

Individual Adult Psychometric Testing and Child Custody Evaluations: If the Shoe Doesn't Fit, Don't Wear It

by

Benjamin D. Garber* and Robert A. Simon**

Abstract

Individual adult psychometric instruments are commonly included in child custody evaluations. The data thus derived are often given great weight by evaluators, guardians ad litem, attorneys, and judicial officers. Despite these instruments' utility in other applications, they are generally not normed with litigating parents, have not established reliability for this application, and therefore cannot provide valid data to the custody determination process. Indeed, the questions to be addressed through child custody evaluation are not about the quality of individual adult functioning, but are, instead, focused on understanding the quality of specific relationship dynamics. The present article recommends that evaluators reconsider the inclusion of individual adult psychometric instruments in child custody evaluations and that the consumers of such reports be better prepared to challenge the admissibility and relevance of the data thus obtained. Recommendations for the development of standardized, relationship- and system-based assessment methodologies and instruments are provided.

"How fair does a trial seem to the public where the defendant stands up and says, 'Your Honor, I want to represent myself? I do not want this attorney. I want to defend myself.' And the judge said, sit down, we have a psychological evaluation of you."¹

Contemporary medical imaging is a life-saving science. CAT scan, MRI, X-ray and ultrasound, for example, are highly reliable and valid means of detecting otherwise invisible structural and functional pathologies. Thus, diagnostic evidence derived via these technologies is commonly admitted in American courts.

Reliability and validity in one application does not guarantee reliability or validity in all applications. Functional Magnetic

* Ph.D., HealthyParent.com, Family Law Consulting, PLLC.

** Ph.D., Forensic Psychology Consulting, www.dr-simon.com.

¹ Transcript of Oral Argument at 7, *Indiana v. Edwards*, 554 U.S. 164 (2008) (Scalia, J.).

Resonance Imaging (fMRI), for example, is a reliable and valid means of assessing cochlear implant candidates,² but is not reliable (and therefore cannot be valid) as a means of detecting lies.³ As a result, the federal courts have commonly allowed fMRI evidence in the former application but precluded it in the latter.⁴

We contend that the same is true of contemporary individual adult psychometric instruments (hereinafter, “tests”⁵). That is, reliable and valid tests can provide helpful data when properly administered with the populations upon which they have been normed and in the clinical settings for which they have been developed, but must not be presumed to be either reliable or valid when administered in the context of contested custody litigation.⁶ With this in mind, we argue that child custody evaluations that rely upon test data risk misleading the court, breaching relevant ethical rules, creating unnecessary, time-consuming and expensive legal straw-men, and doing harm to families and to the vulnerable children whose best interests the courts must serve.

What follows is a critique of the inclusion of tests in child custody evaluations (CCE). Just as the child custody evaluator is charged to examine the “fit” between parental limitations and capacities and the children’s needs,⁷ we address the “fit” between these instruments’ limitations and capacities and the needs of CCE. We conclude that the fit is poor and, therefore, that individual adult psychometrics have little or no place in the process of evaluating family dynamics.

² Andrea Joachim Bartsch et al., *Diagnostic Functional MRI: Illustrated Clinical Applications and Decision-Making*, 23 J. MAGNETIC RESONANCE IMAGING 921 (2006).

³ Daniel D. Langleben & Jane Campbell Moriarty, *Using Brain Imaging for Lie Detection: Where Science, Law and Research Policy Collide*, 19 PSYCHOL., PUB. POL’Y, & L. 222 (2013).

⁴ For example, in *United States v. Semrau*, No. 07-10074 ML/P, 2010 WL 6845092 (W.D. Tenn. June 1, 2010), the federal court excluded fMRI testimony regarding a defendant’s veracity.

⁵ The law defines a test as an instrument that, “ascertains the truth or the quality or fitness of a thing” BLACK’S LAW DICTIONARY 1473 (6th ed. 1990).

⁶ Randy K. Otto et al., *The Use of Psychological Testing in Child Custody Evaluations*, 38 FAM. & CONCILIATION CTS. REV. 312 (2000).

⁷ American Psychological Association, *Guidelines for Child Custody Evaluations in Divorce Proceedings*, 49 AM. PSYCHOL. 677 (1994).

We submit that the subject of a CCE is the dynamic and developing family system, not the individual and, on this basis, that measures of individual functioning are largely irrelevant. This paradigm shift requires most evaluators to retool and reconsider process, product, and presentation.

We submit that a CCE report that includes intimidating and invalid means, percentile ranks, and alpha coefficients risks misleading consumers and harming the children whom we intend to serve.⁸ This risk is compounded when evaluators blindly rely exclusively on computer-generated scoring and interpretations.⁹

Finally, we submit that it is an error to mistake common practice for correct practice.¹⁰ The fact that a majority of today's custody evaluators include individual psychometric adult instruments in CCEs, thereby establishing a "community standard" is not reason to do the same. We recommend instead that, pending the development of reliable and valid measures of systemic functioning, custody evaluators eschew the use of individual adult psychometric tests in favor of cautiously and carefully crafted, systemically-informed, empirically-grounded, and child-centered *qualitative* observations, inferences and recommendations.

I. What Are Individual Adult Psychometric Instruments (Tests)?

Historically, the field of psychology distinguished itself from the field of psychiatry and from other professional mental health

⁸ Allan Postuma, *Current and New Developments in Psychological Testing for Child Custody Disputes*, in HANDBOOK OF CHILD CUSTODY 67 (Mark L. Goldstein ed. 2016) ("The use of psychological testing in these assessments can exacerbate extreme adversarial positions"). See also Steven K. Erickson et al., *A Critical Examination of the Suitability and Limitations of Psychological Tests in Family Court*, 45 FAM. CT. REV. 157 (2007); Stephen D. Hart et al., *Precision of Actuarial Risk Assessment Instruments*, 190 BRIT. J. PSYCHIATRY 60 (2007).

⁹ The Association of Family and Conciliation Courts Model Standards caution that "[e]valuators shall recognize that test data carry an aura of precision that may be misleading," and therefore strongly advise against the use of computer generated test interpretation for both ethical evidentiary reasons. Association of Family and Conciliation Courts, *Model Standards of Practice for Child Custody Evaluation*, 45 FAM. CT. REV. 70 (2007).

¹⁰ Keeping in mind that common practice may meet admissibility standards under *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923), but not under *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993), as discussed below.

fields through the development, administration, and interpretation of tests designed to measure psychological functioning. Alfred Binet, for example, was among the first psychometricians. His Benet-Simon Test¹¹ (known today as the Stanford-Binet Intelligence Test, now in its fifth edition) remains a staple among psychometricians.¹²

Tests must be distinguished from questionnaires.¹³ Whereas tests are subject to formal interpretation and generate diagnoses, questionnaires can be standardized, economical, and time-efficient means of collecting data and generating hypotheses.¹⁴

The field of psychometric assessment has become a thriving commercial industry.¹⁵ The chief contemporary compendium of psychological tests includes reviews of three thousand, five hundred instruments.¹⁶ Although the purpose, nature, means of administration, process of interpretation, reliability, and validity of these instruments vary immensely, they all share certain qualities:

1. *Norms.*

Psychometric instruments are typically developed by administering the instrument to one or more demographically distinct samples. On this basis, population-specific typical and atypical responses are determined. Thus, it is meaningful to draw inferences about a particular 20-year-old Caucasian male college student using a test that was normed using 20-year-old Caucasian, male, college students, but it is entirely unclear what the

¹¹ No known relationship to the second author.

¹² Kirk A. Becker, *History of the Stanford Binet Intelligence Scales: Content and Psychometrics*, Stanford-Binet Intelligence Scales, Fifth Edition: Assessment Service Bull. No. 1 (2003), http://www.hmhco.com/~media/sites/home/hmh-assessments/clinical/stanford-binet/pdf/sb5_asb_1.pdf?la=EN.

¹³ Leigh D. Hagan & Ann C. Hagan, *Custody Evaluations Without Psychological Testing: Prudent Practice or Fatal Flaw?*, 36 J. PSYCHIATRY & L. 67 (2008).

¹⁴ BENJAMIN D. GARBER, *TEN CHILD-CENTERED FORENSIC FAMILY EVALUATION TOOLS: AN EMPIRICALLY ANNOTATED USERS' GUIDE* (2015).

¹⁵ Jonathan Rich, *Psychological Testing: Old Specialty, New Markets*, NAT'L PSYCHOL. (July 1, 2007), <http://nationalpsychologist.com/2007/07/psychological-testing-old-specialty-new-markets/10933.html>.

