

**38 Fam. & Conciliation Courts Rev. 312**

Family and Conciliation Courts Review

July, 2000

Special Issue Child Custody Evaluations

Randy K. Otto<sup>a1</sup> John F. Edens<sup>a2</sup> Elizabeth H. Barcus<sup>a3</sup>

Copyright (c) 2000 by Sage Publications, Inc.; Randy K. Otto, John F. Edens, Elizabeth H. Barcus

**THE USE OF PSYCHOLOGICAL TESTING IN CHILD CUSTODY EVALUATIONS**

Evaluation of families for purposes of assisting the court in making decisions about custody is perhaps the most complicated forensic evaluation. Mental health professionals conducting such evaluations must ensure that their evaluations validly assess areas of concern deemed relevant by the judiciary and legislature. Evaluators sometimes use psychological measures in the evaluation process, and in recent years, a number of tests designed specifically for use in child custody evaluation contexts have been developed. Because some published tests do not meet basic professional standards, child custody evaluators should carefully review any test and its supporting documentation before including it in their examination procedures. In this article, the authors discuss the rationale for using psychological tests in child custody evaluations, describe current testing practices, review and critique contemporary custody evaluation instruments, and offer a template for mental health professionals to use when considering use of a particular test.

Cases of disputed child custody may be among the most challenging for attorneys and judges (Settle & Lowery, 1982). Because decisions regarding custody typically involve issues revolving around child development, parenting behavior, family systems, psychopathology, and emotional well-being, courts sometimes request the input of mental health professionals. This is based on the assumption that mental health professionals will be able to provide valuable information about the children and parents, which, in turn, will form the basis for a more informed and better decision on the part of the legal decision maker.<sup>1</sup>

Child custody evaluation may be the most complex and difficult type of forensic evaluation. In contrast to most examinations in which one person is evaluated, in the typical child custody evaluation, the mental health professional examines a number of persons (e.g., mother, father, child or children, and potential or actual stepparents). Additionally, given the expansive nature of the underlying psycholegal issues (i.e., the best interests of the children and the ability of the parents to meet those interests), the examinees must be assessed regarding a variety of behaviors, capacities, and needs. Finally, because the stakes are so significant (i.e., residential placement of the children \*313 and decision-making authority with respect to their welfare), emotions in cases of contested custody typically run high, further compounding what is an already complicated evaluation process.

Given the potential level of complexity involved in custody evaluations, it is perhaps not surprising that mental health professionals have attempted to develop instruments to assess those criteria that are putative predictors of children's postdivorce adjustment (e.g., parental emotional stability and degree of a child's attachment to the parents). The recent development of custody-specific tests represents perhaps the most overt attempt to operationalize, quantify, and assess those psychological factors that often are judged to be the most relevant in the judicial decision-making process. In this article, we discuss the rationale for using psychological tests in child custody evaluations, describe current test practices of psychologists and others who conduct child custody evaluations for the courts, review and critique contemporary

instruments designed specifically for use in child custody evaluations, and finally, offer a template for mental health professionals to use when they are considering employing a particular test in their evaluations.

## PSYCHOLOGICAL TESTING IN CHILD CUSTODY EVALUATION

### CURRENT TESTING PRACTICES AND LIMITATIONS

Brodzinsky (1993) offered a number of reasons why psychologists and other mental health professionals skilled in psychological testing might include traditional and more contemporary child custody assessment instruments in their evaluations. Particularly for psychologists, testing is associated with their professional identity and is part of their usual and customary practice. Thus, it should come as no surprise that they would use standard psychological tests (e.g., measures of intelligence, achievement, personality, or psychopathology) when conducting child custody evaluations, particularly if they are unfamiliar with the specific decision-making criteria used by the judiciary. Brodzinsky also noted that psychological tests lend an air of objectivity, science, and insight to evaluations (whether deserved or not), which may make them popular with attorneys and judges who may be more responsive to test-based or test-anchored evaluations and opinions. Finally, Brodzinsky noted that test use may be driven by financial incentives, given the amount of time it can take to administer, score, and interpret a particular battery of tests given to parents, children, and others.

**\*314** Although the use of standardized psychometric measures is common, the testing practices of child custody evaluators have been subject to considerable criticism. Brodzinsky (1993) and Bricklin (1992, 1999) noted that the traditional psychological tests used by many child custody evaluators do not address psycholegal issues directly relevant to the child custody question (e.g., parenting ability, the nature and quality of the parent-child relationship, and the willingness of each parent to facilitate a close relationship with the other parent). Use of general psychological measures requires the evaluator, at a minimum, to make an inference from the general construct assessed by the instrument (e.g., psychopathology or intelligence) to a more specific and relevant behavior (e.g., ability to meet the child's emotional and behavioral needs). Grisso (1984) offered a similar critique:

Too often we rely on assessment instruments and methods that were designed to address *clinical* questions, questions of psychiatric diagnosis, when clinical questions bear only secondarily on real issues in many child custody cases. Psychiatric interviews, Rorschachs, and MMPIs might have a role to play in child custody assessment. But these tools were not designed to assess parents' relationships to children, nor to assess parents' child-rearing attitudes and capacities, and *these* are often the central questions in child custody cases. (Cited in Melton, Petrila, Poythress, & Slobogin, 1997, p. 484)

These general concerns are partially reflected in the child custody guidelines promulgated by the Association of Family and Conciliation Courts (AFCC) (n.d.), the American Psychological Association (APA) Committee on Professional Practice Standards (1994), and the American Academy of Child and Adolescent Psychiatry (AACAP) (1994), all of which caution that child custody evaluations should be informed by legal criteria, should be expansive in scope, and should not be simply psychopathology—focused.

Although the majority of mental health professionals conducting child custody evaluations use standard psychological tests, specific custody tests have been developed in the past decade, and they are being used with increasing frequency. Such tests are appealing to mental health professionals and the judiciary because they ostensibly address the specific questions involved in forming an opinion in a custody case, such as, “Does the parent have adequate parenting skills?” or “With which parent is the child most bonded?” Such questions are not easily answered by making inferences from the results of standard measures of psychopathology, intelligence, and personality.

**\*315 SURVEYS OF TESTING PRACTICES**

Despite the apparent widespread use of psychological tests, relatively little data exist in terms of identifying what instruments tend to be employed as part of child custody evaluations. Three surveys have been published that examine child custody evaluators' use of psychological tests, but how accurately these results depict current practice is unclear, so their value is somewhat limited. One limitation is that all of the surveys had a relatively small number of respondents. A second limitation is that psychologists were overrepresented in the studies, and as a result, we know much less about test use by non-psychologist practitioners who conduct child custody evaluations. This, of course, is important, as it seems reasonable to assume that psychologists, given their background and training, are more likely than other mental health professionals (e.g., social workers, psychiatrists, or marriage and family therapists) to use psychological testing in their custody evaluations. Thus, the studies described below may overestimate the use and significance of psychological tests in custody evaluations.

In the earliest published study, Keilin and Bloom (1986) surveyed a national sample of psychologists, psychiatrists, and master's-level practitioners identified through various national and local directories of forensic experts and custody evaluators. A return rate of 63% (190 out of 302) was obtained, although only 82 of the returned surveys provided usable data. Of those persons who responded, 78% were doctoral-level psychologists, 18% were psychiatrists, 3% were master's-level practitioners, and 1% were social workers. No single measure was used by a majority of the respondents when assessing children. Intelligence tests were the tests most frequently employed by the examiners, with 45% of respondents using some measure of intelligence in the majority (85%) of their cases. The next most frequently used measure was the Thematic Apperception Test (TAT) or the Children's Apperception Test (CAT), with 39% of the respondents using such measures in most (75%) of their evaluations. The three next most commonly used tests used with children were miscellaneous projective drawings, the Rorschach Inkblot Technique, and the Bender-Gestalt Visual Motor Test.

