

**THE GOOD, THE BAD AND THE UGLY: A COMPLEX REVIEW OF
PSYCHOLOGICAL TESTING IN CUSTODY EVALUATIONS**

JAN MARIE DeLIPSEY, Ph.D, *Dallas*

Trial Resources, President

Licensed Psychologist

KEITH NELSON, *Dallas*

McCurley, Orsinger, McCurley & Nelson, L.L.P.

State Bar of Texas

29TH ANNUAL ADVANCED FAMILY LAW COURSE

August 18-21, 2003

San Antonio

CHAPTER 26

Jan Marie DeLipsey, Ph.D., P.C.
Trial Resources, President
Licensed Psychologist
4514 Travis, Suite 211
Dallas, Texas 75205
(214)599-0141, fax (214)599-0151

PUBLICATIONS (Partial Listing)

- De Lipsey, J.M. (2000). "Child allegations of Abuse" 2001 Family Law Update. (Aspen Publishers: IN PRESS).
- De Lipsey, J. M. & Larsen, P. (2000). The Child Witness Advanced Family Law (State Bar of Texas Family Law Section : San Antonio, Texas)
- De Lipsey, J. M. (1999) "Profiling Clients and Juries in Custody Cases" Advanced Family Law Drafting Course Proceedings (State Bar of Texas Family Law Section: Houston, Texas)
- De Lipsey, J. M. (1999) "Sentencing Considerations in Sexual Assault Cases" (Texas Criminal Defense Lawyers Association Proceedings: Fort Worth, Texas)
- De Lipsey, J. M. (1999). "Child Sexual Abuse" Expert Witness Manual (State Bar of Texas Family Law Section).
- De Lipsey, J. M., Orsinger, R. R. & Simpson, G. (1999). "How People Function Mentally and Emotionally" Expert Witness Manual (State Bar of Texas Family Law Section).
- De Lipsey, J. M. (1999). "Distinguishing Disorders, Syndromes, and Profiles" Expert Witness Manual (State Bar of Texas Family Law Section).
- Orsinger, R. R. & De Lipsey, J. M. (1999). "Diagnostic and Statistical Manual of Mental Disorders: A Lawyer's View" Expert Witness Manual (State Bar of Texas Family Law Section).
- Taylor, P. & De Lipsey, J. M. (1999). "Other Assessment Techniques" Expert Witness Manual (State Bar of Texas Family Law Section).
- Orsinger, R. R., De Lipsey, J. M. & Simpson, G. (1999). "The Scientific Bases for Psychological Theories" