

Copyright © 2001 Matthew Bender & Company, Inc., a member of the LexisNexis Group

California Family Law Monthly

May 2001

2001 Calif. Family Law Monthly 105

### **Issues in the Use of Psychological Testing in Custody Evaluations**

By Barbara J. Tinsley, Ph.D. and Elaine S. Rosen, J.D., Esq.[n\*]

Courts order child custody evaluations to facilitate the process of determining the health, safety, welfare, and best interest of children in disputed custody and visitation situations. Cal. Rules of Ct., Rule 1257.3 sets forth "[u]niform standards of practice for court-ordered child custody evaluations" that are consistent with Fam. Code § § 211 and 3117. These standards require evaluators to "observe and consider each party in comparable ways and to substantiate (from multiple sources when possible) interpretations and conclusions regarding each child's developmental needs; the quality of attachment to each parent and that parent's social environment; and reactions to the separation, divorce, or parental conflict" [Cal. Rules of Ct., Rule 1257.3(e)(2)]. Furthermore, these standards suggest that the scope of the evaluation process may include psychological testing to confirm or question other information or previous conclusions [Cal. Rules of Ct., Rule 1257.3(e)(1)(B)].

In the context of this mandate, and the complexity involved in making custody evaluations, evaluators often choose to utilize traditional psychological testing of the adults and children involved in a custody dispute. For the mental health professionals who engage in custody evaluation, psychological testing is one of the principal methods of assessment. Estimates of the pervasiveness of psychological testing in custody evaluations suggest that they are used in from 50-80 per cent of all evaluations.

Psychological testing is perceived by psychologists, other mental health professionals, attorneys, and the judiciary as being scientific and impartial. In most cases, the result of psychological testing is a quantified score, which can be interpreted in light of standardized norms. These scores provide what is perceived as an objective determination of the clinical status of the family member, or a concurrent or predictive criteria for children's psychological outcomes.

However, while many evaluators use psychological testing for these reasons, this practice has been widely criticized. It has been noted that traditional psychological tests were not developed to specifically address the psychological and legal issues relevant to child custody decision making. Thus, evaluators are required to extrapolate from the general focus of personality, intelligence, and psychopathology assessments to the much more specific issues related to child custody, such as parents' relationships with children or the ability to parent, which are the foci of such evaluations. It has also been argued that traditional psychological testing, with its goal of diagnosis, has little value in custody evaluations because diagnosis is not relevant to evaluation goals, namely, determining how best to meet children's psychological and physical needs.

Most psychological testing is only able to identify individuals with serious pathological thought processes and behaviors, which is not the case in many child custody evaluations. Additionally, it is crucial to consider whether identifying psychological dysfunction in a parent is always relevant to the custody determination. In other words, if disordered thinking or behavior is identified in a parent, does it impact their ability to parent, and children's development? The answer to this question is not as clear as might be thought. While the implications of some parental pathology such as substance abuse, severe depression, and schizophrenia have been demonstrated to negatively influence children's normal development, not all forms of parental psychological dysfunction do. Moreover, careful studies specifying the effects of many types of adult pathology on children experiencing custody placement have not yet

been done in a manner that offers the ability to specify the clinical criteria to identify a parent as incapable of providing adequate emotional and physical care to children.

The quality of tests and their results are measured by their psychometric properties. Psychometric properties refer to a number of quantifiable indicators of the extent to which a test can be considered reliable (providing consistency in assessment) and valid (providing accuracy in assessment). There are many sources of information about the various tests obtainable to mental health professionals for use in clinical assessments. These include the test manuals published to accompany commercially available tests, and such reference books as the *Mental Measurements Yearbook*. Test manuals are often the most authoritative concerning a test, and provide detailed information concerning the purposes of a test, how it was developed, and how it should be used. Thorough understanding of the information offered in test manuals permits professionals awareness of the strengths and limitations of a particular assessment. Obviously, knowledge of what a test can and cannot offer in terms of assisting in evaluators' decision making processes with respect to custody disputes is crucial. All tests have limitations and shortcomings, and they can be discovered by careful study of test manuals. These test manuals contain technical information and statistics, and may be intimidating to those not trained in deciphering their import. Yet, without understanding this information about tests, evaluators and those processing their evaluations are more likely to be poor users of the information obtained with the tests.

