
*THEORETICAL AND TEST
USAGE ISSUES*

The Responsible Use
of Psychological Testing
in Child Custody Evaluations:
Selection of Tests

James R. Flens

SUMMARY. The responsible use of psychological tests in child custody evaluations requires an advanced understanding of both psychological issues of test selection and legal criteria regarding admissibility of

James R. Flens, PsyD, has a private practice in Brandon, FL. His practice is centered around family law-related evaluations, including evaluations regarding custody, modification and relocation, work product review, and consultation with both the legal community and custodial evaluation professionals. He is also an instrument-rated pilot and factory-trained bicycle mechanic.

Address correspondence to: James R. Flens, PsyD, 1463 Oakfield Drive, Suite 111, Brandon, FL 33511 (E-mail: jayflens@aol.com).

The author wishes to thank Jon Gould, PhD, and Leslie Drozd, PhD, for assistance in preparing this paper.

[Haworth co-indexing entry note]: "The Responsible Use of Psychological Testing in Child Custody Evaluations: Selection of Tests." Flens, James R. Co-published simultaneously in *Journal of Child Custody* (The Haworth Press, Inc.) Vol. 2, No. 1/2, 2005, pp. 3-29; and *Psychological Testing in Child Custody Evaluations* (ed: James R. Flens, and Leslie Drozd) The Haworth Press, Inc., 2005, pp. 3-29. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: docdelivery@haworthpress.com].

Available online at <http://www.haworthpress.com/web/JCC>

© 2005 by The Haworth Press, Inc. All rights reserved.

Digital Object Identifier: 10.1300/J190v02n01_02

expert testimony. This paper discusses the psychological and legal issues associated with test selection and with admissibility of expert testimony pertaining to psychological test data. It is argued that the legal standards of relevance and helpfulness require the methodology underlying an expert's testimony to be both reliable and valid. Therefore, it is essential to select psychological tests with demonstrated reliability and validity. Case law regarding expert testimony and the integration of professional practice guidelines pertaining to the use of psychological tests with ethical standards will be discussed. [Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <<http://www.HaworthPress.com>> © 2005 by The Haworth Press, Inc. All rights reserved.]

KEYWORDS. Assessment, child custody, ethics, evaluation, Daubert, Kumho, test selection

PART 1. SELECTION OF TESTS

Then the king said, "Bring me a sword." So they brought a sword before the king. And the king said, "Divide the living child in two, and give half to one, and half to the other." (1 Kings 3:24-25)

As the first recorded child custody dispute reveals, historical resolution of child custody disputes was a little bit different than it is today. Unlike King Solomon—who had only his sword and his wits—today's custody evaluators operate in two worlds. They operate in the legal world, which encompasses state statutes, case law precedents, and rules of evidence. They also operate in the mental health world, which encompasses the application of forensic methods and procedures (Martindale & Gould, 2004) and an arsenal of tests and measures used in conducting child custody evaluations (Kirkpatrick, 2003; see, e.g., Ackerman, 2001; Condie, 2003; Gould, 1998, 1999; Heilbrun, 2001; Schutz, Dixon, Lindenberger, Child, & Ruther, 1989; Stahl, 1994; Woody, 2000). In this article, I describe the interdependence between legal standards and psychological ethics applied to the selection of psychological tests in child custody evaluations. I argue that evaluators' responsible use of psychological tests begins with an understanding of rules of evidence governing expert testimony and an understanding of legal and psychological concepts of reliability, relevance, and helpfulness.

A BRIEF REVIEW OF PSYCHOMETRIC CONCEPTS

The use of psychological testing in child custody evaluations requires evaluators to possess (or develop) an advanced understanding of psychometric issues (i.e., reliability, validity), the effects of context on the test data, the use of context-specific normative data, and the legal criteria and admissibility standards for psychological data that are found in statutes and case law. It also requires evaluators to consider sources of bias that may affect interpretation of test results, including evaluator biases such as confirmatory bias (Borum, Otto, & Golding, 1993), confirmatory distortion (Martindale, in press) or “psychotic certainty” (Martindale, 2004), and test-taker bias (e.g., response styles including impression management and self-deceptive enhancement; see e.g., Friedman, Lewak, Nichols, & Webb, 2001; Greene, 2000; Paulhus, 1998). A quick review of three common psychometric¹ terms might be appropriate at this time for those who are not familiar with testing terminology (see, e.g., American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Anastasi & Urbina, 1997). *Reliability* refers to the consistency of results, including but not limited to consistency across time, situation, and evaluator; it asks the question, “Does the test consistently measure what it is purported to measure?” *Validity* refers to the accuracy of the test; it answers the question, “Does the test accurately measure what it is purported to measure?” The *Standard Error of Measurement* refers to the margin of error surrounding a test score; it answers the question, “What are the likely upper and lower boundaries of a person’s true score on a test?”

There are important relationships between the reliability and validity of a test. First, a test’s validity cannot be more than its reliability because the reliability coefficient is part of the denominator of the validity equation. Second, a test may be reliable and invalid. That is, a test may measure something consistently, but does not measure the factor accurately. The converse is not true. If a test is valid, it must be reliable. Third, if a test has low reliability, it also has low validity. As Otto, Edens, and Barcus (2000) stated, “[T]he 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” (p. 33).

It is important to understand that the term “reliability” has different meanings when used in the psychological or legal communities. From a psychological perspective, the term reliability means “consistency,” as noted above. From a legal perspective, however, the term reliability re-

fers to accuracy, which is “validity” from the psychological perspective. The different uses of this term may cause confusion when discussing psychological and legal issues.

AREAS OF THE LAW

Custody evaluators should be familiar their state’s rules of evidence, with particular attention paid to rules governing expert testimony and the admissibility of expert testimony. For the purposes of this article, I draw attention primarily to these important areas of the law. However, evaluators also need to be familiar with two other areas of the law. It is important to have knowledge of case law decisions relevant to child custody determinations. Case law decisions are how the Court interprets and clarifies the legal standards (statute, rule). For example, many states may have case law decisions that specifically identify factors to be examined in a relocation case or factors that define a reliability analysis. Florida, for example, codified the relocation criteria espoused by the Court in *Mize v. Mize* (1993) and *Russenberger v. Russenberger* (1996). Various states have used case law to define and clarify the admissibility of expert opinion testimony. California, for example, modified the Frye test with *People v. Kelly* (1976). Tennessee, on the other hand, rejected the use of the Frye test in its opinion of *McDaniel v. CSX Transportation, Inc.* (1997). In that case, the Court expanded the Daubert criteria to make that state’s rule more stringent than the federal standard.