¹⁶ THE TWENTIETH MENTAL MEASUREMENTS YEARBOOK 103 (Janet F. Carlson et al., eds. 2017).

results of the same test mean when administered to a 50-year-old, female, African-American high school drop-out. In short, Philip Stahl and Robert Simon advise that “psychological tests are best used with the reference population with which the test was normed.”¹⁷

The American Psychological Association explicitly requires that psychologists must “use assessment instruments whose validity and reliability have been established for use with members of the population tested.”¹⁸ With this in mind, it is at least confusing to realize that the Millon Clinical Multiaxial Inventory - the second most commonly administered test among custody evaluators¹⁹ —is normed “entirely on clinical samples and [is] applicable only to individuals who evidence problematic emotional and interpersonal symptoms or who are undergoing professional psychotherapy or a psychodiagnostic evaluation.”²⁰

The fact is that we don’t know, *a priori*, who child custody litigants are until they become litigants. Experience suggests that they come from all socio-economic strata, all racial, religious, language and cultural groups, and span the full range of ages, IQs, sizes, shapes, and colors. They obviously share the experience of parenting (although some are not biological parents) and typically a contentious (if not antagonistic) attitude toward a former parenting partner. Unless and until this population is better understood and instruments are developed that represent its normative thinking, feeling, and behavior, it is simply misleading to

¹⁷ PHILIP M. STAHL & ROBERT A. SIMON, *FORENSIC PSYCHOLOGY CONSULTATION IN CHILD CUSTODY LITIGATION* 124 (2013).

¹⁸ AMERICAN PSYCHOLOGICAL ASSOCIATION, *ETHICAL PRINCIPLES OF PSYCHOLOGISTS AND CODE OF CONDUCT*, Principle 9.02[b] (2010), <http://www.apa.org/ethics/code/index.aspx>.

¹⁹ James N. Bow et al., *MMPI-2 and MCMI-III in Forensic Evaluations: A Survey of Psychologists*, 10 J. FORENSIC PSYCHOL. PRAC. 37 (2010).

²⁰ THE MILLON INVENTORIES, SECOND EDITION: A PRACTITIONER’S GUIDE TO PERSONALIZED CLINICAL ASSESSMENT 76 (Theodore Millon & Caryl Bloom, eds. 2008). Even though test scoring and interpretation services such as The Caldwell Report and Pearson Assessments publish a so-called child custody interpretation, these publishers do not clearly and plainly note that the interpretations in their reports are not empirically and independently derived. They are, instead, interpolations based on the insight and wisdom of some very learned and accomplished psychometricians, but their intellect does not replace norms that are empirically derived from a broad and representative population of child custody litigants.

make statements about a custody litigant by comparison to the responses of other entirely distinct normative groups.

2. *Reliability.*

Test data are meaningless until they have been shown to be stable across time, contexts, and/or administrators. Reliability describes these indices. For example, we know that Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) measures of adult intelligence achieve reliability coefficients approaching .90 over two to twelve week periods and across evaluators.²¹ We do not know, however, whether any particular test is reliable among custody litigants.

Of particular relevance to the question of test reliability among custody litigants is the larger question of the stability of the population in general. Assessing people in the midst of crisis tends not to capture their typical functioning. The intense social, emotional, and financial pressures associated with contested custody litigation can induce or exacerbate acute and reactive anxiety, anger, and regression among otherwise healthy and high functioning adults.²² Thus, it remains to be seen whether a properly normed test could demonstrate test-retest reliability across the period before, during, and after the close of litigation.

3. *Validity.*

Reliability is necessary but it is not sufficient. A test that is reliable must still be shown to actually measure the construct that it purports to measure. For example, a tape measure is a very reliable means of measuring height. That measure may be moderately meaningful as a means of inferring age (that is, adults tend to be taller than children), but is likely an entirely invalid means of measuring intelligence.

Validity can be demonstrated in a number of ways. The measure can be compared to measures obtained using other, established instruments (concurrent validity), or by comparing test

²¹ DAVID WECHSLER, WECHSLER ADULT INTELLIGENCE SCALE—FOURTH EDITION ADMINISTRATION AND SCORING MANUAL 47-48 (2008).

²² Meredith S. Sears et al., *Spillover in the Home: The Effects of Family Conflict on Parents' Behavior*, 78 J. MARRIAGE & FAM. 127 (2016).

data with concurrent or predicted behavior (construct or predictive validity).

Even given a properly normed and reliable instrument, professionals who conduct CCEs are still faced with the mountainous problem of criterion validity. That is, if the criterion against which a test is to be validated is the “best interests of the child,” but this criterion cannot be generically defined, then the task would appear to be impossible.

II. What Is Child Custody Evaluation (CCE)?

CCE is the assessment of the psychosocial variables that characterize a family system and bear upon the well-being of the child or children. CCE is usually conducted under court order when parents litigate the future allocation of their parenting rights and responsibilities.