Some of the more commonly accepted dimensions of the psychometric qualities of tests are various types of reliability and validity. Reliability refers to the stability of test scores. Several forms of reliability in tests are often measured. For example, test-retest reliability is concerned with the stability of test scores over time; that is, to what extent will individuals achieve similar scores when taking the test twice, with separations of various lengths of time? This type of reliability is critical in custody evaluations due to the volatility of the psychological state of parents being evaluated in custody disputes. Another type of reliability, internal consistency reliability, measures the extent to which different items within the same test will yield the same results.

Validity, the measurement of the accuracy of a test, is also assessed in a variety of ways. Convergent validity refers to the extent to which a test provides results that are similar to the results of other tests that claim to measure the same psychological construct. Concurrent criterion-related validity measures how well scores on a newly developed test match scores on well-used, psychometrically acceptable measures. For example, individuals' scores on newly developed assessments of parenting skills can be compared to their scores on older, traditional assessments of parenting that have been widely-used in the field for some time. A second type of validity, predictive validity concerns the extent to which individuals' scores on a test will help predict their assessment at a later point on some meaningful dimension. This is a very critical type of psychometric criteria for assessments used in custody evaluations. One of the main goals of custody evaluations is to attempt to predict the best rearing environment that is likely to provide for children's emotional, intellectual, and physical well-being. Thus, psychological assessment is most valuable to evaluators when it addresses the qualities of the relationship between past and present parenting, and child development.

Another important type of validity includes content validity. Content validity measures the extent to which a test reflects and measures attainment of its assessment objectives. In other words, content validity refers to the systematic evaluation of test content to determine whether it covers a representative sample of the behavior domain covered.

Other psychometric properties of tests, in addition to reliability and validity include the standard error of measurement. This is the margin of error used to interpret a test score. An individual's true score on a test is never known, because some degree of measurement error is always present in any assessment. Measurement errors occur because all human behavior varies from time to time, and because all psychometric measuring devices are imprecise to some degree.

Cultural considerations moderate the interpretation of test scores of personality and behavior. Some beliefs, attitudes and behaviors may be regarded as normative (i.e., normal and reasonable) in some cultures, while being regarded as deviant in others. One example of such beliefs, attitudes, and behaviors relevant to custody disputes is parent discipline of children. Within various subcultures in the U.S., a variety of attitudes about physical punishment exist, and methods of physical punishment that might be considered extremely punitive and abusive in one subculture might be perceived in another as necessary to inculcate the extent of compliance in children that will protect them from very dangerous neighborhood violence and danger.

The history of psychometric tests has witnessed many criticisms of various instruments as unfair or unrepresentative with respect to issues of gender, culture, or other demographic features. In the selection of the final items to compose a test, a number of steps can be taken to try to eliminate such biases. For example, bias review panels can be assembled that represent professionals or lay persons from varying backgrounds. Members of such bias review

panels are given lists of preliminary test items and instructed to determine whether any of the items could be perceived as irrelevant, confusing, or inappropriate on the basis of gender, religion, or ethnic group membership. Such panels often identify items that could be endorsed in a pathological direction because of beliefs that are normative with a particular subculture. Obviously, the utilization of a bias review panel does not guarantee the elimination of bias in the test. However, it does ensure that each item has been carefully considered by a group of individuals that professionally and experientially reflects the diversity of the population likely to be administered this test.

Other approaches to reducing ethnic/cultural bias in tests is to minimize the use of language in such tests. The rationale for this approach is that language is a major component of culture. Another method is to try out potential test items with diverse groups and select those items that minimize differences among the groups for inclusion in the final form of the test.