The third area is knowledge of state statutes defining the best interest of the child standard and other concepts relevant to child custody determinations. These standards and concepts inform the evaluator about what specifically can and should be addressed in the evaluation itself. The Michigan Standard, for example, is often considered as the model set of guidelines or criteria the Court uses to determine the best interests of the child (see, e.g., Otto, Buffington-Vollum, & Edens, 2003) (see Table 1).

Rules of Evidence

Rules of Evidence define what can and cannot be admitted into evidence. There are Federal Rules of Evidence (FRE) that apply to federal courts and there are state rules of evidence that apply to state courts. Most state courts have adopted rules that closely resemble the FRE. It is

TABLE 1. Michigan's Child Custody Statute for Determining "Best Interests of the Child"

The "Michigan Standard" ⁴
<ul style="list-style-type: none"> • The love, affection, and other emotional ties existing between the parties involved and the child; • The capacity and disposition of the parties involved to give the child love, affection, and guidance and continuation of educating and raising the child in his or her religion or creed, if any; • The capacity and disposition of the parties involved to provide the child with food, clothing, medical care, or other remedial care recognized and permitted under the laws of this state in lieu of medical care, and other material needs; • The length of time the child has lived in a stable, satisfactory environment and the desirability of maintaining continuity; • The permanence, as a family unit, of the existing or proposed custodial home; • The moral fitness of the parties involved; • The mental and physical health of the parties involved; • The home, school, and community record of the child; • The reasonable preferences of the child, if the court deems the child to be of sufficient age to express a preference; • The willingness and ability of each of the parents to facilitate and encourage a close and continuing parent-child relationship between the child and the other parent; • Any other factor considered by the court to be relevant to particular child custody dispute.

strongly recommended that custody evaluators become aware of relevant statutes, codes, rules of court, and case law. Although most states have evidence codes that are quite similar in structure and intent to the FRE, not all states follow the FRE. It is important, therefore, that evaluators know their state's evidence code in relevant areas. For the purposes of this article, the FRE will form the basis of our discussion.

Two of the most important sections of the evidence code that are used to determine the admissibility of expert testimony address the relevance of the evidence and the helpfulness to the judge of the evidence. State Evidence Codes will have at least one definition of relevance. The examples described below are from the FRE (see Table 2). The first rule (FRE 401) defines “relevant evidence” as any information that may make the existence of a fact more or less likely. The assumption is that the testimony provided to the court will help in determining a fact, and that without the testimony the determination of the fact would be less probable. All testimony is admissible unless the testimony does not help make a fact more or less likely. Then, the testimony is deemed as not relevant and, therefore, inadmissible (Rule 402). However, some evidence may be ruled as inadmissible if it is harmful, confusing, misleading, a waste of time, or a repetition of facts already in evidence (Rule 403).

Another important part of any evidence code (FRE 702; see Table 2) addresses opinions and testimony provided by experts. State codes, following the structure of the FRE, will often provide both a definition of

TABLE 2. Federal Rules of Evidence: Relevance and Helpfulness

Relevance (FRE 401, 402, 403) and Helpfulness (FRE 702)

Rule 401 Definition of “Relevant Evidence”: “Relevant evidence” means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.

Rule 402 Relevant Evidence Generally Admissible; Irrelevant Evidence Inadmissible: All relevant evidence is admissible, except as otherwise provided by the Constitution of the United States, by Act of Congress, by these rules, or by other rules prescribed by the Supreme Court pursuant to statutory authority. Evidence which is not relevant is not admissible.

Rule 403 Exclusion of Relevant Evidence on Grounds of Prejudice, Confusion, or Waste of Time: Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.

Rule 702 Testimony by Experts: If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

expert testimony and a description of the court might identify expert from no-expert testimony. As discussed below, the evidence rules governing expert testimony are drawn both from Rules of Evidence (see Table 2) and from case law.

Prior to 1923, admissibility of expert testimony was governed by the court's review of an expert's credentials and a review of the potential testimony to determine if this testimony would be helpful. If it was determined that the testimony would be helpful, the expert was then allowed to testify (see, e.g., *Congress & Empire Spring Co. v. Edgar*, 1878; *Winans v. New York & Erie Railroad Co.*, 1858). Beginning in 1923, however, the standard for admissibility of expert testimony was governed by the "General Acceptance Test" articulated in *Frye v. U.S.* (1923). In that case, a federal appellate court opined:

Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs. (p. 1014)

Echoing the court's focus on helpfulness, the Federal Court held in *Jenkins v. U.S.* (1962) that, "The test, then, is whether the opinion offered will be likely to aid the trier in the search for the truth" (p. 643).

Scholarly debate and diverging decisions in the Federal Court addressing whether the General Acceptance Test or an analysis of the reliability of the proffered testimony were the relevant admissibility standards led the Supreme Court of the United States to reexamine the criteria for admissibility of expert testimony (Goodman-Delahunty, 1997; Krauss & Sales, 1999; Shuman & Sales, 1999). In 1993, the U.S. Supreme Court ruled in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993) that the FRE focus on reliability was the proper standard for examining admissibility of expert testimony. This ruling has become known as the "Daubert standard" or "Daubert criteria." The *Daubert* Court defined "scientific knowledge" as follows:

"The adjective 'scientific' implies a grounding in the methods and procedures of science. Similarly, the word 'knowledge' connotes more than subjective belief or unsupported speculation" (p. 590).