CCE is a relatively new development in the history of contested custody litigation. For generations, courts have relied upon one-size-fits-all, gender-specific heuristics to determine post-divorce custody rights.²³ Under English common law, for example, children were considered their father’s property. More recently, the tender years doctrine determined that children under a certain age belonged in their mother’s care.²⁴ It is only in the last fifty years with the adoption of the best interests of the child standard that the courts have begun to consider each child’s individual needs. It is in this context that CCE was born.

CCE is an empirically-driven, child-centered examination of the family system. It serves to provide the court with an understanding of how best to understand the unique needs of each child. The CCE evaluator seeks to provide the court with observations, inferences, and—depending upon jurisdictional mandates and the court’s specific order—recommendations relative to the future allocation of parenting rights and responsibilities.

Contemporary standards and guidelines relevant to the conduct of CCE call for the evaluator to employ multiple and diverse methods so as to consider “alternative plausible hypothe-

²³ Richard A. Warshak, *Securing Children’s Best Interests While Resisting the Lure of Simple Solutions*, 56 J. DIVORCE & REMARRIAGE 57 (2015).

²⁴ J. Herbie DiFonzo, *From the Rule of One to Shared Parenting: Custody Presumptions in Law and Policy*, 52 FAM. CT. REV. 213 (2014).

ses.”²⁵ In practice, survey data find that evaluators commonly include individual adult and child interviews, individual, dyadic and/or systemic observations, a review of relevant historical documents, personal references, and individual adult psychometric testing.²⁶

James Bow and Francella Quinnell report that 91% of surveyed evaluators commonly include individual adult psychometric tests in CCEs.²⁷ Marc Ackerman and Tracy Brey Pritzl observed among CCE evaluators that “more testing is [being] performed more frequently and a wider variety of tests are [being] administered than was the case ten years ago.”²⁸

This apparent trend toward more universal and varied adult testing is despite the fact that none of the relevant CCE standards or guidelines explicitly require the inclusion of testing,²⁹ and despite the fact that psychologist-evaluators are responsible

²⁵ See, e.g., American Academy of Matrimonial Lawyers, *Child Custody Evaluation Standards* 4.3(a) (2011), <http://www.aaml.org/library/publications/child-custody-evaluation-standards-doc>; see also AMERICAN PSYCHOLOGICAL ASSOCIATION, *supra* note 18, Principle 10; Association of Family and Conciliation Courts, *Model Standards of Practice for Child Custody Evaluation*, 45 FAM. CT. REV. 70, at 79, Principle 5.4 (2007).

²⁶ Marc J. Ackerman & Melissa C. Ackerman, *Custody Evaluation Practices: A Survey of Experienced Professionals (Revisited)*, 28 PROF. PSYCHOL.: RES. & PRAC. 137 (1997); Marc J. Ackerman & Tracy Brey Pritzl, *Child Custody Evaluation Practices: A 20-Year Follow-up*, 49 FAM. CT. REV. 618 (2011); William G. Keilin & Larry J. Bloom, *Child Custody Evaluation Practices: A Survey of Experienced Professionals*, 17 PROF. PSYCHOL.: RES. & PRAC. 338 (1986); Francella A. Quinnell & James N. Bow, *Psychological Tests Used in Child Custody Evaluations*, 18 BEHAV. SCI. L. 491 (2001).

²⁷ James N. Bow & Francella A. Quinnell, *Psychologists' Current Practices and Procedures in Child Custody Evaluation: Five Years After American Psychological Association Guidelines*, 32 PROF. PSYCHOL.: RES. & PRAC. 261(2001).

²⁸ Ackerman & Pritzl, *supra* note 26, at 621. But see the debate over these findings in Marc J. Ackerman & Tracy Brey Pritzl, *Reality Is . . . A Response to Martindale, Tippins, Ben-Porath, Austin, and Wittmann*, 50 FAM. CT. REV. 508 (2012); David A. Martindale et al., *Assessment Instrument Selection Should Be Guided by Validity Analysis not Professional Plebiscite: Response to a Flawed Survey*, 50 FAM. CT. REV. 502 (2012).