Another potential weakness of tests is that examinees may not respond honestly and accurately. For example, some test-takers may give responses they believe are socially desirable. Some tests include items or sets of items designed to identify these individuals. One type of such items are Inconsistent Response items, which indicate whether there is an unusually wide discrepancy in an individual's response to pairs of items with similar content--pairs of items such as "I am an attractive person" and "I look fine just the way I am". Such discrepancies are sometimes due to haphazard or careless responding. Sometimes, however, they reflect some peculiarity in an individual's life circumstance, such as involvement in a custody dispute.

Another class of items designed to pinpoint individuals who may be systematically giving responses they believe are socially desirable are "Self-criticism" items. Such test items are mildly derogatory statements, such as "I get angry sometimes"--common frailties that most people would admit when responding candidly. If an individual's response to Self-criticism items is too low, their overall test score may be invalid. An individual who denies too many such statements is probably being defensive and making a deliberate effort to present a favorable picture of himself or herself, and the low score on these items suggests that the overall test score may be artificially elevated by this defensiveness. The low Self-criticism score is a sign that further investigation, perhaps through interview questions or study of the individual's records, should be initiated.

Increasingly, mental health professionals involved in custody evaluation are developing specific custody tests to address the more precise questions in evaluations, such as the adequacy of parenting skills, or the nature and quality of the child's various relationships with caregivers. While these tests have the advantage of more precisely speaking to the parenting issues relevant in custody disputes, they are much newer and have much less reliability and validity data than older, more traditional psychological assessments. Thus, an unfortunate paradox exists with respect to the tests available for custody evaluations. On one hand, the best assessment tools in terms of technical standards are less specific for answering the questions asked in the evaluation. The tests that more precisely answer the questions involved in custody determination, however, are not as technologically advanced or reliable as the more traditional and better evaluated assessment tools that measure more general concepts such as personality and intelligence.

In light of the advantages and disadvantages of using psychological testing discussed above, instead of asking the question of *whether* testing in custody evaluations is appropriate, the better question to ask is: *when* is testing appropriate in evaluations? One obvious caveat for evaluators and those in the judiciary reviewing their reports might be to carefully consider the links between for whom and how a specific test was developed, and the exact question being asked about a parent or child. For example, personality tests developed to measure hostility in psychiatric patients might not be the proper test of choice in measuring parental hostility in the context of a high conflict custody dispute.

As an alternative to psychological testing, it has been argued that interviews or observations of behavior provide more intrinsically accurate sources of information as a basis for evaluators' custody recommendations. However, interviews and observational methods do not, for the most part, offer external scores or criteria or validity estimates, as do psychological tests, for which a range of responses has already been specified in advance and analyzed systematically.

Some observational procedures for assessing parent-child interaction have been standardized. There are a number of tasks that have been presented for family members to perform individually or in dyads or other combinations, such as building a block construction or making origami, that permit systematized objective assessment of specified verbal and nonverbal behavior. Observations of the process by which members of a dyad decide how to proceed, the degree of cooperation and directions exchanged, limit-setting, nurturance, the extent to which participants exhibit competition, and which family member adopts a dominant role, are examples of observational constructs that can be systematically

observed by watching families engage in such tasks, and from which implications for the childrearing environment that parents provide can be derived.

In summary, psychological testing in child custody evaluations is probably best used as confirmatory information, with clear discussion of its limitations, in the context of other forms of evidence concerning parents' ability to provide appropriate childrearing environments. Psychological assessment using standardized tests encompasses much more than mere knowledge about the implications of the scores obtained. Those involved in using assessments must also be knowledgeable about aspects of individuals that lead them to respond to psychological tests in certain ways and how they lead their lives correspondingly. Various tests measure dimensions of attitudes and behavior in different ways and with different emphases, even when those tests might appear similar in format, due to the variety of approaches possible in test construction. Mental health professionals using standardized tests have a responsibility, as defined by the American Psychological Association guidelines, to be familiar with the psychometric qualities of the tests they use, including aspects of the tests' reliability and validity, as well as cultural issues associated with testing.

#### FOOTNOTES:

Footnote \*. Barbara Tinsely is an instructor with the Dept. of Psychology at University of California, Riverside, CA. Elaine Rosen is the Director, Dept. of Law and Public Policy, University of California, Riverside Extension.