But in order to qualify as scientific knowledge, an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation (i.e., “good grounds”) based on what is known. In short, the requirement that an expert’s testimony pertain to scientific knowledge establishes a standard of evidentiary reliability. (p. 590)

The *Daubert* Court identified the judge as a gatekeeper for admissibility of expert testimony. Judges now had the responsibility of examining the underlying scientific methodology for its reliability. If the methodology was judged reliable, then information that flowed from that methodology and the opinions upon which expert testimony was based were allowed. The standard envisioned was to be a flexible set of guidelines the trial Court *could* use (as opposed to “should”) in determining the admissibility of expert testimony. More specifically, the Court noted:

Faced with a proffer of expert scientific testimony, then, the trial judge must determine at the outset, pursuant to Rule 104(a), whether the expert is proposing to testify to (1) scientific knowledge that (2) will assist the trier of fact to understand or determine a fact in issue. This entails a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue. We are confident that federal judges possess the capacity to undertake this review. Many factors will bear on the inquiry, and we do not presume to set out a definitive checklist or test. But some general observations are appropriate. (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 1993, pp. 592-593)

The following (flexible) guidelines were offered by the Court and became known as the *Daubert Standard*:

[Testability or Falsifiability] Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested. “*Scientific methodology today is based on generating hypotheses and testing them* (emphasis added) to see if they can be falsified; indeed, this methodology is what distinguishes science from other fields of human inquiry.” (p. 593)

[Peer Review] Another pertinent consideration is whether the theory or technique has been subjected to peer review and publication. Publication (which is but one element of peer review) is not a sine qua non of admissibility; it does not necessarily correlate with reliability, and in some instances well-grounded but innovative theories will not have been published. Some propositions, moreover, are too particular, too new, or of too limited interest to be published. But submission to the scrutiny of the scientific community is a component of “good science,” in part because it increases the likelihood that substantive flaws in methodology will be detected. . . . The fact of publication (or lack thereof) in a peer reviewed journal thus will be a relevant, though not dispositive, consideration in assessing the scientific validity of a particular technique or methodology on which an opinion is premised. (pp. 593-594)

[Error Rate and Standards of Control] Additionally, in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error, and the existence and maintenance of standards controlling the technique’s operation. (p. 594)

[General Acceptance] Finally, “general acceptance” can yet have a bearing on the inquiry. A “reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community.” Widespread acceptance can be an important factor in ruling particular evidence admissible, and “a known technique which has been able to attract only minimal support within the community,” may properly be viewed with skepticism. (p. 594)

The inquiry envisioned by Rule 702 is, we emphasize, a flexible one. Its overarching subject is the scientific validity and thus the evidentiary relevance and reliability—of the principles that underlie a proposed submission. The focus, of course, must be solely on principles and methodology, not on the conclusions that they generate. (pp. 594-595)

The Court went on to state:

To summarize: “General acceptance” is not a necessary precondition to the admissibility of scientific evidence under the Federal Rules of Evidence, but the Rules of Evidence—especially Rule 702—do assign to the trial judge the task of ensuring that an expert’s testimony both rests on a reliable foundation and is relevant to the task at hand. Pertinent evidence based on scientifically valid principles will satisfy those demands. (p. 597)

In 1997, the U.S. Supreme Court further extended their thinking on *Daubert* in *General Electric Co. v Joiner* (1997). The *Joiner* decision focused attention on the need for the expert to show how opinions expressed were connected to the data upon which the opinions are based. No longer was an expert’s say-so appropriate. An expert had to show a relationship between reliable data and expressed opinion:

But conclusions and methodology are not entirely distinct from one another. Trained experts commonly extrapolate from existing data. But nothing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the ipse dixit² of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered. (p. 146)

In other words, the Court may not allow an expert to opine something simply because the expert “says it’s so.” There must be something more than just the expert’s word tying the data and the opinion. The focus of the *Daubert* Court, noted in Footnote 8 from the Court’s decision, was on scientific knowledge because that was the nature of the testimony offered into evidence in that case: “Rule 702 also applies to ‘technical, or other specialized knowledge.’ Our discussion is limited to the scientific context because that is the nature of the expertise offered here” (p. 590).

The third prong in what has come to be called the *Daubert* trilogy was a 1999 U.S. Supreme Court case that expanded the *Daubert* standard beyond scientific knowledge to include all expert testimony. In the case *Kumho Tire Co. v. Carmichael* (1999), the Court noted the following:

The *Daubert* “gatekeeping” obligation applies not only to “scientific” testimony, but to all expert testimony. Rule 702 does not distinguish between “scientific” knowledge and “technical” or “other specialized” knowledge, but makes clear that any such knowledge might become the subject of expert testimony. It is the Rule’s word

“knowledge,” not the words (like “scientific”) that modify that word, that establishes a standard of evidentiary reliability. *Daubert* referred only to “scientific” knowledge because that was the nature of the expertise there at issue. (p. 138)

We conclude that *Daubert*’s general holding—setting forth the trial judge’s general “gatekeeping” obligation—applies not only to testimony based on “scientific” knowledge, but also to testimony based on “technical” and “other specialized” knowledge. We also conclude that a trial court may consider one or more of the more specific factors that *Daubert* mentioned when doing so will help determine that testimony’s reliability. But, as the Court stated in *Daubert*, the test of reliability is “flexible,” and *Daubert*’s list of specific factors neither necessarily nor exclusively applies to all experts or in every case. Rather, the law grants a district court the same broad latitude when it decides how to determine reliability as it enjoys in respect to its ultimate reliability determination. (p. 141)

The Court clarified that the focus of attention on FRE 702 should be on the term “knowledge” rather than on “scientific” (see Table 2). The Court made clear that it was concerned about underlying reliable methodology as the foundation for expert testimony that is sound, reliable, and generally accepted and concerned about the opinion itself, whether such testimony came from a medical doctor or from a tire specialist. To borrow from the Clinton Presidential Campaign, “It’s the methodology, stupid.”

The *Daubert*, *Joiner*, and *Kumho* cases were clarifications of the FREs, and therefore did not apply directly to the states. Many states, however, have adopted the *Daubert* standard. Several states have continued their reliance on the Frye test, or an expanded version of the Frye test. California, for example, has the Kelly-Frye test (*People v. Kelly*, 1976) and Florida has the Ramirez-Frye test (*Ramirez v. State*, 1995). Both of these states have expanded the Frye test to determine if the methodology underlying the expert’s opinion is reliable, valid, and helpful to the Court.

The relevance of these case law precedents to child custody evaluations is that the methodology underlying the evaluator’s opinion must be reliable, relevant, and helpful to the court. Therefore, the prudent custody evaluator should select assessment tools that are both reliable and valid.

Test Selection and Relevance

In 1971, a case came before the U.S. Supreme Court that had nothing whatsoever to do with custody work, but the reverberations of which have been dramatically felt by evaluators. *Griggs et al. v. Duke Power Company* (1971) was a case involving procedures employed in the selection, placement, and promotion of personnel in an industrial setting. In deciding the case, the court ruled that any testing procedures must be demonstrably reasonable measures of (or predictors of) job performance. The lesson to be taken from the *Griggs* decision is that the selection of psychological tests must be reasonably linked to assessment of factors identified as the focus of the evaluation.

ETHICS, EXPERT TESTIMONY, AND THE SELECTION OF PSYCHOLOGICAL TESTS AND MEASURES

Rules of Evidence place a burden on psychologists—and other professions—to provide expert testimony that is reliable, relevant, and helpful. Expert testimony must reveal both a reliable methodology used in an evaluation and how the opinions drawn from the data derived from the use of the reliable methodology are connected to the data. Psychological ethics also place emphasis on reliability and relevance. For example, Section 2.04 of the American Psychological Association's (APA) *Ethical Principles of Psychologists and Code of Conduct* (APA, 2002; see also APA, 1992) describes that basis for scientific and professional judgments: "Psychologists' work is based upon established scientific and professional knowledge of the discipline." Psychologists also have a responsibility to keep up with changes in the field as noted in Section 2.03 (Maintaining Competence): "Psychologists undertake ongoing efforts to develop and maintain their competence" (p. 1064).

The APA Ethics Code section addressing use of psychological assessment techniques reveals a focus on reliability and relevance similar to the focus expressed in the evidence code standards discussed above. Similar to the *Joiner* concern about insuring that opinions are connected to reliable data, Standard 9.01(a) of the Ethics Code describes the need for psychologists to base their opinions on information and techniques sufficient to substantiate their findings: "(a) Psychologists base the opinions contained in their recommendations, reports, and diagnostic or evaluative statements, including forensic testimony, on information and techniques sufficient to substantiate their findings" (p. 1071; see also

Standard 2.04, Bases for Scientific and Professional Judgments). Parallel to the Supreme Court decision in *Griggs* cited above, the APA Ethics Code further admonishes psychologists to use tests that are relevant for the purpose for which it is intended to be used. Section 9.02 (Use of Assessments) states, “(a) Psychologists administer, adapt, score, interpret, or use assessment techniques, interviews, tests, or instruments in a manner and for purposes that are appropriate in light of the research on or evidence of the usefulness and proper application of the techniques” (p. 1071).

Remember that the lesson to be taken from the *Griggs* decision when applied to child custody work is that evaluators should focus their attention and their assessment efforts on functional abilities that bear directly upon the attributes, behaviors, attitudes, and skills that published research suggests are reliably associated with effective parenting and co-parenting. Examining an attribute in the absence of evidence of its connection to parenting effectiveness and related factors leaves a psychologist open to criticism on several fronts. For the custody evaluator, test selection and the data derived from the use of those tests must always be directly or indirectly addressing questions about parenting effectiveness, child development, or parent-child fit and co-parenting issues (Gould, this volume).

Admissibility of expert testimony is often dependent upon a showing that the methodology is reliable and that the opinions expressed by the expert are reasonably connected to the data. When psychologists select tests whose reliability and validity have not been established for use with members of the population tested, it is possible that legal standards of reliability and relevance would not permit testimony drawn from those tests to be admitted. The use of a test that has no demonstrated reliability and validity in the population for which it is being used may be viewed as an unreliable methodology. Opinions based upon unreliable methodology are, by definition, inadmissible. The requirement stated in Standard 9.02(b) to “describe the strengths and limitations of test results and interpretation” when “validity or reliability (of a test) has not been established” (APA, 2002, p. 1071) may be a critical component of any custody evaluation. The evaluator may need to explain how information drawn from a test of unknown reliability provides any probative value or how the presentation of information that appears to be scientifically derived yet is based upon an unreliable methodology is not “substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading (the court or a) . . . waste of time” (FRE 403, see Table 2).

Standard 9.06 might be viewed as parallel to concerns expressed in *Daubert* about scientific knowledge. *Daubert* was concerned, in part, about expert testimony based upon the notion that “an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation—i.e., ‘good grounds,’ based on what is known. In short, the requirement that an expert’s testimony pertain to ‘scientific knowledge’ establishes a standard of evidentiary reliability” (*Daubert* at 590).

I believe that the proper interpretation of psychological test data includes understanding test factors, test-taking abilities, and situational factors such as personal, linguistic, and cultural differences that might affect the accuracy of interpretations. There is a significant literature on test factors that may affect individual test responses in a forensic context. It is incumbent upon the evaluator to consider test factors and test-taking abilities when interpreting test results. It is also important to explain in the body of a report how each of these test factors may have affected the confidence in the meaning of the test data and the certainty of conclusions drawn from those data. Framed within the *Daubert* language, evaluators must apply the field’s scientific knowledge when interpreting psychological test data in order to increase the probative value of expert testimony. As noted in the *Daubert* (1993) decision, “Scientific methodology today is based on generating hypotheses and testing them . . .” (p. 593). Conveniently for custody evaluators, the appropriate use of psychological testing in child custody evaluations (or any situation, for that matter) specifically involves generating and testing hypotheses. Consistent with this position is Section 9.06 (Interpreting Assessment Results) of the Ethics Code (APA, 2002):

When interpreting assessment results, including automated interpretations, psychologists take into account the purpose of the assessment as well as the various test factors, test-taking abilities, and other characteristics of the person being assessed, such as situational, personal, linguistic, and cultural differences, that might affect psychologists’ judgments or reduce the accuracy of their interpretations. They indicate any significant limitations of their interpretations. (See also Standards 2.01b and c, Boundaries of Competence, and 3.01, Unfair Discrimination) (p. 1072)

Another Ethical Standard tied to expert testimony is Section 9.09 (Test Scoring and Interpretation Services):

- (a) Psychologists who offer assessment or scoring services to other professionals accurately describe the purpose, norms, validity, reliability, and applications of the procedures and any special qualifications applicable to their use.
- (b) Psychologists select scoring and interpretation services (including automated services) on the basis of evidence of the validity of the program and procedures as well as on other appropriate considerations. (See also Standard 2.01b and c, Boundaries of Competence.)
- (c) Psychologists retain responsibility for the appropriate application, interpretation, and use of assessment instruments, whether they score and interpret such tests themselves or use automated or other services. (p. 1072)

Standard 9.09(a) appears to parallel concerns about the reliability and relevance of testimony (see above). Courts may need to be informed about the purpose for selecting a particular test and how the interpretation of data from that test may be reasonably tied to the concerns before the court. A problem in the use of interpretive scoring programs provided by testing services is that the ethical criteria of 9.09(b) may be impossible to meet. Presently, the algorithms (i.e., program logic and decision rules) used to generate the statements in the computer generated test interpretations (CGTI) are proprietary secrets and not available for review by the evaluator. Therefore, it is not possible for evaluators to know how to answer important questions about how the program generates the statements found in CGTIs. This issue created enough concern that a letter, co-authored by three psychologists, was sent to the APA Ethics Committee for clarification. The response from the APA Ethics Committee Chairperson suggested two questions that psychologists should consider regarding the use of any CGTI program (Behnke, 2004): Given the purpose for which the service is utilized, what evidence of the program's validity do I require so that I may benefit, and not harm, my client? and What information about a program need I have in order to take responsibility for what my assessment contains? In addition to these two questions, I submit that evaluators might want to consider the following additional questions before using the interpretive statements offered on CGTI:

- Is the program an actuarial interpretation program or an automated interpretation program?
- What is the level of significance regarding the test scores?