²⁹ AMERICAN PSYCHOLOGICAL ASSOCIATION, *supra* note 18; American Academy of Child and Adolescent Psychiatry Guidelines, *Practice Parameters for Child Custody Evaluation*, 36 J. AM. ACAD. CHILD & ADOLESCENT PSYCHIATRY 57S (1997); American Academy of Matrimonial Lawyers, *supra* note 25; Association of Family and Conciliation Courts, *supra* note 25.

to be aware of the psychometric qualities of the instruments they administer.³⁰ As one example, consider the American Academy of Child and Adolescent Psychiatry caution that:

In most cases, psychological testing of the parents is not required. Psychological tests, such as the Minnesota Multiphasic Personality Inventory, the Thematic Apperception Test, or the Rorschach were not designed for use in parenting evaluations. Their introduction into a legal process leads to professionals battling over the meaning of raw data and attorneys making the most of findings of “psychopathology,” but has little utility for assessing parenting.³¹

III. Testing as Hypothesis-Generation and Bias-Checking

The debate among professionals about the use of tests in CCEs can become almost as intense as the conflict between the parents we are tasked to assess. On one hand, some professionals take the position that testing is a necessary component of an adequate CCE: “failure to use these metrics in a comparative analysis poses jeopardy to the scientific merit of the evaluation and may prove a disservice to the several audiences who look to the [mental health evaluator] for credible advice founded in scientific methodology.”³²

On the other hand, there are those who recognize the limitations of the use of tests in CCEs and their associated potential to do harm: “Ultimately, psychological tests are not yet sophisticated or developed enough to be used with a population of child

³⁰ Jonathan W. Gould et al., *Responsible Use of Psychological Tests: Ethical and Professional Practice Concerns*, in *THE OXFORD HANDBOOK OF CHILD PSYCHOLOGICAL ASSESSMENT* 222 (Donald H. Saklofske et al., eds., 2013).

³¹ American Academy of Child and Adolescent Psychiatry Guidelines, *supra* note 29. The authors note that the domain and field of psychological testing finds its home primarily in the field of psychology, not psychiatry. Therefore, it may not be surprising that guidelines promulgated by the AACAP would explicitly indicate that psychological testing is not required. We include their guideline language, however, because it sets forth one of the rationales, supported by the authors, that describe some of the issues we are concerned with regarding the use of psychological testing. With this in mind, we further note that the guidelines for child custody evaluations of the American Psychological Association do not, directly or indirectly, imply that the use of psychological testing is necessary in conducting child custody evaluations.

³² Hagan & Hagan, *supra* note 13, at 86, 101.

custody litigants in a manner that generates dispositive data regarding the psychological characteristics of parents as they pertain to parenting the children in the family dynamic.”³³

The most common rationale takes a middle road, recommending the use of tests *not* as diagnostic indices of individual adult psychological functioning, but instead as an implicit “back-seat driver.” In this view, individual adult psychometric measures help the evaluator to keep his or her eyes on the road by generating, “hypotheses, which then must be subjected to verification from alternative data sources.”³⁴ Without tests, CCE evaluators are thought to be more vulnerable to their inevitable blind spots and biases: “It is important that psychologists use [individual adult psychometric] testing in child custody evaluations to generate and test hypotheses . . . [f]ailing to test/generate hypotheses creates potential for confirmatory biases and confirmatory distortion.”³⁵

We recognize the invisible and invasive effects of confirmational bias³⁶ and the evaluator’s associated need for checks and balances with regard to many types of cognitive and implicit bias. We take this as the rationale behind the mandate to incorporate “multiple and diverse methods”³⁷ and the value inherent in interviewing references, reviewing histories, and interviewing parties. These and other qualitative data sources serve the function of backseat driver. Used properly, they should open the evaluator’s eyes to alternative hypotheses, suggest new avenues of inquiry, and help to minimize evaluator bias and blind spots. They help assure that conclusions are reached only when there is data con-

³³ STAHL & SIMON, *supra* note 17, at 134.

³⁴ Mary Johanna McCurley et al., *Protecting Children from Incompetent Forensic Evaluations and Expert Testimony*, 19 J. AM. ACAD. MATRIM. LAW. 277, 300 (2005).

³⁵ James N. Bow et al., *Testing in Child Custody Evaluations – Selection, Usage and Daubert Admissibility: A Survey of Psychologists*, 6 J. FORENSIC PSYCHOL. PRAC. 17, 34 (2006).

³⁶ JONATHAN W. GOULD & DAVID A. MARTINDALE, *THE ART AND SCIENCE OF CHILD CUSTODY EVALUATIONS* (2007).

³⁷ Association of Family and Conciliation Courts, *supra* note 25, at guideline 5.4.

vergence, since unitary data points are far more likely to result in inaccurate and biased inferences and conclusions.³⁸

Despite the logical but misleading appeal of the objectivity of quantitative data, the fact that testing generates quantitative data does not make it more objective, more valuable, or a better check or balance. As appealing as quantitative data may be in the morass of emotion and conflict typical of most CCEs, the numbers are only as useful as they are reliable and valid. Whereas police records, medical records, and third party references can alert the examiner to alternative hypotheses, testing does so with the additional risk of misleading the naïve consumer and by generating diagnoses.

