates that more psychometric information can sometimes lead to *less accurate* clinical judgments. Moreover, the cherry picking of invalid data in an impressionistic manner is wholly inconsistent with science and violates the tenets of *Daubert*. These pseudoscientific opinions have no place in the courtroom and should be vigorously challenged by counsel.

CONCLUSION

Psychologists provide a valuable service in family courts across the nation by offering expert opinions in cases often fraught with complex emotional issues. The use of psychological tests can aid psychologists in this endeavor. Nevertheless, considerable caution is warranted, as many of these tests—particularly most projective techniques—are inappropriate for the assessment of adults or children involved in family court litigation. Opinions formed by improper assessment techniques that incorporate these problematic tests are themselves vulnerable to evidentiary challenges and should be excluded, as they violate legal and scientific principles. Attorneys should familiarize themselves with commonly used psychological tests and vigorously challenge expert testimony derived from those that are scientifically unsound.

REFERENCES

Ackerman, M., & Ackerman, M. (1997). Custody evaluation practices: A survey of experienced professionals (revisited). *Professional Psychology: Research and Practice*, *28*, 137–145.

Acklin, M. W., McDowell II, C. J., Verschell, M. S., & Chan, D. (2000). Interobserver agreement, intraobserver reliability, and the Rorschach comprehensive system. *Journal of Personality Assessment*, *74*(1), 15.

American Psychological Association. (1985). *Standards for educational and psychological testing*. Washington, DC: Author.

American Psychological Association. (1994). Guidelines for child custody evaluations in divorce proceedings. *American Psychologist*, *49*, 677–680.

Baer, R. A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: A meta-analytic review. *Psychological Assessment*, *14*, 16–26.

Bagby, R. M., 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*, *11*, 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.

Ball, J. D., Archer, R. P., & Imhof, E. A. (1994). Time requirements of psychological testing: A survey of practitioners. *Journal of Personality Assessment*, *63*, 239–249.

Bartlett-Simpson, B., Kneeshaw, S., & Schaefer, C. (1993). The use of anatomical dolls to assess child sexual abuse: A critical review. *International Journal of Play Therapy*, *2*(2), 35–51.

Boat, B. W., & Everson, M. D. (1988). Use of anatomical dolls among professionals in sexual abuse evaluations. *Child Abuse & Neglect*, *12*(2), 171–179.

Bricklin, B. (1990a). *Bricklin Perceptual Scales manual*. Furlong, PA: Village Publishing.