With respect to assessment of adults, the most commonly used test was the Minnesota Multiphasic Personality Inventory (MMPI). Seventy percent of the respondents reported using this instrument in child custody evaluations, and those who used it employed it in almost all (88%) of their cases. The next most frequently used instruments were the Rorschach Inkblot Technique (42%) and the TAT (38%), and evaluators who employed these instruments \*316 reported using them in a majority of their cases. Measures of adult intelligence also were occasionally employed, with the Wechsler Adult Intelligence Scale (WAIS) being used by 29% of the respondents. Those who used the WAIS reported employing it in a majority (67%) of their cases.

Ackerman and Ackerman (1997) replicated the Keilin and Bloom (1986) survey to obtain a more current picture of the types of psychological tests employed in custody evaluations. In the decade between these two surveys, a number of new or revised standard psychological tests had been developed and marketed, as well as several instruments specifically designed for application in cases of child custody. The investigators mailed 800 questionnaires to psychologists identified by various psychological and legal associations. Of these, 336 were returned, but only 201 (25%) fit the study's selection criteria. In contrast to the Keilin and Bloom survey, all respondents were doctoral-level psychologists.

Similar to the findings of Keilin and Bloom (1986), intelligence tests and projective measures continued to be the instruments most frequently used with children. Fifty-eight percent of the respondents reported using intelligence tests in their evaluations. Those using such tests indicated employing them in 45% of their evaluations. Thirty-seven percent used either the CAT or the TAT (in 53% of their evaluations). With respect to assessment of adults, the MMPI/MMPI-2 remained the most frequently used assessment instrument, with 92% of the respondents employing a version of this test in the large majority (91%) of their evaluations. The Rorschach Inkblot Technique remained the second most frequently used test with adults—48% of the respondents reported employing the test, and those who used it did so in more than half (64%) of their cases. The next most frequently used tests were the revised WAIS and the Million Clinical Multiaxial Inventory (MCMI-II/MCMI-III), with 43% and 34% of the examiners reporting using these tests in their custody evaluations, respectively.

In terms of the more recently developed tests specifically designed to assess children during custody evaluations, more than one third of the respondents (35%) reported using the Bricklin Perceptual Scales (BPS) (Bricklin, 1990a). Furthermore, those who used the BPS did so in the majority (66%) of their cases. Another child custody assessment instrument, the Perception of Relationships Test (PORT) (Bricklin, 1989), was also used by a significant proportion of the respondents (16%) in a large number of their cases (64%).

Few of the respondents reported using specific custody measures designed for use with families or adults. Eleven percent of the respondents reported using the Ackerman-Schoendorf Scales for Parent Evaluation of Custody (ASPECT) (Ackerman & Ackerman, 1992), but those who used it \*317 did so in essentially all (89%) cases. The only other custody-specific measures endorsed were the Parent Awareness Skills Survey (PASS) (Bricklin, 1990b), used by 8% of the respondents (who employed it in 94% of their cases), and the Custody Quotient (CQ) (Gordon & Peek, 1989), used by 4% of the respondents (in 57% of their cases).

LaFortune and Carpenter (1998) recently surveyed mental health professionals about the tests and strategies they employed in their custody evaluations. They solicited data from 286 participants identified through a variety of sources (e.g., state agencies and telephone advertisements) and obtained a geographically diverse sample of 165 respondents. In the sample, 89% of respondents were psychologists, 6% licensed professional counselors, 1% licensed marriage and family therapists, and 3% licensed social workers (1% of the sample was unidentified).

Respondents reported the frequency of use of various assessment methods on a 1 (*never*) to 5 (*always*) scale. Regarding psychological tests used to assess adults, parenting scales, such as the ASPECT and the Bricklin measures,<sup>2</sup> were second in frequency of use (mean response level of 3.28) only to the MMPI (mean response level of 4.19). Unfortunately, frequency of use for individual custody tests was not reported by the authors. Nevertheless, it appears that these newer, more specific instruments enjoyed a fairly significant rate of use among these respondents. Data regarding instruments used to assess children apparently were not collected, so it is unclear whether a similar high rate of use would have been found.

Findings from the above surveys indicate that, although traditional psychological tests continue to be employed with a fair degree of frequency by custody evaluators, the new generation of custody assessment measures also are used with some regularity. Although these tests have good face validity (i.e., their item content makes sense and appears to assess factors relevant to child custody decision making), significant questions remain regarding their utility and validity, and their appropriateness for use in custody evaluations at the present time.

## **CRITERIA FOR TEST USE**

Exactly what criteria should be considered in determining whether any given test is an appropriate assessment procedure in a child custody evaluation? Grisso's (1986) model for conducting forensic evaluations provides a helpful framework for considering the potential utility of psychological testing in the context of custody evaluations. Evaluators considering including a particular test first should identify the psychological constructs that they are to assess. These constructs can be specified by examining the law regarding \*318 child custody and identifying the psychological, emotional, and behavioral constructs on which the law is focused. Next, evaluators should select methods of assessment, including potential tests, that ostensibly measure the specific construct of interest or related constructs. Because tests that are valid and appropriate in one context may be poorly suited for other applications or in other contexts, decisions about the value or appropriateness of a test cannot be made independent of its proposed application. Therefore, a test that reliably and validly assesses a particular construct (e.g., psychopathology) in one setting (e.g., psychiatric hospitals) may not be especially useful in another type of setting (e.g., custody evaluations).

Next, examiners should consider whether the particular test meets professional standards for psychological tests and measures in general (APA, 1992; American Educational Research Association, 1999) as well as standards specific to forensic evaluation (Heilbrun, 1992). If a proffered test meets the above requirements then its use in the context of a child custody evaluation can be considered appropriate. Reviewed below are those standards that mental health professionals should consider when evaluating the appropriateness of a particular test for use in a child custody evaluation.

### **Ethical Principles of Psychologists and Code of Conduct**

Section 2 of the Ethical Principles of Psychologists and Code of Conduct (APA, 1992) directs professional practice with respect to psychological evaluation and assessment and clearly applies to all psychologists.<sup>3</sup> These standards specify that psychologists must (a) be familiar with an instrument's reliability, validity, norms, and administration; (b) use assessment techniques in ways that are appropriate in light of research or other evidence; and (c) be aware of cases in which particular assessment techniques or norms may not be applicable or may require adjustment in administration or interpretation. This latter standard seems particularly important given concerns about the possible misuse of instruments that primarily assess issues of concern within a clinical context (i.e., intelligence, psychopathology, or personality).

### **Standards for Educational and Psychological Testing**

The *Standards for Educational and Psychological Testing* (American Educational Research Association, 1999) represents an interdisciplinary attempt to establish principles for the development and use of psychological and educational tests. Arguably, these standards apply to all test users, regardless of discipline. The *Standards for Educational and Psychological Testing* is comprehensive and addresses issues such as technical standards for \*319 test development and validation, standards for professional practice and use, standards for test use with special populations (e.g., minority individuals and individuals with disabilities), and other issues such as test administration, test scoring, reporting results, the rights of test takers, and the responsibilities of test users.

### **Forensic Practice Guidelines**

In addition to the general guidelines described above, there are a number of more specific guidelines that are relevant for consideration of test inclusion. Custody evaluation guidelines promulgated by the AFCC (n.d.), the APA Committee on Professional Practice Standards (1994), and the AACAP (1994) all make reference to psychological testing and its use in the context of child custody evaluations. Consistent with the more specific testing standards and guidelines described above, both the AFCC and APA standards urge evaluators to use psychological tests only as appropriate; the organizations also note the tests' limitations in the context of child custody disputes.

In contrast, the AACAP (1994) standards direct that psychological testing is usually not helpful in the context of child custody evaluations and typically results in “professionals battling over the meaning of raw data” (p. 65S). That the AACAP guidelines identify psychological test results as the only kind of data gathered in the context of a child custody evaluation that is vulnerable to a “battle of the experts” suggests that this commentary derives more from guild issues than a legitimate concern over the potential misuse of test data.