CCEs are explicitly not intended to generate individual adult psychiatric diagnoses.³⁹ This is in part because psychiatric diagnoses have no clear meaning for parenting or co-parenting capacity. Mary Johanna McCurley et al. opine, for example, “It is impossible to determine from test results alone if a parent’s measured response patterns are related, either directly or indirectly, to parenting competencies.”⁴⁰ Rachel Birnbaum et al. similarly observe that, “No psychological tests have predictive validity relating to parenting capacity, specific parenting time schedules, and/or sole custody compared with joint custody.”⁴¹ Finally, Virginia Luftman et al. report that “traditional psychological tests do not address parenting ability, the nature of the parent-child relationship, and the parent’s abilities to communicate or foster the child’s relationship with the other parent.”⁴²

³⁸ See generally Nancy Levit, *Listening to Tribal Legends: An Essay on Law and the Scientific Method*, 58 *FORDHAM L. REV.* 263, 270 (1989) (“A theory must be consistent with the generally accepted body of knowledge, both within its own discipline and in other areas Theories that rely on and relate to comprehensive and converging evidence from other disciplines are more likely to be valid.”).

³⁹ Association of Family and Conciliation Courts, *supra* note 25, at 4.6(c). The AFCC guidelines require that “[e]valuators recognize that the use of diagnostic labels can divert attention from the focus of the evaluation (namely, the functional abilities of the litigants whose disputes are before the court) and that such labels are often more prejudicial than probative.” *Id.* at 4.6(c).

⁴⁰ McCurley et al., *supra* note 34, at 299-300.

⁴¹ RACHEL BIRNBAUM ET AL., *CHILD CUSTODY ASSESSMENTS: A RESOURCE GUIDE FOR LEGAL AND MENTAL HEALTH PROFESSIONALS* 62 (2008).

⁴² Virginia H. Luftman, *Practice Guidelines in Child Custody Evaluations for Licensed Clinical Social Workers*, 33 *CLINICAL SOC. WORK J.* 327 (2005).

Moreover, when CCEs generate diagnoses via testing, custody litigation becomes, “a pathology hunt”⁴³; that is, a competition to identify the parent with more or worse labels who is thereby deemed unworthy of parenting. In short, diagnosis is the medical model’s individualistic, reductionistic, and static picture of an individual. CCE is a dynamic strength-based picture of a growing family system. With this in mind, it is startling to recognize that many custody evaluators “viewed the major purposes of testing as ruling out psychopathology and assessing personality functioning.”⁴⁴

IV. The UMDA, CCEs and Testing

The Uniform Marriage and Divorce Act (UMDA), drafted by the National Conference of Commissioners on Uniform State Laws, lays the foundation for the child custody laws in most jurisdictions.⁴⁵ The UMDA establishes that the court shall determine the best interests of the child on the basis of “all relevant factors, including . . . the mental and physical health of all individuals involved.”⁴⁶ Some take this clause as sufficient reason to include testing as part of a CCE at least in those jurisdictions that explicitly incorporate the UMDA in relevant legislation.⁴⁷ We contend that the UMDA contains no explicit or implicit need to incorporate individual adult psychometric tests. Beyond those principles of test development and statistical meaning discussed above, there is no standard or guideline that dictates how mental health professionals determine the quality of individual or family functioning. Clinicians representing every guild, background, and

⁴³ Vivienne Roseby, *Uses of Psychological Tests in a Child-Focused Approach to Child Custody Evaluations*, 29 *FAM. L.Q.* 97, 98 (1995).

⁴⁴ Bow et al., *supra* note 35, at 18.

⁴⁵ UNIF. MARRIAGE & DIVORCE ACT (amended 1973), 9A U.L.A. 282 (1998).

⁴⁶ While jurisdictions have their own language, definitions, and criteria for what constitutes the best interest of children, the language generally makes reference to elements of the mental and psychosocial health of all individuals involved. Thus, to a reasonable extent, a best interest determination must take account of mental, emotional, and psychosocial factors.

⁴⁷ Daniel J. Hyman, *Use of the Personality Assessment Inventory in Child Custody Evaluation*, 5 *OPEN ACCESS J. FORENSIC PSYCHOL.* 120 (2013).

type of training make these judgments every day without the use of tests. Why, then, would CCE be any different?