Bricklin, B. (1990b). *Parent Awareness of Skills Survey manual*. Furlong, PA: Village Publishing.
Bricklin, B. (1993). *Tests manuals supplement #9*. Furlong, PA: Village Publishing.
Bricklin, B., & Elliot, G. (1991). *Parent Perception of Child Profile*. Furlong, PA: Village Publishing.
Burley, T., Handler, L., & Hammer, E. F. (1997). Personality factors in the accurate interpretation of projective tests. In E. F. Hammer (Ed.), *Advances in projective drawing interpretation* (pp. 359–377). Springfield, IL: Charles C. Thomas Publisher, Ltd.
Butcher, J., Dahlstrom, W., Graham, J., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
Conte, J. R., Sorenson, E., Fogarty, L., & Rosa, J. D. (1991). Evaluating children's reports of sexual abuse: Results from a survey of professionals. *American Journal of Orthopsychiatry*, *61*(3), 428–437.
Costantino, G., Colon-Malgady, G., Malgady, R. G., & Perez, A. (1991). Assessment of attention deficit disorder using a thematic apperception technique. *Journal of Personality Assessment*, *57*, 87–95.
Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).
Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.
Eaton, L. (2004, May 23). For arbiters in custody battles, wide power and little scrutiny. *New York Times*. Retrieved October 26, 2006, from http://www.nytimes.com.
Emery, R. E., Otto, R. K., & O'Donohue, W. T. (2006). A critical assessment of child custody evaluations: Limited science and a flawed system. *Psychological Science in the Public Interest*, *6*, 1–29.
Emery, R. E. (2005). Parental alienation syndrome: Proponents bear the burden of proof. *Family Court Review*, *43*, 8–13.
Entwisle, D. (1972). To dispel fantasies about fantasy-based measures of achievement and motivation. *Psychological Bulletin*, *77*(6), 377–391.
Erickson, S. (2003). Psychological testimony on trial: Questions arise about the validity of popular testing methods. *New York State Bar Journal*, *75*(6), 19–26.
Everson, M., & Boat, B. (1990). Sexualized doll play among young children: Implications for the use of anatomical dolls in sexual abuse evaluations. *Journal of the American Academy of Child and Adolescent Psychiatry*, *29*(5), 736–742.
Exner, J. E. (2002). *The Rorschach, basic foundations and principles of interpretation*. Hoboken, NJ: Wiley.
Fineman, S. (1977). The achievement motive and its measurement. Where are we now? *British Journal of Psychology*, *68*, 1–22.
Freedenfeld, R. N., Ornduff, S. R., & Kelsey, R. M. (1995). Object relations and physical abuse: A TAT analysis. *Journal of Personality Assessment*, *64*(3), 552–568.
Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).
Garb, H. N., Wood, J. M., Lilienfeld, S. O., & Nezworski, T. (2005). Roots of the Rorschach controversy. *Clinical Psychology Review*, *25*, 97–118.
Grisso, T. (2005). Commentary on "Empirical and ethical problems with custody recommendations": What now? *Family Court Review*, *43*, 223–228.
Groth-Marnat, G. (2003). *Handbook of psychological assessment* (4th ed.). Hoboken, NJ: Wiley.
Grove, W., Barden, C., Garb, H., & Lilienfeld, S. (2002). Failure of Rorschach-Comprehensive-System-based testimony to be admissible under the *Daubert-Joiner-Kumho* standard. *Psychology, Public Policy, and Law*, *8*, 216–234.
Guarnaccia, V., Dill, C., Sabatino, S., & Southwick, S. (2001). Scoring accuracy using the Comprehensive System for the Rorschach. *Journal of Personality Assessment*, *77*, 464–474.
Hagen, M., & Castagna, N. (2001). The real numbers: Psychological testing in custody evaluations. *Professional Psychology: Research and Practice*, *32*, 269–271.
Halon, R. L. (2001). The Millon Clinical Multiaxial Inventory-III: The normal quartet in child custody cases. *American Journal of Forensic Psychology*, *19*(1), 57–75.
Hamel, M., Shaffer, T., & Erdberg, P. (2000). A study of nonpatient preadolescent Rorschach protocols. *Journal of Personality Assessment*, *75*, 280–294.
Hartman-Crouch, T. S. (2000). The Minnesota Multiphasic Personality Inventory-2 in the context of child custody litigation. (Doctoral Dissertation, California School of Professional Psychology, 2000). *Dissertation Abstracts*, *61*, 2762.
Harris, D. (1963). *Children's drawings as measures of intellectual maturity*. New York: Harcourt Brace Jovanovich.
Heinze, M. C., & Grisso, T. (1996). Review of instruments assessing parenting competencies used in child custody evaluations. *Behavioral Sciences & the Law*, *14*, 293–313.
Hibbard, S. (2003). A critique of Lilienfeld et al.'s (2000) "The scientific status of projective techniques." *Journal of Personality Assessment*, *80*, 260–271.
Hibbard, S., Farmer, L., Wells, C., & Difillipo, E. (1994). Validation of Cramer's defense mechanism manual for the TAT. *Journal of Personality Assessment*, *63*, 197–210.

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

Hunsley, J., Lee, C. M., & Wood, J. M. (2003). Controversial and questionable assessment techniques. In S. O. Lilienfeld, J. M. Lohr, & S. J. Lynn (Eds.), *Science and pseudoscience in contemporary clinical psychology*. New York: Guilford Press.

John A. v. Bridget M., 791 N.Y.S.2d 421 (2005).

Jurow, G., & Schaul, B. (2005). Custody evaluations: Recommendations about recommendations. *New York Law Journal*, *234*, 4–10.

Kahill, S. (1984). Human figure drawing in adults: An update of the empirical evidence, 1967–1982. *Canadian Psychology*, *25*, 269–292.

Kendall-Tackett, K. A., & Watson, M. W. (1992). Use of anatomical dolls by Boston-area professionals. *Child Abuse & Neglect*, *16*, 423–428.