Heilbrun (1992) discussed the role of psychological testing in forensic assessment and offered guidelines for selecting tests to use in such applications. Those concerns that are applicable to consideration of tests in child custody matters are discussed below. First, Heilbrun suggested that any test forming the basis of an opinion in a forensic evaluation should be commercially available; accompanied by a manual describing its development, psychometric properties, and administration procedures; and listed or reviewed in a readily available resource or reference such as the *Mental Measurements Yearbook*. With the above, all parties will have access to information regarding the test and its utility and can prepare appropriately for direct examination and cross-examination of mental health professionals who use the

instruments. Second, because test reliability limits test validity, Heilbrun recommended that adequate levels of reliability be demonstrated, with special caution being exercised when using instruments with relevant reliability estimates lower than .80. Third, it is suggested that the test should be relevant to the legal issue at hand and that, when possible, this should be established by validation research published in refereed journals. Fourth, to ensure generalizability \*320 from the testing situations, Heilbrun noted that instruments should be administered in a standardized manner. Fifth, the examiner should consider differences between the examinee and the population with which a test was developed and/or differences between the purposes for which a test was developed and the purposes for which it is being used. Finally, Heilbrun directed that, where appropriate, the effect of response styles (e.g., “faking good,” malingering, and social desirability) on test results should be considered by the examiner, with interpretations being offered accordingly.

Together, the above authorities provide reasonable guidelines for evaluating the utility of tests that may be used in the context of child custody proceedings. In the next section, we review instruments that have been developed specifically for the purposes of child custody evaluation. A brief description of the instrument and manual is provided first, followed by an evaluation of its strengths and weaknesses and a discussion of the extent to which the test presently meets professional standards.<sup>4</sup>

### CHILD CUSTODY EVALUATION INSTRUMENTS

There are a myriad of tests that are potentially useful in the context of a child custody evaluation. Tests that might be useful in child custody evaluation fall into one of three broad categories: clinical assessment instruments, forensically relevant instruments, and forensic assessment instruments (for further discussion and explanation, see Heilbrun, Rogers, & Otto, 2000). Clinical assessment instruments assess general psychological constructs (e.g., psychopathology, intelligence, academic achievement, normal personality), are developed for therapeutic applications, and most typically, are used in non—forensic settings. Nonetheless, to the degree that these instruments assess general constructs that may be relevant to decisions revolving around child custody, their use in child custody evaluations is appropriate. For example, in those jurisdictions where emotional stability of the parties is one factor to be considered in making decisions about the custody and placement of children, use of the MMPI-2 as one way of assessing psychopathology and emotional stability as it may be related to parenting is appropriate. Similarly, if an examiner uses the Child Behavior Checklist in her or his assessment of the child's current adjustment and response to the divorce, and as a way of understanding the child's current needs, then this too would be appropriate use of a general, clinical assessment instrument in the context of a forensic (child custody) evaluation. As a final example, if a psychologist uses the Parenting Stress Index to assess how interactions with the child affect the parent, this too would appear to be appropriate use of a validated clinical assessment instrument in the context of child custody evaluation.

\*321 Forensically relevant instruments are assessment techniques that assess constructs or issues that most typically arise in the course of forensic evaluations, but they are not limited to forensic evaluation. Tests of defensiveness, malingering, and psychopathy (e.g., Paulhus Deception Scales, Structured Interview of Reported Symptoms, and Psychopathy Checklist—Revised) are examples of such instruments. Perhaps with the exception of measures of general defensiveness in responding (for an example, see Paulhus, 1999), forensically relevant instruments are unlikely to prove helpful in the large majority of child custody evaluations.

Forensic assessment instruments are developed specifically for application in forensic settings, and their purpose is to assess constructs relevant to particular legal issues. At the current time, there are five child custody evaluation instruments commercially published, the majority of which have been developed by Barry Bricklin, a Pennsylvania psychologist.

Rogers and Webster (1989) observed that, in many forensic evaluation situations, the best validated tests and assessment instruments are general clinical tests, which have the least direct legal relevance (i.e., the constructs they assess are not directly related to the legal issue at hand). This observation also applies in the child custody evaluation context, where the general clinical assessment instruments that are used typically have better validity data than child custody

evaluation measures, but the constructs they assess (e.g., psychopathology, intelligence, academic achievement, or normal personality) are not directly legally relevant (although they may be relevant nonetheless). The review that follows is limited to forensic assessment instruments devoted to child custody evaluation. That the utility of general assessment techniques in the context of child custody evaluation is not addressed here should not be interpreted as indicating that the instruments are of no value in child custody evaluation. Rather, such a task is more deserving of a book than an article.

As the following review demonstrates, although test developers have made some important progress in the assessment of custody-relevant psychological constructs, the existing instruments suffer from a number of limitations. This review, however, is not intended to minimize the important contributions test developers have made; rather, it is designed to alert custody evaluators and consumers to the limitations of some of these instruments and the need for further research.

### Perception of Relationships Test

The PORT (Bricklin, 1989) is a projective or unstructured psychological test based on human figure drawings and their placement. It purportedly \*322 assesses (a) the degree to which a child seeks “psychological closeness” with each parent and (b) the types of interactions the child has with each parent. The PORT consists of seven drawing tasks and is identified as being appropriate for use with children age 3 and older. Children's drawings of themselves, their parents, and their families are scored so that a primary caretaking parent can be identified. Bricklin notes that the PORT also has the potential to be a useful tool in detecting physical and sexual abuse insofar as it can detect “psychological consequences of abuse” (p. 44). Scoring guidelines, which are very complex for some items, are provided in the PORT manual. Some evidence supporting the reliability and validity of the PORT also is provided in the manual.

There are several problems with the PORT that raise significant concerns regarding its use at the present time. First, the PORT manual is poorly organized and poorly written. A description of the PORT, scoring procedures, and supporting data are interspersed with the author's commentary on general child custody evaluation issues. Descriptions of research protocols are sometimes vague and incomplete, and information regarding the PORT's psychometric properties is presented both in the PORT manual (Bricklin, 1989) and in a photocopied paper (Bricklin & Elliott, 1997). Thus, the clinician interested in quickly accessing important information such as the test's reliability and validity data will be frustrated.

Bricklin (1989) (Bricklin & Elliott, 1997) did not report any data regarding item scoring and interrater reliability (i.e., the extent to which different examiners will obtain similar scores when administering the test to an examinee). Thus, important questions remain regarding how consistently examiners will score and interpret test results. This issue appears all the more significant in light of the amount of attention Bricklin devotes in the manual to projective interpretation of drawing techniques. No studies examining the reliability or validity of the PORT have been published in peer-reviewed journals. Bricklin (1989) prefaced a brief section on test-retest reliability (i.e., the extent to which an examinee would obtain a similar score across repeated administrations of the test) with a discussion minimizing its significance: “There are no real reasons to expect the measurements reported here to exhibit any particular degree of stability, since they should vary in accordance with changes in the child's perceptions” (p. 64). Despite this assertion, one would hope that, at a minimum, test results would remain approximately the same over at least brief test-retest periods. Also, reliability across different testing conditions (e.g., which parent brings the child to the evaluation) would seem to be crucial to determining the validity of the interpretations offered. Yet only minimal data regarding such issues are provided. Data regarding test-retest reliability are limited to a total of 18 cases in which children \*323 were tested twice with the PORT, with periods between testing ranging from 4 to 8 months. Bricklin reported that, although specific scores changed in some cases, in only 2 of the 18 cases did the “chosen parent” change.

Although validity data for the PORT are somewhat better, they leave much to be desired. Most of Bricklin's data address the convergent validity<sup>5</sup> of the PORT. Bricklin (1989) reported findings of one study in which the primary caretaking parent identified by the PORT was identical to the parent of choice identified by the Bricklin Perceptual Scales (Bricklin,

1990a; see below) in 19 of 23 (83%) cases. Also described is a study in which clinicians made judgments about 30 children and their parents, based on extensive clinical data. In 28 of the 30 cases, the PORT-identified primary caretaking parent was consistent with the clinicians' decisions regarding who would be the better primary caretaker. On pages 61 and 62 of the PORT manual, Bricklin (1989) appears to describe a study in which PORT scores and judgments of clinicians who viewed 30 parent-child interactions were compared. According to Bricklin, the PORT primary caretaking parent was in agreement with the clinicians' opinions regarding parent of choice "with better than 90 percent accuracy." Bricklin described a second set of studies in which he compared PORT scores to judges' decisions regarding custody. According to the author, in 80 of 87 cases (92%), the judges granted custody to the parent identified as the primary caretaking parent by the PORT. In a later published supplement (Bricklin & Elliott, 1997), references are made to a number of other studies, but the descriptions of each study are so brief (i.e., one to two lines per study) as to prevent adequate consideration of them. Of these studies, perhaps most interesting is the description of a sample of 1,038 cases in which an agreement rate of 89% was obtained when PORT primary caretaking parents were compared to the judgments of "independent psychologists based on all clinical and life-history data available." These findings, however, essentially appear to be psychologists reporting to Bricklin how frequently the PORT results were consistent with their clinical impressions in custody cases they evaluated.