V. Admissibility

For an evaluation to have value in the courts, it must fulfill the criteria for admissibility of the jurisdiction in which the proceeding is taking place. Generally speaking, there are two evidentiary standards for expert scientific testimony⁴⁸: The *Frye* standard⁴⁹ or the *Daubert* standard.⁵⁰ *Frye* emphasizes a general acceptance standard. *Daubert* requires that the court establish the scientific validity of the testimony.⁵¹ The evaluator has a proactive duty to anticipate these hurdles: "In selecting methods and procedures, evaluators shall be aware of the criteria concerning admissibility and weight of evidence employed by courts in their jurisdictions."⁵²

Hurdling the lower bar set by *Frye* is easy, although meeting *only* a general acceptance standard leaves many of the relevant ethical and practice concerns discussed above unsatisfied and the evaluator vulnerable to associated criticisms and complaints. Getting over the higher bar set by *Daubert* is much harder. In the case of testing and CCEs, it may even be impossible.⁵³

⁴⁸ There are some variations to these standards in some states. For example, California uses the *Kelley/Frye* standard. Adopting the *Frye*, 293 F. 1013 (D.C. Cir. 1923), rule in *People v. Kelly* 549 P.2d 1240 (1976), the California Supreme Court noted that the *Frye* rule was generally a conservative one, requiring a consensus in the scientific community before testimony concerning a new and novel technique or methodology would be admissible. "The primary advantage, however, of the *Frye* test lies in its essentially conservative nature. For a variety of reasons, *Frye* was deliberately intended to interpose substantial obstacles to the unrestrained admission of evidence based upon new scientific principles." *Id.* at 1246.

⁴⁹ *Frye*, 293 F. at 1014.

⁵⁰ *Daubert.*, 509 U.S. at 588. The interested reader is directed to Daniel W. Shuman & Alex S. Berk, *Judicial Impact: The Best Interests of the Child and the Daubert and Frye Evidentiary Frameworks*, in PARENTING PLAN EVALUATIONS: APPLIED RESEARCH FOR THE FAMILY COURT 563 (Kathryn Kuehnle & Leslie Drozd eds. 2012).

⁵¹ See *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 150 (1999).

⁵² Association of Family and Conciliation Courts, *supra* note 25, at standard 5.6.

⁵³ Indeed, in states that do not use the *Daubert* standard, we encourage child custody evaluators to function and operate as if the *Daubert* standard is

More than a decade ago, James Flens cautioned that,

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.⁵⁴

Jonathan Gould more specifically warned that “most of the psychological tests and measures used in child custody evaluations might have difficulty surviving a *Daubert* challenge because the tests have not been developed for use in child custody assessments.”⁵⁵

VI. Summary and Recommendations

We believe that a logical, empirical, and statistically-informed analysis of the use of individual adult psychometric tests in child custody evaluations is sufficient to caution against the practice. These tests are not appropriately normed, have yet to demonstrate reliability, and cannot demonstrate criterion validity for inclusion in CCEs. They are not necessary to satisfy the criteria of practice standards, practice guidelines, or the UMDA. They are unlikely to withstand the intense scrutiny of a proper *Daubert* challenge. Above and beyond these technicalities, however, we posit that the variables assessed by these test instruments are not the variables that are fundamental in a child custody evaluation. Child custody evaluations are not about individuals. Child custody evaluations are first and foremost about family systems.

the jurisdictional norm. While not mandatory, this ensures a more cautious and conservative approach and helps to make more likely that a specific technique or approach is admissible in non-*Daubert* jurisdictions. We see this more conservative approach as being not only appropriate in forensic settings, we also see it as more respectful of the family given the abundance of caution that accompanies the *Daubert* standard.

⁵⁴ James R. Flens, *The Responsible Use of Psychological Testing in Child Custody Evaluations: Selection of Tests*, 2 J. CHILD CUSTODY: RES., ISSUES, & PRAC. 3, 13 (2005).

⁵⁵ Jonathan Gould, *Use of Psychological Tests in Child Custody Assessment*, 2 J. CHILD CUSTODY: RES., ISSUES & PRACTICES 49, 58 (2005).

Nevertheless, we recognize that the practice of including individual adult psychometric tests in CCEs is substantially ingrained within the community of child custody evaluators, particularly those who are trained as psychologists. Psychometric testing is central to the livelihood of many mental health professionals, and far too familiar to the courts to be abruptly discontinued. With these realities in mind, we recommend at least that evaluators and consumers approach test data with well-informed caution.