Koocher, G. P., Goodman, G. S., White, C. S., Friedrich, W. N., Sivan, A. B., & Reynolds, C. R. (1996). Psychological science and the use of anatomically detailed dolls in child sexual-abuse assessments. *Psychological Bulletin*, *188*, 199–222.

Lally, S. J. (2001). Should human figure drawings be admitted into court? *Journal of Personality Assessment*, *76*, 135–149.

Lampel, A. K. (1999). Use of the Millon Clinical Multiaxial Inventory-III in evaluating child custody litigants. *American Journal of Forensic Psychology*, *17*, 19–31.

Lie, G., & Inman, A. (1991). The use of anatomical dolls as assessment and evidentiary tools. *Social Work*, *36*, 396–399.

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, *1*(2), 27–66.

Loevinger, J. (1987). *Paradigms of personality*. New York: Freeman.

McCann, J. T., Flens, J. R., Campagna, V., Collman, P., Lazzaro, T., & Connor, E. (2001). The MCMI-III in child custody evaluations: A normative study. *Journal of Forensic Psychology Practice*, *1*, 27–44.

Medoff, D. (1999). MMPI-2 validity scales in child custody evaluations: Clinical versus statistical significance. *Behavioral Sciences and the Law*, *17*, 409–411.

Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (1997). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (2nd ed.). New York: Guilford Press.

Meyer, G. J. (2001). Evidence to correct misperceptions about Rorschach norms. *Clinical Psychology: Science and Practice*, *8*, 389–396.

Millon, T. (1997). *Manual for the Millon Clinical Multiaxial Inventory-Three (MCMI-III)* (2nd ed.). Minneapolis, MN: National Computer Systems.

Murray, H. (1943). *Thematic Apperception Test manual*. Cambridge, MA: Harvard University Press.

Ornduff, S. R., & Kelsey, R. M. (1996). Object relations of sexually and physically abused female children: A TAT analysis. *Journal of Personality Assessment*, *66*(1), 91–105.

Otto, R. K., Buffington-Vollum, J. K., & Edens, J. F. (2003). Child custody evaluation: Research and practice. In I. B. Weiner (Series Ed.) & A. Goldstein (Vol. Ed.), *Comprehensive handbook of psychology: Vol. 11. Forensic psychology* (pp. 179–208). New York: Wiley.

Otto, R. K., & Collins, R. P. (1995). Use of the MMPI-2/MMPI-A in child custody evaluations. In Y. Ben-Porath, J. R. Graham, G. C. Hall, R. D. Hirschman, & M. S. Zaragoza (Eds.), *Forensic applications of the MMPI-2* (pp. 222–252). Thousand Oaks, CA: Sage.

Otto, R. K., Edens, J. F., & Barcus, E. H. (2000). The use of psychological testing in child custody evaluations. *Family & Conciliation Courts Review*, *38*, 312–340.

Pistole, D. R., & Ornduff, S. R. (1994). TAT assessment of sexually abused girls: An analysis of manifest content. *Journal of Personality Assessment*, *63*, 211–222.

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.

Riethmiller, R. J., & Handler, L. (1997). The great figure drawing controversy: The integration of research and clinical practice. *Journal of Personality Assessment*, *69*, 488–496.

Rogers, R., & Ewing, C. (2003). The prohibition of ultimate opinions: A misguided enterprise. *Journal of Forensic Psychology Practice*, *3*, 65–75.

Rogers, R., Salekin, R., & Sewell, K. (1999). Validation of the Millon Clinical Mutiaxial Inventory for axis II disorders: Does it meet the *Daubert* standard? *Law and Human Behavior*, *23*, 425–443.

Rossini, E., & Moretti, R. (1997). Thematic Apperception Test (TAT) interpretation: Practice recommendations from a survey of clinical psychology doctoral programs accredited by the American Psychological Association. *Professional Psychology: Research and Practice*, *28*, 393–398.

Sattler, J. M. (1998). *Clinical and forensic interviewing of children and families: Guidelines for the mental health, education, pediatric, and child maltreatment fields*. La Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Scribner, C. M., & Handler, L. (1987). The interpreter's personality in draw-a-person interpretation: A study of interpersonal style. *Journal of Personality Assessment*, *51*, 112–122.