Although early reviews of the PORT described it as a promising addition to the field (Brodzinsky, 1993), more recent reviews (Carlson, 1995; Conger, 1995; Heinze & Grisso, 1996; Melton et al., 1997) consistently have pointed out the limitations noted above as well as several others that prevent it from meeting basic scientific and legal standards. Problems identified can be classified as relating to (a) a poorly articulated theoretical foundation, (b) a non-standardized administration format, (c) the absence of normative data, (d) subjective scoring procedures, (e) the absence of reliability data, (f) minimal validity data, and (g) failure to assess parents' functional abilities in the process of making custody determinations and recommendations.

**\*324** First, the basic premise on which the PORT is based, that children's unconscious preferences are more important than their verbalized statements, is not testable via empirical analysis using falsifiable hypotheses. There simply are no data suggesting that this assertion is true or, even if true, that the PORT adequately assesses this unconscious preference. Such a limitation calls into question the relevance of the PORT to custody decision making at all. Second, the PORT allows for considerable deviation in how it is administered. Such deviations, even for a test that otherwise has adequate psychometric properties, may invalidate any data that could have been obtained via standardized procedures. Third, the sample reported in the PORT manual is inadequately described and is too small. There are no normative data available for the PORT, and differences across age groups are not reported. Fourth, the criteria by which some of the PORT items are to be scored are subjective. This leads to significant concerns that different examiners would not obtain similar results when testing the same child (i.e., poor or unknown interrater reliability). Fifth, Bricklin's description as to why traditional reliability estimates (e.g., test-retest) are not applicable to the PORT is simply inaccurate. If a child's unconscious ratings of the preferred parent are so inconsistent as to render them unstable over even brief periods of time, then the entire premise on which the PORT is based (i.e., assessing the preferred parent) is essentially irrelevant to making any decision regarding a child's long-term placement.

Sixth, the validity data provided in the PORT manual are inadequate. Data indicating that the PORT results were consistent with judicial decisions are problematic for at least two reasons. First, it is not clear whether the PORT results influenced the decisions being made. If so, the obtained percentage of agreement suffers from criterion contamination (i.e., the outcome being predicted was actually influenced by the prediction being made, therefore artificially elevating the rate of agreement). A second problem relates to the criterion itself. If the utility of the PORT rests in its ability to predict what a judge ultimately will decide independently, what is the purpose of administering the test at all? A much more relevant outcome to predict would be whether the results of the PORT actually identify the parent truly better suited to serve as the custodian. Unfortunately, no such data have been published, if they exist.



In summary, the PORT has significant limitations in almost every area that is relevant to the development of a useful test. The absence of an adequate manual, normative data, and any published studies examining its psychometric properties suggests that the PORT does not meet basic requirements established by the latest version of the *Standards for Educational and Psychological Testing* or guidelines for use of psychological tests in forensic settings (Heilbrun, 1992).

### \*325 Bricklin Perceptual Scales

The BPS (Bricklin, 1990a) is purported to measure a child's perception of each parent in the areas of competence, supportiveness, follow-up consistency, and possession of admirable traits. Using a card and stylus, the child indicates how well each parent completes a particular behavior (e.g., being patient) or task (e.g., helping with schoolwork) by puncturing a line along a continuum from *not so well to very well*; children's responses also are obtained verbally. Each parent is rated on the same 32 items, so the child completes a total of 64 items that are distributed over a total of 4 scales (see above). The parent who receives more positive ratings on the greater number of cards is considered the parent of choice. The underlying assumption is that children's verbal reports about their parents may be distorted and that a non-verbal assessment strategy will provide a better estimate of their "true" preferences.

Many of the limitations of the PORT noted earlier also apply to the BPS and raise significant concerns regarding its use in custody decision making. First, no studies examining the utility of the BPS have been published in peer-reviewed journals, and the information provided in the test manual is inadequate regarding several important issues, particularly reliability and validity. For example, the BPS manual provides no data regarding interrater reliability. It appears likely, however, that this form of reliability would be high given the straightforward manner in which the items are scored. Similarly, scale consistency (i.e., the extent to which individual items on a scale reliably measure the same psychological construct) is not addressed in the BPS manual. Given the small number of items on some scales (e.g., the follow-up consistency scale has only three items, and the admirable traits scale has seven items), it is likely that scale consistency will be low (scales with few items tend to have relatively low internal consistency), thus limiting scale validity.

Similar to the rationale provided for the PORT, Bricklin (1990a) prefaced his discussion of the data regarding test-retest reliability of the BPS by down-playing its significance. "There are no reasons to expect the measurements reported here to exhibit any particular degree of stability, since they should vary in accordance with changes in the child's perceptions" (p. 42). Although it is true that less-than-perfect test-retest reliability is not problematic if a test measures a construct that changes over time, at a minimum it should be demonstrated that test results are stable over relatively brief periods of time. Data regarding test-retest reliability are presented in a confusing manner (see p. 42). It appears, however, that in 12 custody cases in which testing was repeated within a period of no more than 7 months, the parent of choice identified by the BPS remained the same.

\*326 Validity data for the BPS also are presented in a confusing manner in the manual, with additional data provided in a photocopied publication that serves as a manual supplement for a number of Bricklin measures (Bricklin & Elliott, 1997). Validity data fall into one of three types: agreement between BPS scores and other measures of parent-child involvement, agreement between BPS scores and judges' decisions, and examiners' opinions regarding the accuracy and utility of the BPS.

Bricklin (1990a) cited two studies that revealed agreement between the BPS and the PORT as evidence of the validity of each test (also see above). In 46 of 55 cases (84%), the BPS and the PORT identified the same parent of choice (with 50% agreement representing chance). As further evidence of the validity of the BPS, Bricklin cited a number of studies that examined the relationship between BPS scores and perceptions of parenting behavior. For example, Bricklin administered the BPS and a set of questionnaires to 23 children and their parents. The questionnaires queried the children and parents regarding the parents' child-rearing behaviors and responsibilities. Parents who were identified by the children as taking on more child-rearing responsibilities were identified as the parent of choice in a manner similar to that used with the PORT. In 21 of 23 cases (91%), the same parents of choice were identified by the children's

responses to the questionnaires and the BPS. Of some interest is Bricklin's (1990a) claim that this high (but less-than-perfect) rate of agreement provides even greater support for the validity of the BPS than a higher rate of agreement might have demonstrated. He argues that children's responses to the BPS, because they are nonverbal, represent less conscious feelings about their parents, whereas their responses to questionnaires represent more conscious feelings. Thus, perfect agreement between the two measures (one that assesses conscious feelings and one that assesses unconscious feelings) would prove problematic.

Reviewers have consistently criticized the unsubstantiated claim that nonverbal responses are more valid indicators of preference, noting that reliance on purportedly unconscious indicators makes empirical validation of the BPS difficult if not impossible to obtain (Heinze & Grisso, 1996; Melton et al., 1997; Shaffer, 1992). With the same sample, Bricklin (1990a) also examined the rates of agreement between the parents of choice as identified by the children's BPS responses and the parents' perceptions of their involvement with their children. In 13 of 17 (76%) cases, there was agreement between each parent's perception and the parent of choice identified by the BPS. Finally, Bricklin described 29 cases of contested custody in which he conducted evaluations using the BPS. In 27 of these cases (93%), the parent identified as the parent of choice by the BPS was awarded custody by the \*327 judge. It is not reported whether Bricklin offered an ultimate opinion on custody that was consistent with BPS scores.