Timothy Tippins and Jeffrey Wittmann provide precisely this degree of well-informed caution with regard to the conduct of CCEs in general.⁵⁶ They advise that the data obtained in the course of conducting a child custody evaluation should be understood at four distinct levels. At level I, the evaluator reports direct observations (e.g., Mother told Billy to clean up). At level II, the evaluator ties direct observations to established scientific constructs (e.g., Mother is authoritarian). At level III, the evaluator ties these inferences to the question before the court (e.g., Billy will benefit from an authoritarian parent's care). At level IV, the evaluator leaps from inference to address the ultimate question, that is, the future allocation of parenting rights and responsibilities (e.g., Billy should be placed primarily in his mother's care).⁵⁷

⁵⁶ Timothy M. Tippins & Jeffrey P. Wittmann, *Empirical and Ethical Problems with Custody Recommendations: A Call for Clinical Humility and Judicial Vigilance*, 43 FAM. CT. REV. 193 (2005).

⁵⁷ We note several responses to Tippins and Wittmann's position, see Ackerman & Pritzl, *supra* note 26; Nicholas Bala, *Tippins and Wittmann Asked The Wrong Question: Evaluators May not Be "Experts," but They Can Express Best Interests Opinions*, 43 FAM. CT. REV. 554 (2005); Linda Dessau, *A Short Commentary on Timothy M. Tippins and Jeffrey P. Wittmann's "Empirical and Ethical Problems with Custody Recommendations: A Call for Clinical Humility and Judicial Vigilance,"* 43 FAM. CT. REV. 266 (2005); Jonathan W. Gould & David A. Martindale, *A Second Call for Clinical Humility and Judicial Vigilance: Comments on Tippins and Wittmann (2005)*, 43 FAM. CT. REV. 253 (2005); Joan B. Kelly & Janet R. Johnston, *Commentary on Tippins and Wittmann's "Empirical and Ethical Problems with Custody Recommendations: A Call for Clinical Humility and Judicial Vigilance,"* 43 FAM. CT. REV. 233 (2005), and Tippins and Wittmann's rebuttal. Timothy M. Tippins & Jeffrey P. Wittmann, *A Third Call: Restoring the Noble Empirical Principles of Two Professions*, 43 FAM. CT. REV. 270 (2005).

Borrowing from Tippins and Wittman, we recommend that testing data must be understood similarly. It is perfectly reasonable to report direct observations about how a parent behaved when taking a test and his or her specific responses (level I). It may even be reasonable to generate hypotheses about those behaviors as they may be relevant generic constructs such as parenting or co-parenting (level II). However, we strongly believe that leaping from these generic constructs to draw inferences about how the child's needs might best be served (level III) and how the ultimate question before the court should be resolved (level IV) is statistically, empirically, and ethically untenable.

Thus, one might reasonably report the content of a parent's responses to MMPI items (level I). One might even cautiously report the associated validity and clinical scaled scores (level II), although doing so without sliding down the slippery slope into diagnosis and inferences based on diagnosis about parenting (level III) and placement (level IV) likely requires the high wire balance of a Wallenda.⁵⁸ Doing so without tripping over critical issues associated with norming, reliability and validity, and thereby admissibility may simply be impossible.

For now, we recommend that the best practice is to omit individual psychometric tests from custody evaluations. We look forward to the development of forensically normed, reliable, and valid measures of systemic functioning. The only empirically sound systemic measures that we know of at present are built around attachment theory⁵⁹ and the associated standardized assessment methodology known as The Strange Situation.⁶⁰ Despite concerns about forensic application of these methods,⁶¹

⁵⁸ David A. Martindale, *Diagnoses in Child Custody Evaluation Reports*, 34 *MATRIM. STRATEGIST* 1 (2016).

⁵⁹ JOHN BOWLBY, *ATTACHMENT AND LOSS: ATTACHMENT* (Vol. 1 1969); JOHN BOWLBY, *ATTACHMENT AND LOSS: SEPARATION* (Vol. 2 1973).

⁶⁰ Mary Ainsworth & Barbara Wittig, *Attachment and Exploratory Behavior of One Year-Olds in a Strange Situation*, in *DETERMINANTS OF INFANT BEHAVIOR IV* 111 (Brian M. Foss ed.1969).

⁶¹ Benjamin D. Garber, *Attachment Methodology in Custody Evaluation: Four Hurdles Standing Between Developmental Theory and Forensic Application*, 6 *J. CHILD CUSTODY* 38 (2009).

there is promise.⁶² Unless and until such instruments are validated for use in CCEs, evaluators are cautioned that if the shoe doesn't fit, don't wear it.

⁶² Patricia M. Crittenden et al., *Assessing Attachment for Family Court Decision Making*, 15 J. FORENSIC PRAC. 237 (2013).