- Are there different levels of significance for different scales?
- At what point does the program actually generate a statement for a particular scale?
- Are there different statements depending on the level of elevation for any given scale?
- Does the program take into consideration profile configurations or combinations of elevated scales, or are the statements based on single-scale elevation?
- Does the program take into consideration the response style when offering the CGTI statements?
- Does the program use context-specific normative data to facilitate interpretive statements?

These questions raise additional issues about what (and how much) “research” and/or “evidence of the usefulness” of a test is necessary and/or sufficient for use in a forensic evaluation in general, and more specifically, a child custody evaluation. This discussion is beyond the scope of the present article. Suffice it to say that the custodial evaluator should be prepared to address a variety of questions about the use of a CGTI report in the event the custody evaluation is challenged on legal or psychological grounds.

DEVELOPMENT OF CRITERIA FOR SELECTION AND USE OF PSYCHOLOGICAL TESTS IN CHILD CUSTODY CONTEXT

As mental health professionals have considered how best to assist courts in determinations of psychological and psychiatric issues, several authors have proposed criteria for the selection of psychological tests and measures used in court-related activities. Heilbrun (1992) developed a list of eight criteria that can be used by the evaluator in determining a test’s appropriateness for use in forensic evaluation. More recently, Otto and colleagues (2000) have taken Heilbrun’s eight-step model and adapted it specifically for the child custody context. The Heilbrun and Otto et al. models are quite similar, as can be seen in Table 3.

Depending on which model you choose, the answers to these questions can typically be found in the test manual and the relevant literature regarding the specific methodology. In addition, the literature is likely to include criticisms of a particular test or methodology. It is strongly recommended that copies of the relevant literature regarding specific assessment methodologies be readily available to the evaluator

TABLE 3. Model Criteria for Selection of Psychological Tests

Heilbrun (1992)	Otto et al. (2000)
Test must be commercially available	Is the test commercially published
Test must have published manual describing development, psychometric properties, and procedures for administration	Is a comprehensive test manual available
Test-retest reliability is at least 0.80	Are adequate levels of reliability demonstrated
There is ongoing research exploring its usefulness (validity)	Have adequate levels of validity been demonstrated
The test must be relevant to the legal issue or to a psychological construct underlying a legal issue	Is the test valid for the purposes in which it will be used
There is standard administration	What are the qualifications necessary to use this instrument
Test is reviewed in peer-reviewed journals	Has the instrument been peer reviewed
Test must have measures of response style	[The Otto et al. model does not address this issue] ⁵

for future reference, and also for defense of the evaluator's selection decisions. Critiques of the various methodologies used by the evaluator should also be included in this collection of literature. As can be imagined, the witness stand is the least desirable place to learn about a body of (substantial) literature that is highly critical of the chosen methodology.

It is important to remember that child custody evaluations take place in an adversarial legal system. It is not uncommon for the custody evaluator's report, opinions, and recommendations to be reviewed by another expert hired by the side dissatisfied with the report. This "battle of experts" can be unpleasant, and is not the appropriate forum to learn that

the psychological tests employed by the evaluator have been bitterly attacked or criticized in the literature. It is therefore important for the prudent custody evaluator to have a balanced understanding of the literature on a particular assessment methodology that includes both supportive and critical reviews.

There are resources available regarding critical reviews of most published psychological tests. For example, the *Mental Measurements Yearbook* and *Tests in Print* (both published by the Buros Institute of Mental Measurement; for more information, go to www.unl.edu/buros/) are two excellent sources for critical reviews. The literature also includes analyses of several psychological tests regarding their admissibility under different legal criteria (i.e., Frye test, Daubert's four-prong standard) and psychological issues of test selection. This includes Human Figure Drawings (Lally, 2001), the MCMI-III (McCann, 2002; see Rogers, Salekin, & Sewell, 2000, for an opposing view; and Dyer & McCann, 2000, for a reply), the MMPI-2 (see, e.g., Otto, 2002; Otto & Collins, 1995), the Rorschach (Gacono, Evans, & Viglione, 2002; McCann, 1998; for an opposing view, see, e.g., Grove & Barden, 1999; Grove, Barden, Garb, & Lilienfeld, 2002; Wood, Nezworski, Lilienfeld, & Garb, 2003; and for a reply see, e.g., Ritzler, Erard, & Pettigrew, 2002a, 2002b), and child custody-specific tests (e.g., the Bricklin scales, ASPECT, Parent-Child Relationship Inventory; see Ackerman, this volume; Connell, this volume; Otto et al., 2000; Yañez & Fremouw, 2004) and other parenting assessment instruments (i.e., Child Abuse Potential Inventory, Parenting Stress Index; Yañez & Fremouw, 2004).

Psychological test usage in the child custody context has not gone without criticism (Brodzinsky, 1993; Grisso, 1986, 2003; Melton, Petrilla, Poythress, & Slobogin, 1997). Early criticism involved inappropriate use of tests and diagnostic impressions that were misleading and pejorative (Grisso, 1986, 2003). Other criticisms involved overutilization of psychological tests without psycho-legal relevance (Brodzinsky, 1993; Melton et al., 1997). It would be appropriate for the responsible and competent evaluator to have an awareness of the literature regarding the pros and cons of test usage. Recent research, however, has found that current child custody evaluation practices do not support these concerns (Ackerman & Ackerman, 1997; Bow & Quinnell, 2001, 2002; Horvath, Logan, & Walker, 2002; Quinnell & Bow, 2001).

Several studies have looked at assessment methodology within the child custody context. These studies can be divided into two types:

those that utilized survey methodology (Ackerman & Ackerman, 1997; Bow & Quinnell, 2001; Karras & Berry, 1985; Keilin & Bloom, 1986; Quinnell & Bow, 2001) and those that reviewed the methodology contained within child custody reports (Bow & Quinnell, 2002; Horvath et al., 2002). The first study was conducted by Keilen and Bloom almost 20 years ago. The results of their survey showed that approximately 75% of respondents used testing with parents and children. The MMPI-2 was the most common test utilized with parents, used by approximately two-thirds of the respondents. This study serves as the jumping off point for all subsequent studies investigating psychological test usage in child custody evaluations.