In a photocopied test manual supplement (Bricklin & Elliott, 1997), a number of other studies examining the validity of the BPS are described, but the descriptions of the studies are so sketchy (i.e., one to two lines per study) as to prevent adequate consideration of them. Of particular interest is the authors' report that in a sample of 1,765 cases, the BPS parent of choice was consistent with the judgments of psychologists who had access to "clinical and life history data." Similar to the large sample reported in the PORT section above, these findings appear to essentially be psychologists reporting to Bricklin how frequently the BPS results were consistent with their clinical impressions in custody cases they evaluated, with no appreciation for the limited value of this approach or the problem of criterion contamination.

Published reviews of the BPS (Hagin, 1992; Heinze & Grisso, 1996; Melton et al., 1997; Shaffer, 1992) have noted many of the problems identified above as well as other limitations indicating that use of the BPS is inappropriate at this time. For example, the instructions for administration and scoring are confusing and suggest that a standardized administration format may be abandoned under various circumstances (Heinze & Grisso, 1996; Shaffer, 1992). Also, the absence of normative data is especially problematic, particularly given the likelihood of responses being affected by changes in developmental capacity (Heinze & Grisso, 1996). Also, the language may be too complicated for some children to understand, particularly those with limited cognitive abilities. Another limitation is that the BPS may be biased in favor of mothers, in that in one unpublished study of the BPS (Speth, 1993, cited in Heinze & Grisso, 1996) fathers scored significantly lower than mothers on the supportiveness subscale. Also, gender differences in response patterns were found in this study, with adolescent boys endorsing fewer items in favor of fathers than adolescent girls. In summary, the criticisms noted above raise serious concerns about the BPS. Conservative custody evaluators should consider refraining from using this instrument until better supporting data become available.

### **Parent Perception of Child Profile**

The Parent Perception of Child Profile (PPCP) (Bricklin & Elliott, 1991) is an instrument designed to assess parents' understanding and awareness of a child's development and needs in eight basic areas: interpersonal relations, daily routine, health history, developmental history, school history, fears, personal hygiene, and communication style. The instrument queries parents \*328 about their knowledge and understanding of the child in these important life areas. The rationale behind the PPCP is that a parent who is more knowledgeable about a child's needs and development will be able to do a better job of parenting than one who is less knowledgeable. According to the test authors, the PPCP can be administered by the examiner in an interview format, or it can be self-administered by the parent. The manual directs that data need not be gathered in all eight categories, and the examiner can decide which issues are most critical for a particular child and parent.

It would appear that the PPCP requires that the parents' perceptions of the child be compared to some standard or criterion if the evaluator is to draw a conclusion about the accuracy of the parents' impressions. However, the authors note that the need for this corroborative information will vary from case to case, depending on an examiner's concerns about the parents' degree of interest in the child. In those cases in which corroborative information is considered necessary, the manual directs that "other source information" can be provided by the child or various significant others (e.g., babysitters and teachers). Validity of the parents' responses is also partially assessed in a recall format, with some questions being repeated later in the interview process.

Multiple limitations of the PPCP are apparent and preclude its recommendation for use at present. First, the manual is lacking and incomplete. Scoring directions, reliability data, or validity data are not provided in the PPCP manual (Bricklin & Elliott, 1991), and the summary score sheet is confusing. Second, there are no studies examining the utility of the PPCP published in peer-reviewed journals. The only data provided in the test manual are concerned with the recall format designed to identify parents who may respond inconsistently as a result of simply guessing about their children's behaviors.

Two reviews of the PPCP (Hiltonsmith, 1995; Kelley, 1995) have noted that the PPCP may "elicit valuable information regarding parents' knowledge of their children" (Hiltonsmith, 1995, p. 737). However, both also point out significant limitations of this instrument and warn against considering it to be a psychometric test rather than a semistructured interview. The primary limitation, noted above, relates to the absence of any type of reliability (internal consistency, test-retest, or interrater) or validity (content, construct, or predictive) data in the PPCP manual or peer-reviewed publications. The lack of such information obviously prohibits any meaningful conclusions from being made about scores obtained from the PPCP.

Aside from the absence of data regarding basic psychometric properties, reviewers have noted significant concerns regarding the administration and scoring of the PPCP. First, the manual directions encourage examiners to include or omit items on a case-by-case basis, thereby resulting in a nonstandardized process of data collection. Second, the PPCP may be either evaluator- \*329 administered or self-administered, which raises concerns as to whether results would be comparable across these different administration formats. Third, item scoring and interpretation rely on considerable clinical judgment.

In summary, although the PPCP appears to be a face-valid instrument that reflects the clinical intuition of its developers, its utility as a psychometric instrument (as opposed to a semistructured clinical interview) is presently unknown (Kelley, 1995). Because the PPCP does not meet basic requirements of the *Standards for Educational and Psychological Testing*, it is recommended that examiners not use it. Those who find the general approach of the PPCP intriguing might consider using it to form the basis of a structured interview but should refrain from scoring it, given the above limitations.

### **Parent Awareness Skills Survey**

The PASS is described as a "clinical tool designed to illuminate the strengths and weaknesses in awareness skills a parent accesses in reaction to typical child care situations" (Bricklin, 1990b, p.4). The PASS consists of 18 typical child care situations or dilemmas and represents a sampling of relevant parenting behaviors that can be applied to children of various ages. It appears to be rooted in the commonsense notion that strengths and weaknesses in parents' child-rearing abilities can be assessed, in part, by querying parents about how they would respond to various child care scenarios. After being presented with a particular child-rearing situation, the parent offers his or her response, with follow-up questioning by the examiner as necessary. Responses are then evaluated and scored.

No data regarding norms, test reliability, or test validity are offered in the manual, and none appear in any peer-reviewed publications. Although scoring directions and guidelines are offered in the PASS manual, the authors note 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. The PASS allows for wide latitude in scoring since its main purpose is to discover the relative (rather than absolute) strengths and weaknesses any individual or compared set of respondents manifest. (Bricklin, 1990b, p. 11)

This description seems to suggest that little can be inferred from PASS results in terms of absolute abilities. However, the author goes on to claim that the PASS can be administered to one parent, with scores having interpretive significance (Bricklin, 1990b, p. 16). The lack of clear scoring guidelines, the author's suggestion that not all items be administered in a particular case, and \*330 the lack of data regarding interrater reliability raise concerns about the validity of inferences derived from responses to the PASS items.

Of additional concern is Bricklin's (1990b) suggestion that the PASS can be used in a psychoeducational manner.

The PASS can be used to strengthen parental communication skills by simply allowing an interested parent to read this manual. In having the opportunity to think about sample responses offered by members of the Bricklin Associates research team, the parent can expand his or her available options. (p. 6)

Although whether the test authors recommend doing this in a therapeutic or forensic context is unclear, certainly the utility of any information gleaned from the instrument for purposes of forensic evaluation would be even further compromised if it were to be collected after an examinee had read the manual.

Before the PASS can be considered for use in forensic settings, empirical data supporting the validity of inferences regarding parent awareness skills developed from PASS scores and evidence supporting the reliability of scores are necessary. It is, however, an interesting attempt to quantify a complicated assessment process (Bischoff, 1995; Cole, 1995).

### **Ackerman-Schoendorf Scales for Parent Evaluation of Custody**

The ASPECT (Ackerman & Schoendorf, 1992) is a rating instrument designed to assess the relative child-rearing abilities of parents. The ASPECT is not a test; rather, it is most appropriately described as a test battery that incorporates observations of and interviews with parents and children, responses to parenting questionnaires, and the results of a variety of test instruments. The tests involved consist of the MMPI/MMPI-2 (parents only), Rorschach Inkblot Technique (parents and children), measures of intelligence (parents and children), achievement measures (children only), and projective storytelling devices (children only). The ASPECT is composed of three subscales (Observational, Social, Cognitive-Emotional) that yield an overall summary index of parenting effectiveness for each parent, referred to as the Parent Custody Index (PCI) (for a more detailed description of the instrument, see Otto & Collins, 1994).