Shaffer, T. W., Erdberg, P., & Haroian, J. (1999). Current nonpatient data for the Rorschach, WAIS-R, and MMPI-2. *Journal of Personality Assessment*, *73*, 305–316.

Siegel, J. C. (1996). Traditional MMPI-2 validity indicators and initial presentation in custody evaluations. *American Journal of Forensic Psychology*, *14*, 55–63.

Skinner, L. J., & Berry, K. K. (1993). Anatomically detailed dolls and the evaluation of child sexual abuse allegations: Psychometric considerations. *Law and Human Behavior*, *17*, 399–421.

Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin*, *112*, 140–154.

Strong, D. R., Greene, R. L., Hoppe, C., Johnston, T., & Olesen, T. (1999). Taxometric analysis of impression management and self-deception on the MMPI-2 in child-custody litigants. *Journal of Personality Assessment*, *73*, 1–18.

Tillbrook, C., Mumley, D., & Grisso, T. (2003). Avoiding expert opinions on the ultimate legal question: The case for integrity. *Journal of Forensic Psychology Practice*, *3*(3), 77–87.

Tippins, T. M. (2005). Custody evaluations, part XI: "Daubert" and the rise of empiricism. *New York Law Journal*, 3–12.

Tippins, T. M., & Wittmann, J. P. (2005). Empirical and ethical problems with custody recommendations: A call for clinical humility and judicial vigilance. *Family Court Review*, *43*, 193–218.

Vitacco, M. J., & Rogers, R. (in press). Psychopathy and response styles in sex offenders. In F. Saleh, A. Grudzinskas, & J. Bradford (Eds.), *Sex offenders: Identification, risk assessment, treatment, and legal issues*. New York: Oxford University Press.

Waller, E. M., & Daniel, A. (2004). Purpose and utility of child custody evaluations: From the perspective of judges. *Journal of Psychiatry & Law*, *32*(1), 5–27.

Weiner, I. B. (2005). The utility of Rorschach assessment in clinical and forensic practice. *Independent Practitioner*, *25*(2), 76–83.

Westen, D. (1991). Clinical assessment of object relations using the TAT. *Journal of Personality Assessment*, *56*, 56–74.

Winter, D. G., & Stewart, A. J. (1977). Power motive reliability as a function of retest instructions. *Journal of Consulting and Clinical Psychology*, *45*, 436–440.

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

Wood, J. M., Nezworski, M. T., Garb, H., & Lilienfeld, S. (in press). The controversy over Exner's Comprehensive System for the Rorschach: The critics speak. *Independent Practitioner*.

Wood, J. M., Nezworski, M. T., Lilienfeld, S., & Garb, H. (2003). *What's wrong with the Rorschach? Science confronts the controversial Inkblot Test*. San Francisco: Jossey-Bass.

Wood, J. M., Nezworski, M. T., & Stejskal, W. J. (1996). The Comprehensive System for the Rorschach: A critical examination. *Psychological Science*, *7*(1), 3–10.

Dr. Erickson is the MIRECC fellow at Yale University, an attorney, and a law guardian for the 4th Judicial Department of New York State. He received his law degrees from SUNY Buffalo Law School and Harvard Law School and his Ph.D. from SUNY Buffalo. His research focuses on health law, criminal law, and public policy.

Scott O. Lilienfeld is Associate Professor of Psychology at Emory University in Atlanta, Georgia. He received his B.A. from Cornell University in 1982 and his Ph.D. from the University of Minnesota in Psychology (Clinical) in 1990. He was on the faculty of SUNY Albany from 1990 to 1994, when he moved to Emory, where he has been on the faculty ever since. He is editor-in-chief of the Scientific Review of Mental Health Practice, Associate Editor of Applied and Preventive Psychology, and sits on the editorial boards of numerous journals. His primary areas of interest are personality disorders, personality assessment, psychiatric classification and diagnosis, and evidence-based practices in clinical psychology.

Michael J. Vitacco is the Associate Director of Research at Mendota Mental Health Institute in Madison, WI. He received his Ph.D. from the University of North Texas and competed a forensic fellowship at the University of Massachusetts Medical School. His primary interest is in the area of forensic assessment with emphases in psychopathy, malingering, and violence risk prediction.