In a follow-up to the Keilen and Bloom (1986) study, Ackerman and Ackerman (1997) found that custody evaluators spent an average of 5.2 hours doing psychological testing. Tables 2 and 3 [pp. 139-140] of their study show the frequency of test usage with children and adults. The Hagen and Castagna (2001) study raised an issue about the concept of “standard of practice” based on the Ackerman and Ackerman data. In a reanalysis of the Ackerman data, the authors found that none of the tests—except the MMPI-2—reached a level of usage consistent with a “standard of practice.” LaFortune and Carpenter (1998) made several conclusions from their data, the most concerning of which is the following:

Even with this large number of diverse findings, a number of general conclusions emerge from the data. . . . Eighth, although tests play a smaller role than interviews and observation, many experts employ procedures with little or no know [sic] valid basis for informing custody decisions. (p. 221)

Quinnell and Bow (2001) compared the results of their survey study to the Ackerman and Ackerman and the Keilen and Bloom studies, and noted the following:

First, participants in the study only ranked psychological testing as moderately important (fourth and sixth) among ten main custody evaluation procedures. . . . These findings suggest that psychological testing is no longer the primary procedure in custody evaluations; but instead is used to supplement other procedures or to create “working hypotheses,” as defined by Heilbrun (1995). (p. 498)

Otto et al. (2000) noted the following regarding child custody-specific tests (e.g., Bricklin Scales, ASPECT) after reviewing the above noted studies:

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 their appropriateness for use in custody evaluations at the present time. (p. 317)

After reviewing these child custody-specific tests (e.g., Bricklin Scales, ASPECT), Otto et al. noted:

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. (p. 336)

In the first of two content analysis studies, Horvath et al. (2002) noted a somewhat surprising finding that custody evaluators may actually not use enough psychological testing in child custody evaluation cases. Specifically, they noted:

On the basis of this review of custody evaluations and others (Logan, Walker, Jordan, & Horvath, in press), we found that there are clearly a few areas frequently neglected by evaluators, including assessment of domestic violence and child abuse, adequate assessment of parenting skills, assessment of health status, formal psychological testing, and using multiple methods of information gathering. (p. 563)

In addition to the need for standardized interviews, it also appears that evaluators need to be encouraged to include psychological testing or behavioral assessment instruments in their evaluations. There is a substantial risk to the intended objectivity of child custody recommendations when there are no independent anchors for opinions such as those that can be obtained through the use of validated instruments (p. 563). In the second content analysis study, Bow and Quinnell (2002) found similar results to those of the Quinnell and Bow (2001) study, and noted, "In general, psychological test find-

ings were not given undue weight and reviewed as one data source” (p. 174).

Suggested Models of Test Selection

There are essentially two different models for the selection of psychological tests in a child custody evaluation. The descriptions of these two models are intentionally somewhat exaggerated and polarized for the sake of discussion. In the Scientist-Practitioner Model,³ the evaluator selects testing methodologies based on the psycho-legal issues involved in the specific case and the psychometric qualities of the particular tests. Using the Scientist-Practitioner Model permits the evaluator to answer questions regarding a test’s appropriateness for use in evaluation from both legal (i.e., relevance and helpfulness) and psychological perspectives (i.e., reliability and validity). Alternatively, in the Lemming Syndrome Model the evaluator selects testing methodologies based on the frequency of test usage described in the literature. The Lemming Syndrome Model allows the evaluator to answer questions regarding a test’s appropriateness with the statement, “The research says that lots of people use it.” This statement is offered by the evaluator to suggest that the reliability and validity of a test is both measured and increased by popular vote.

It is important to remember that a psychological test has little value if it does not measure something consistently or accurately. As noted above, both the Heilbrun (1992) and Otto et al. (2000) models recommend that a test should be commercially available/published and have an accompanying test manual. These two criteria, however, do not guarantee that the test is reliable, valid in general, or valid for specific use in a child custody evaluation. Publishers of psychological tests are in the business of making money. Colorful brochures and catchy phrases are marketing tools that should have no bearing on an evaluator’s decision to use a test. As noted by Martindale (2001), “The frequency with which certain instruments are utilized may be attributable more to marketing and related phenomena than the psychometric integrity” (p. 500).

CONCLUSION

The decision to use psychological testing in the child custody context is a complex matter that involves both an understanding of and integration with legal and psychological concepts. The issues of relevance and

helpfulness are two essential concepts regarding the admissibility of evidence in the legal arena. This is true with all expert testimony and not just that of the child custody evaluator. The child custody evaluation report, underlying methodology, and ultimate testimony are all considered evidence. Therefore, the report, the methodology underlying the evaluation, and the custody evaluator's testimony are all subject to the evidence code of a particular jurisdiction. As such, we can only provide information to the Court if the evidence is relevant to the issue at hand and helpful to the Court. Although this is not something we get to decide, our decisions about the selection of psychological testing can make it easier—or less likely—for the Court to accept our work product. Ensuring our methods and procedures comport with the essential elements of the law allows our work products and testimony to aid the Court in resolving these challenging child custody matters. Relevant and helpful evaluations do not, however, guarantee the expert's work product and testimony will go unchallenged on other legal (and psychological) grounds.

Equally as important to the legal issues of relevance and helpfulness are the psychological issues associated with test selection. Although these issues were discussed following the discussion of legal issues, the psychological issues involved in test selection I believe are of even more importance. Undoubtedly, the use of unreliable or invalid data collection methods cannot only undermine or compromise the evaluator's findings, but also raise legal challenges to the admissibility of the work product and resulting expert testimony. The use of reliable and valid psychological testing is one of multiple procedures underlying the evaluator's methodology. A review of the findings from the Daubert trilogy made it clear that the focus of the Court's inquiry into helpfulness will be the methodology underlying the expert's opinion. Using reliable and valid psychological testing is unarguably important to ensure the methodology underlying the work product passes muster with legal issues of relevance and helpfulness and psychological issues of reliability and validity.