The ASPECT manual is easy to read and well laid out. Both the internal consistency of the various ASPECT scales and the interrater reliability of the ASPECT are presented in the manual. Internal consistency of the scales is moderate, whereas interrater reliability is high (Otto & Collins, 1994). In \*331 contrast to the easy-to-read format of the manual, the authors' presentation of the validity data is confusing. Data supporting the validity of the ASPECT at the current time are limited to the degree of association between PCI score and case outcome. No research regarding the ASPECT has been published in peer-reviewed journals.

Early reviews of the ASPECT (Brodzinsky, 1993) noted that it is a promising instrument that incorporates many of the content domains typically addressed in a custody evaluation. However, other reviewers have noted serious limitations in terms of the basic conceptualization of the ASPECT (Arditti, 1995; Heinze & Grisso, 1996; Melton, 1995; Melton et al., 1997; Wellman, 1994). Melton (1995) offered three primary concerns, two of which relate to the content validity of the instrument. First, some of the items selected for inclusion have no clear relation to custody outcomes (e.g., IQ differences between the parent and child) and only minimally address issues that have been shown empirically to be related to these outcomes (see, e.g., Maccoby & Mnookin, 1992). Second, key factors relevant to the final custody decision (e.g., third-party interviews) are not incorporated into the assessment process, thereby limiting the scope of information used to obtain the PCI score. Third, and finally, by offering a PCI score at all, the ASPECT encourages clinicians to offer ultimate issue opinions regarding legal decisions that are outside of their area of expertise, something that Melton and his colleagues (Melton, 1995; Melton et al., 1997) are opposed to conceptually.

Other significant limitations have been noted by Melton (1995) and by other reviewers with regard to the basic psychometric properties of the ASPECT. Of critical concern is the limited data regarding predictive validity (Arditti, 1995; Heinze & Grisso, 1996; Wellman, 1994). The limited data that do exist only relate to whether the ASPECT can predict judges' custody decisions. As noted above, this is a particularly weak criterion measure in that it is unclear for what purpose a psychological battery that determines judges' decisions ultimately is to be used. Furthermore, there is evidence that the data reported in the manual are unrepresentative of typical custody outcomes, in that nearly half of the cases resulted in the father's obtaining custody. Other limitations noted include the limited size ( $N = 200$ ) and sociodemographic characteristics of the normative sample (disproportionately well-educated Caucasians evaluated by private practitioners) and the lack of clarity in the manual regarding how interrater reliability was established.

### Custody Quotient

The CQ (Gordon & Peek, 1989) initially was developed as a research instrument to assess parenting skills in a custody evaluation and provide a **\*332** single standardized score based on ratings of the parent obtained from various sources. The parent is rated in the following areas: (a) emotional needs, (b) physical needs, (c) no dangers, (d) good parenting, (e) parent assistance, (f) planning, (g) home stability, (h) prior caring, (i) acts and omissions, (j) values, (k) joint custody, and (l) frankness (Fabry & Bischoff, 1992). Parents are rated on a 3-point scale (0 = *weak* to 2 = *highly competent*) across these content areas, with scores being summed and standardized to derive a composite score, or custody quotient.

The CQ manual contains information about its development rationale and research foundations, standardization data, ethical considerations, administration and scoring directions, and remediation information (Fabry & Bischoff, 1992). There is also an eight-page scoring protocol for recording responses, ratings, and demographic information about the parents and children. It should be noted that when one of the authors (Edens) attempted to obtain a copy of the CQ manual to review for the present article, he was informed that it is no longer commercially available (Wilmington Institute, personal communication, November 3, 1999).<sup>6</sup>

The CQ is a potentially useful organizational tool that may help the evaluator structure information from a variety of sources to make a well-informed decision regarding the best interests of the child. However, the CQ suffers from several of the same limitations that have been noted for the other tests reviewed here. First, the psychometric properties of the CQ have not been well established. Although reviewers (Bischoff, 1992; Fabry & Bischoff, 1992) have noted that the content validity of the CQ seems to be adequate, there are no published data regarding its predictive or external validity. Reliability information is not available for the CQ, with the exception of some preliminary test data on interrater reliability. The authors found that interrater reliability ranged from 50% to 100% on five factors. There were, however, only 10 professionals in the sample, including the authors. Also, generalization is a particular problem with the CQ because the normative sample was composed of residents of a single metropolitan area and is not representative of the U.S. population. The CQ authors recognized this problem and suggested in the manual that users develop local norms before employing the test in other settings. Finally, the CQ contains information that can be used to obtain and prescribe

remediation recommendations for parents who lack skills, knowledge, and behavior associated with parenting. However, the theoretical rationale for these recommendations is not provided in the manual.

In summary, reviewers regard the CQ as a potentially useful research instrument, but suggest that it should not be used alone in making a custody decision (Bischoff, 1992; Fabry & Bischoff, 1992). A more conservative evaluation suggests that its use in any forensic capacity is unjustified. \*333 Although the CQ could be a useful measure of parenting skills if more conclusive data were available regarding the statistical properties of the test, based on current professional standards, it would seem inappropriate for use at this time (particularly given the absence of a manual).

## CONSIDERING USE OF TESTS IN CHILD CUSTODY EVALUATIONS

Offered below is a list of questions that mental health professional should ask themselves when considering whether a particular test should be used in the context of a child custody evaluation.<sup>7</sup> We emphasize that the questions presented below are offered only as guidelines and that the significance of any one factor may vary across situations. Mental health professionals should always use their professional judgment when considering the nature and scope of their evaluation processes.

*1. Is the test commercially published?* Ideally, tests used by child custody evaluators will be commercially published. The commercial publication serves as a review process and also ensures a reasonable level of availability and uniformity of test stimulus materials and protocols. Thus, absent unusual circumstances, custody evaluators should avoid using instruments that are not published by an established test publisher.<sup>8</sup> Those who do so risk using a test that has not been subjected to standard review and is flawed in one or more important ways.

*2. Is a comprehensive test manual available?* Test manuals are an authoritative resource that should fully describe the development, standardization, administration, scoring, and psychometric properties of the instrument (including its reliability and validity). Mental health professionals should proceed very cautiously when the manual does not adequately describe the test's development and validation, as this increases the possibility of inappropriate application, administration, scoring, and interpretation. Indeed, both the *Standards for Educational and Psychological Testing* (American Educational Research Association, 1999) and the guidelines developed by Heilbrun (1992) provide some authority, suggesting that tests that lack adequate manuals do not meet minimal standards and should not be used.

*3. Are adequate levels of reliability demonstrated?* Mental health professionals should only use instruments with known and adequate levels of reliability. The particular type of reliability that is most important may vary across tests and situations, but important factors to consider include internal \*334 consistency, interrater reliability, test-retest reliability, and reliability of parallel forms.<sup>9</sup> Because the reliability of a measure limits its validity, tests with poor reliability are tests with poor validity, and tests with unknown reliability are tests with unknown validity (see, generally, Cohen & Swerdlik, 1999). In addition to concerns about the validity of such instruments, examiners who use tests with unknown or poor reliability can expect (and reasonably so) vigorous cross-examinations regarding the opinions based on the test results.

*4. Have adequate levels of validity been demonstrated?* A test must not only be reliable; it must also be valid. Does the test really assess what its authors claim? In the abstract, test validity can be examined in a number of ways (e.g., discriminant validity, predictive validity, concurrent validity, or face validity). Child custody evaluators using instruments must independently assess the validity of measures that they are using. A test is not valid simply because it is published or because its authors purport it to be valid. The validity of opinions that are based on tests with unknown validity is unknown, and opinions based on tests with poor validity are invalid.

5. *Is the test valid for the purpose in which it will be used?* Test validity should be considered by the child custody evaluator not only in the abstract. The more important question for the custody evaluator to ask is whether the test being considered is valid for the purpose for which it will be used. For example, although it is generally accepted that the Wechsler scales provide valid estimates of intellectual ability, few would argue that the Wechsler scales validly inform an examiner's understanding of a father's relationship with his child or understanding and appreciation of his child's emotional adjustment (extremely low scores notwithstanding). This analysis, of course, is of particular significance when custody evaluators consider using tests or measures that assess general psychological constructs that are not specific or are tangentially related to issues of child custody (e.g., measures of intelligence, academic achievement, psychopathology, or personality). As described in more detail above, these cautions should not be interpreted as generally precluding the use of such measures in child custody evaluations. For example, examiners may adopt an approach whereby they assess a general construct and then draw inferences about behaviors and capacities that are more relevant and specific to issues of custody. They should make this inference clear to the consumer of their evaluations (i.e., attorneys, judges, and parents), however, and note any associated limitations. In contrast, however, using general clinical instruments that focus on psychopathology, intelligence, and academic achievement, while neglecting to assess factors more \*335 central to the custody issue, violates essentially all professional practice standards.

Child custody evaluators also must consider whether a test that is of proven validity for a specific purpose is valid with a particular examinee. Issues of ethnicity, race, setting, and age may affect test results and test validity, and child custody evaluators using tests must be informed of the norms and standardization procedures and make decisions about their use with a particular examinee accordingly, with any limitations or caveats being made clear to the consumer. Testing standards, child custody evaluation standards, and general practice guidelines for all custody evaluators require sensitivity to issues of potential bias.

6. *Has the instrument been peer reviewed?* The importance and value of the peer-review process cannot be overestimated. Reliability and validity studies should be published in refereed journals because this allows for the most objective examination of the test and its properties. Absent highly unusual circumstances, child custody evaluators (and mental health professionals conducting any kind of forensic evaluation) should only use tests that have been subjected to the peer-review process and whose properties have been investigated by others, not only their authors. Rules of evidence in all jurisdictions require that techniques used by an expert meet certain criteria, and opinions based, in full or in part, on invalid procedures can be barred (see, e.g., Rule 702, Federal Rules of Evidence; *Daubert v. Merrell Dow Pharmaceuticals*, 1993; *Frye v. U.S.*, 1923). Moreover, child custody evaluators who use instruments of unknown or questionable reliability and validity are doing a disservice to the profession, the legal system, and the families they serve.

7. *What are the qualifications necessary to use this instrument?* Even a valid and appropriate test can produce invalid results if it is administered, scored, or interpreted by a professional who lacks the requisite knowledge, training, and experience. In those cases in which a test meets all the requisite criteria for use in a child custody evaluation, the professional must next consider whether he or she has the requisite general training and knowledge (e.g., in psychometrics and psychological assessment) and more specific training and knowledge (with respect to administration, scoring, and interpretation of the particular test). Simply having access to a test does not mean that one is qualified to use it. Examiners can be embarrassed, and their opinions barred, if the court concludes that they do not have the requisite qualifications to administer, score, or interpret tests that formed the basis of their opinions in custody evaluations.

#### \*336 SUMMARY AND CONCLUSION

Child custody evaluations may be among the most complicated forensic evaluations. As a result of their expertise in matters of child development, parenting, family relations, and psychopathology, mental health professionals are called on to provide the judiciary with insights about children and their parents in a number of important spheres. In conducting such evaluations, child custody evaluators may consider using standardized psychological tests that assess

general constructs that may be related to the specific behaviors and capacities in which the court is interested, or they may consider using forensic assessment instruments developed specifically for purposes of child custody evaluation.

Mental health professionals are obligated by their professional practice standards and child custody evaluation guidelines to use valid techniques that will best inform the court about children and their parents. Child custody evaluators should cautiously consider use of any test. Although the validity of many standard measures (i.e., those that assess personality, psychopathology, intelligence, and academic achievement) has been established, their validity and utility in the context of child custody evaluation is another matter. Certainly, those measures that validly assess general constructs that may be related to parent-child interactions, parent functioning, and child functioning can be included in child custody evaluations, but only if the examiner makes clear the connection or nexus between the general construct being assessed and the legally relevant issue.

In contrast to the above, essentially all of the newer forensic assessment instruments that have been developed specifically for purposes of child custody decision—making have significant limitations. Although those who have developed these instruments have made an important first step, it is clear that further research examining the reliability and validity of these forensic instruments is necessary before they become part of the custody evaluator's assessment process. What is less clear, however, is whether this research will occur. In essentially every published review of these custody assessment instruments, concerns about their reliability and validity have been identified, and the need for research has been made clear. Unfortunately, child custody evaluators continue to wait for that research.

### \*338 REFERENCES

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

Ackerman, M., & Schoendorf, K. (1992). *ASPECT: Ackerman-Schoendorf Scales for Parent Evaluation of Custody—manual*. Los Angeles: Western Psychological Services.

American Academy of Child and Adolescent Psychiatry. (1994). Practice parameters for child custody evaluation. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 57S-68S.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing* (3rd. ed.). Washington, DC: American Educational Research Association.

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

American Psychological Association Committee on Professional Practice Standards. (1994). Guidelines for child custody evaluations in divorce proceedings. *American Psychologist*, 49, 677-682.

Arditti, J. A. (1995). Ackerman-Schoendorf Scales for Parent Evaluation of Custody. In J. C. Conoley & J. C. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 20-22). Lincoln, NE: Buros Institute of Mental Measurements.

Association of Family and Conciliation Courts. (n.d.). *Model standards of practice for child custody evaluation*. Milwaukee, WI: Author.

Bischoff, L. G. (1992). The custody quotient. In J. J. Kramer & J. L. Conoley (Eds.), *The eleventh mental measurements yearbook* (pp. 254-255). Lincoln, NE: Buros Institute of Mental Measurements.



- Bischoff, L. G. (1995). Parent Awareness Skills Survey. In J. L. Conoley & J. C. Impara (Eds.), *The twelfth mental measurements yearbook* (p. 735). Lincoln, NE: Buros Institute of Mental Measurements.
- Bricklin, B. (1989). *Perception of Relationships Testmanual*. Furlong, PA: Village Publishing.
- Bricklin, B. (1990a). *Bricklin Perceptual Scales Manual*. Furlong, PA: Village Publishing.
- Bricklin, B. (1990b). *Parent Awareness Skills Survey manual*. Furlong, PA: Village Publishing.
- Bricklin, B. (1992). Data-based tests in child custody evaluations. *American Journal of Family Therapy*, 20, 254-265.
- Bricklin, B. (1999). The contributions of psychological tests to child custody evaluations. In R. M. Galatzer-Levy and L. Kraus (Eds.), *The scientific basis of child custody decisions* (pp. 120-156). New York: John Wiley.
- Bricklin, B., & Elliott, G. (1991). *Parent Perception of Child Profile Manual*. Furlong, PA: Village Publishing.
- Bricklin, B., & Elliott, G. (1997). *Critical child custody evaluation issues: Questions and answers. Test manuals supplement for BPS, PORT, PASS, PPCP*. Furlong, PA: Village Publishing.
- Brodzinsky, D. M. (1993). On the use and misuse of psychological testing in child custody evaluations. *Professional Psychology: Research and Practice*, 24, 213-219.
- Carlson, J. N. (1995). Perception-of-Relationships Test. In J. C. Conoley & J. C. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 746-747). Lincoln, NE: Buros Institute of Mental Measurements.
- \*339 Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to tests and measurement* (4th ed.). Mountain View, CA: Mayfield.
- Cole, D. E. (1995). Parent Awareness Skills Survey. In J. L. Conoley & J. C. Impara (Eds.), *The twelfth mental measurements yearbook* (p. 735). Lincoln, NE: Buros Institute of Mental Measurements.
- Conger, J. (1995). Perception-of-Relationships Test. In J. C. Conoley & J. C. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 747-748). Lincoln, NE: Buros Institute of Mental Measurements.
- Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S.Ct. 2786 (1993).
- Fabry, J. J., & Bischoff, R. A. (1992). The Custody Quotient. In D. Keyser & R. Sweetwater (Eds.), *Test critiques* (Vol. 9, pp. 145-148). Austin, TX: PRO-ED.
- Felner, R. D., Rowlison, R. T., Farber, S. S., Primavera, J., & Bishop, T. A. (1987). Child custody resolution: A study of social science involvement and impact. *Professional Psychology: Research and Practice*, 5, 468-474.
- Frye v. U.S., 293 F. 1013 (D.C. Cir. 1923).
- Gordon, R., & Peek, L. A. (1989). *The Custody Quotient: Research manual*. Dallas, TX: Wilmington Institute.
- Grisso, T. (1984, June). *Forensic assessment in juvenile and family cases: The state of the art*. Keynote address at the Summer Institute on Mental Health Law, University of Nebraska, Lincoln.