The psychometric issues of reliability and validity are essential ingredients in the value of any psychological test. Value, here, can be measured by both how consistent (i.e., reliable) and accurate (i.e., valid) the specific assessment methodologies are as data collection methods. The APA (2002) Ethics Code speak to the issues of reliable and valid test instruments, as do writers addressing the use of psychological testing in a forensic context. The attention to the selection of reliable and valid assessment instruments transcends philosophical arguments about

the (alleged) differences in models used by child custody evaluators. This appears to be a more simplistic argument: opinions and recommendations based on inconsistent and inaccurate data are of little or no value to anyone, let alone the Court. The statement by Otto et al. (2000) noted above deserves repeating at this point: “[T]he 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” (p. 33).

The use of an unreliable assessment method provides inaccurate data and erroneous opinions and recommendations that form the basis of the evaluator’s work product. Testimony resulting from this data, opinions, and recommendations will be misleading and unhelpful to the Court, the parents in dispute, and the child at issue. Therefore, the selection of both reliable and valid assessment instruments is crucial to developing a solid foundation from which the opinions and recommendation of the evaluator are based, and presented in the form of the work product or testimony.

NOTES

1. Psychometric qualities refer to the statistical properties of a test.
2. ipse dixit: “he himself said it; an assertion by one whose sole authority for it is the fact that he himself said it.” Gifis, S. H. (1991). *Law Dictionary* (3rd ed., p. 252). Hauppauge, NY: Barron’s Educational Series, Inc.
3. The term “Scientist Practitioner” is used with permission from Katherine Kuehnle, PhD. See, e.g., Kuehnle (1996, 1998).
4. *Michigan Child Custody Act of 1970*, MCL § 722.23 (1993 amended)
5. Apparently, this was an oversight, as the authors consider the need for response style measures essential in forensic assessment methodology (R. Otto, personal communication, October 6, 2004).

REFERENCES

- Ackerman, M. J. (2001). *Clinician’s guide to child custody evaluations* (2nd ed.). New York: John Wiley & Sons, Inc.
- Ackerman, M. J. (2005). The Ackerman-Schoendorf Scales for Parent Evaluation of Custody (ASPECT): A review of research and update. *Journal of Child Custody*, 2(1/2), 179-193.
- Ackerman, M. J., & Ackerman, M. C. (1997). Custody evaluation practices: A survey of experienced professionals (revised). *Professional Psychology: Research and Practice*, 28(2), 137-145.

- 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*(12), 1597-1611.
- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, *57*(12), 1060-1073.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Behnke, S. (2004). Test-scoring and interpretation services. *APA Monitor*, *35*(3), 58-59.
- Borum, R., Otto, R., & Golding, S. (1993). Improving clinical judgment and decision making in forensic evaluation. *The Journal of Psychiatry & Law*, *21*, 35-76.
- Bow, J. N., & Quinnell, F. A. (2001). Psychologists current practices and procedures in child custody evaluation: Five years after American Psychological Association guidelines. *Professional Psychology: Research and Practice*, *32*(3), 261-268.
- Bow, J. N., & Quinnell, F. A. (2002). A critical review of child custody evaluation reports. *Family Court Review*, *40*(2), 164-176.
- Brodzinsky, D. M. (1993). On the use and misuse of psychological testing in child custody evaluations. *Professional Psychology: Research and Practice*, *24*(2), 213-219.
- Condie, L. O. (2003). *Parenting evaluations for the court. Care and protection matters*. New York: Kluwer Academic/Plenum Publishers.
- Congress & Empire Spring Co. v. Edgar, 99 U.S. 645 (1878).
- Connell, M. (2005). Review of "The Ackerman-Schoendorf Scales for Parent Evaluation of Custody" (ASPECT). *Journal of Child Custody*, *2*(1/2), 195-209.
- Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).
- Dyer, F. J., & McCann, J. T. (2000). The Millon clinical inventories, research critical of their forensic application and Daubert criteria. *Law & Human Behavior*, *24*(4), 487-497.
- Friedman, A. F., Lewak, R., Nichols, D. S., & Webb, J. T. (2001). *Psychological assessment with the MMPI-2*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Frye v. U.S., 293 F. 1013 (D.C. Cir. 1923).
- Gacono, C. B., Evans, F. B., & Viglione, D. J. (2002). The Rorschach in forensic practice. *Journal of Forensic Psychology Practice*, *2*(3), 33-53.
- General Electric Co. v. Joiner, 522 U.S. 136 (1997).
- Goodman-Delahunty, J. (1997). Forensic psychological expertise in the wake of *Daubert*. *Law and Human Behavior*, *21*(2), 121-140.
- Gould, J. W. (1998). *Conducting scientifically crafted child custody evaluations*. Thousand Oaks, CA: Sage Publication.
- Gould, J. W. (1999). Scientifically crafted child custody evaluations. Part two: A paradigm for forensic evaluation of child custody determination. *Family and Conciliation Courts Review*, *37*(2), 159-178.
- Gould, J. W. (2005). Use of psychological tests in child custody assessment. *Journal of Child Custody*, *2*(1/2), 49-69.
- Greene, R. L. (2000). *MMPI-2/MMPI: An interpretive manual* (2nd ed.). Boston: Allyn & Bacon.
- Griggs et al. v. Duke Power Company, 401 U.S. 424 (1971).