- Grisso, T. (1986). *Evaluating competencies: Forensic assessments and instruments*. New York: Plenum.
- Hagin, R. A. (1992). Bricklin Perceptual Scales. In J. L. Conoley & J. C. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 117-118). Lincoln, NE: Buros Institute of Mental Measurements.
- Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior*, 16, 257-272.
- Heilbrun, K., Rogers, R., & Otto, R. K. (2000, March). *Forensic assessment: Current status and future directions*. Paper presented at the Biennial Meeting of the American Psychology Law Society, New Orleans, Louisiana.
- Heinze, M. C., & Grisso, T. (1996). Review of instruments assessing parenting competencies used in child custody evaluations. *Behavioral Sciences and the Law*, 14, 293-313.
- Hiltonsmith, R. W. (1995). Parent Perception of Child Profile. In J. C. Conoley & J. C. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 736-738). Lincoln, NE: Buros Institute of Mental Measurements.
- Keilin, W. G., & Bloom, L. J. (1986). Child custody evaluation practices: A survey of experienced professionals. *Professional Psychology: Research and Practice*, 17, 338-346.
- Kelley, M. L. (1995). Parent Perception of Child Profile. In J. C. Conoley & J. C. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 738-739). Lincoln, NE: Buros Institute of Mental Measurements.
- LaFortune, K. A., & Carpenter, B. N. (1998). Custody evaluations: A survey of mental health professionals. *Behavioral Sciences and the Law*, 16, 207-224.
- Maccoby, E. E., & Mnookin, R. H. (1992). *Dividing the child: Social and legal dimensions of child custody*. Cambridge, MA: Harvard University Press.
- Melton, G. B. (1995). Ackerman-Schoendorf Scales for Parent Evaluation of Custody. In J. C. Conoley & J. C. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 22-23). Lincoln, NE: Buros Institute of Mental Measurements.
- \*340 Melton, G. B., Petrila, J., Polythress, N. G., & Slobogin, C. (1997). *Psychological evaluations for the courts: A handbook for attorneys and mental health professionals and lawyers* (2nd ed.). New York: Guilford.
- Melton, G. B., Weithorn, L. A., & Slobogin, C. (1985). *Community mental health centers and the courts: An evaluation of community-based forensic services*. Lincoln: University of Nebraska Press.
- Munsinger, H. L., & Karlson, K. W. (1994). *Uniform Child Custody Evaluation System*. Odessa, FL: Psychological Assessment Resources.
- Otto, R. K., & Collins, R. (1994). Use of the MMPI-2/MMPI-A in child custody evaluations. In Y. S. Ben-Porath, J. R. Graham, G.C.N. Hall, R. D. Hirschman, & M. S. Zaragoza (Eds.), *Forensic applications of the MMPI-2* (pp. 222-252). Beverly Hills, CA: Sage.
- Paulhus, D. L. (1999). *Manual for the Paulhus Deception Scales*. North Tonawanda, New York: Multi-Health Systems.
- Reidy, T., Silver, R., & Carlson, A. (1989). Child custody decisions: A survey of judges. *Family Law Quarterly*, 23, 75-87.

- Rogers, R., & Webster, C. (1989). Assessing treatability in mentally disordered offenders. *Law and Human Behavior*, 13, 19-29.
- Settle, S. A., & Lowery, C. R. (1982). Child custody decisions: Content analysis of a judicial survey. *Journal of Divorce*, 6, 125-138.
- Shaffer, M. B. (1992). Bricklin Perceptual Scales. In J. L. Conoley & J. C. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 118-119). Lincoln, NE: Buros Institute of Mental Measurements.
- Wellman, M. M. (1994). Ackerman-Schoendorf Scales for Parent Evaluation of Custody. In D. Keyser & R. Sweetwater (Eds.), *Test critiques* (Vol. 10, pp. 13-19). Austin, TX: PRO-ED.

#### Footnotes

- a1 *Randy K. Otto is an associate professor in the Department of Mental Health Law and Policy, Florida Mental Health Institute, University of South Florida. He is president of the American Academy of Forensic Psychology and serves on the editorial boards of Psychological Assessment, Law and Human Behavior, Behavioral Sciences and the Law, Correctional Mental Health Reporter, and the Journal of Behavioral Health Services and Research.*
- a2 *John F. Edens is an assistant professor in the Forensic Clinical Psychology Program at Sam Houston State University, where he has taught and conducted research since 1998. After obtaining a doctorate in clinical psychology from Texas A&M University in 1996, he completed a 2-year postdoctoral fellowship in forensic psychology in the Department of Mental Health Law and Policy at the University of South Florida.*
- a3 *Elizabeth H. Barcus is a doctoral student in the Forensic Clinical Psychology Program at Sam Houston State University. She received a master's degree in clinical psychology from Wheaton College in 1996.*
- 1 There is a small body of literature on the relative importance that judges and lawyers assign to psychological evaluations in the decision-making process. Reidy, Silver, and Carlson (1989) reported perhaps the most optimistic finding regarding judges' perceptions of the role of mental health experts. As part of a survey of important decision-making factors, 156 judges rated "testimony of court-appointed psychologist" as the fourth most important criterion, following "desires of children, age 15," "custody investigation report," and "testimony of the parties." Other researchers, however, have reported that less weight is assigned to evidence provided by those in the mental health field. In a sample of 43 superior court judges eligible to hear family law cases and 74 attorneys involved in custody cases, Felner, Rowilson, Farber, Primavera, and Bishop (1987) reported that only 20% of the attorneys and only 2% of the judges listed the recommendations of a mental health professional as one of the "five most critical" criteria to consider when making custody decisions. When asked to indicate the ideal decision-making situation, only 7% of both judges and attorneys listed recommendations of mental health professionals as among the five most relevant criteria. In a content analysis of custody decision-making of 57 judges and 23 trial commissioners. Settle and Lowery (1982) reported that "professional advice" was ranked 12th on a list of 20 potential criteria employed by judges and trial commissioners making custody decisions. Melton, Weithorn, and Slobogin (1985) reported a similarly bleak picture, with nearly half of their sample of judges reporting that mental health testimony in custody cases was no more than occasionally useful.
- 2 Unfortunately, the authors were not more descriptive in terms of which specific instruments developed by Bricklin were identified by the respondents.
- 3 Although psychologists are not only mental health professionals who conduct testing and evaluations, guidelines from this discipline's ethics code are specifically referenced here because (a) psychologists are most likely to include tests in their child custody evaluations and (b) the American Psychological Association (APA) code specifically addresses issues of psychological testing.
- 4 We do not review standard tests (e.g., measures of intelligence, academic achievement, psychopathology, or personality) that may be included in child custody evaluations, for two reasons. First, the validity of these instruments has been reviewed

extensively in other venues (see, e.g., *Mental Measurements Yearbook* or *Test Critiques* for such reviews); second, the limitations of these tests, including the degree of inference necessary when using them in the context of child custody evaluations, have been covered adequately here and by other authors (e.g., Brodzinsky, 1993; Bricklin, 1992, 1999). We also do not review evaluation formats that have been developed for use in child custody evaluation (e.g., Uniform Child Custody Evaluation System; see Munsinger & Karlson, 1994), because these formats are not tests but, rather, are structured clinical evaluations.

- 5 Convergent validity refers to the extent to which a test provides results that are similar to the results of other tests that purport to measure the same psychological construct.
- 6 We include a review of the Custody Quotient because some examiners apparently continue to employ it in their evaluations (Ackerman & Ackerman, 1997).
- 7 This section is adapted, in part, from Heilbrun, Rogers, and Otto (2000).
- 8 Evaluators should keep in mind that test publishers sometimes market, sell, and distribute tests that they have not published. Therefore, inclusion of an instruments in a test publisher's catalog does not necessarily mean that the test has been published by that firm.
- 9 This form of reliability refers to the extent to which, if two versions of a test exist, they have been demonstrated to obtain similar results with the same examinee.

38 FAMCCR 312

---

End of Document

© 2018 Thomson Reuters. No claim to original U.S. Government Works.