- Grisso, T. (1986). *Assessing competencies: Forensic assessment and instruments*. New York: Plenum Press.
- Grisso, T. (2003). *Assessing competencies: Forensic assessment and instruments* (2nd Ed.). New York: Plenum Press.
- Grove, W. M., & Barden, R. C. (1999). Protecting the integrity of the legal system: The admissibility of testimony from mental health experts under Daubert/Kumho analyses. *Psychology, Public Policy, and Law*, 5(1), 224-242.
- Grove, W. M., Barden, R. C., Garb, H. N., & Lilienfeld, S. O. (2002). Failure of Rorschach-comprehensive-system-based testimony to be admissible under the Daubert-Joiner-Kumho standard. *Psychology, Public Policy, and Law*, 8(2), 216-234.
- Hagen, M. A., & Castagna, N. (2001). The real numbers: Psychological testing in child custody evaluations. *Professional Psychology: Research and Practice*, 32(3), 269-271.
- Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior*, 16(3), 257-272.
- Heilbrun, K. (1995). Child custody evaluation: Critically assessing mental health experts and psychological tests. *Family Law Quarterly*, 29, 63-78.
- Heilbrun, K. (2001). *Principles of forensic mental health assessment*. New York: Kluwer Academic/Plenum Publishers.
- Horvath, L. S., Logan, T. K., & Walker, R. (2002). Child custody cases: A content analysis of evaluations and practice. *Professional Psychology: Research and Practice*, 33(6), 557-565.
- Karras, D., & Berry, K. K. (1985). Thus evaluations: A critical review. *Professional Psychology: Research and Practice*, 16(1), 76-85.
- Keilin, W. G., & Bloom, L. J. (1986). Child custody evaluation practices: A survey of experienced professionals. *Professional Psychology: Research and Practice*, 17, 338-346.
- Kirkpatrick, H. D. (2003). A floor, not a ceiling: Beyond guidelines—an argument for minimum standards of practice in conducting child custody and visitation evaluations. *Journal of Child Custody*, 1(1), 61-75.
- Krauss, D. A., & Sales, B. D. (1999). The problem of “helpfulness” in applying Daubert to expert testimony: Child custody determinations in family law as an exemplar. *Psychology, Public Policy, and Law*, 5(1), 78-99.
- Kuehnle, K. (1998). Child sexual abuse evaluations: The scientist-practitioner model. *Behavioral Sciences and the Law*, 16(1), 5-20.
- Kuehnle, K. (1996). *Assessing allegations of child sexual abuse*. Sarasota, FL: Professional Resource Press.
- Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).
- LaFortune, K. A., & Carpenter, B. N. (1998). Custody evaluations: A survey of mental health professionals. *Behavioral Sciences and the Law*, 16, 207-224.
- Lally, S. J. (2001). Should human figure drawings be admitted into court? *Journal of Personality Assessment*, 76(1), 135-149.
- Martindale, D. A. (2001). Cross-examining mental health experts in child custody litigation. *The Journal of Psychiatry & Law*, 29, 483-511.
- Martindale, D. A. (2004, October 6). *Child custody evaluations: Legal, clinical, & ethical issues*. Department of Mental Health Law & Policy, Louis de la Parte Florida Mental Health Institute, University of South Florida.

- Martindale, D. A. (2005). Confirmatory bias and confirmatory distortion. *Journal of Child Custody*, 2(1/2), 31-48.
- Martindale, D. M., & Gould, J. W. (2004). The forensic model: Ethics and scientific methodology applied to custody evaluations. *Journal of Child Custody*, 1(2), 1-22.
- McCann, J. T. (1998). Defending the Rorschach in court: An analysis of admissibility using legal and professional standards. *Journal of Personality Assessment*, 70(1), 125-144.
- McCann, J. T. (2002). Guidelines for forensic application of the MCMI-III. *Journal of Forensic Psychology Practice*, 2(3), 55-69.
- McDaniel v. CSX Transportation, Inc., 955 S.W.2d 257 (1997).
- Melton, G. B., Petrilla, J., Poythress, N. G., & Slobogin, C. (1997). *Psychological evaluations for the courts* (2nd ed.). New York: Guilford Press.
- Mize v. Mize, 621 So.2d 417 (Fla. 1993).
- Otto, R. K. (2002). Use of the MMPI-2 in forensic settings. *Journal of Forensic Psychology Practice*, 2(3), 71-91.
- Otto, R. K., Buffington-Vollum, J. K., & Edens, J. F. (2003). Child custody evaluation. In I. B. Weiner (Series Ed.) & A. M. Goldstein (Vol. Ed.), *Handbook of psychology: Vol. 11 Forensic psychology* (pp. 179-208). New York: John Wiley & Sons, Inc.
- Otto, R. K., & Collins, R. P. (1995). Use of the MMPI-2/MMPI-A in child custody evaluations. In S. E. Hobfoll (Series Ed.) & Y. S. Ben-Porath, J. R. Graham, G. C. N. Hall, R. D. Hirschman, & M. S. Zaragoza (Vol. Eds.), *Applied psychology: Individual, social, and community issues: Vol. 2. Forensic applications of the MMPI-2* (pp. 222-252). New York: John Wiley & Sons, Inc.
- Otto, R. K., Edens, J. F., & Barcus, E. H. (2000). The use of psychological testing in child custody evaluations. *Family and Conciliation Courts Review*, 38(3), 312-340.
- Paulhus, D. L. (1998). *Paulhus Deception Scales (PDS): The Balanced Inventory of Desirable Responding-7*. North Tonawanda, NY: MHS.
- People v. Kelly, 549 P.2d 1240 (1976).
- Quinnell, F. A., & Bow, J. N. (2001). Psychological tests used in child custody evaluations. *Behavioral Sciences in the Law*, 19, 491-501.
- Ramirez v. State, 651 So. 2d 1164, 1995.
- Ritzler, B., Erard, R., & Pettigrew, G. (2002a). Protecting the integrity of Rorschach expert witnesses: A reply to Grove and Barden (1999) re: The admissibility of testimony under Daubert/Kumho analyses. *Psychology, Public Policy, and Law*, 8(2), 201-215.
- Ritzler, B., Erard, R., & Pettigrew, G. (2002b). A final reply to Grove and Barden: The relevance of the Rorschach comprehensive system for expert testimony. *Psychology, Public Policy, and Law*, 8(2), 235-246.
- Rogers, R., Salekin, R. T., & Sewell, K. W. (2000). The MCMI-III and the Daubert standard: Separating rhetoric from reality. *Law & Human Behavior*, 24(4), 501-506.
- Russenberger v. Russenberger, 669 So.2d 1044 (Fla. 1996).
- Schutz, B. M., Dixon, E. B., Lindenberger, J., & Ruther, N. J. (1989). *Solomon's sword. A practical guide to conducting child custody evaluations*. San Francisco: Jossey-Bass Publications.
- Shuman, D. W., & Sales, B. D. (1999). The impact of *Daubert* and its progeny on the admissibility of behavioral and social science evidence. *Psychology, Public Policy and Law*, 5(1), 3-15.

- Stahl, P. M. (1994). *Conducting child custody evaluations. A comprehensive guide*. Thousand Oaks, CA: Sage Publication.
- Winans v. New York & Erie Railroad Co., 62 U.S. 88, (1858).
- Wood, J. M., Nezworski, M. T., Lilienfeld, S. O., & Garb, H. N. (2003). *What's wrong with the Rorschach?* San Francisco, CA: Jossey-Bass.
- Woody, R. H. (2000). *Child custody. Practice standards, ethical issues & legal safeguards for mental health professionals*. Sarasota, FL: Professional Resource Press.
- Yañez, Y. T., & Fremouw, W. (2004). The application of the Daubert standard to parental capacity measures. *American Journal of Forensic Psychology*, 22(3), 5-28.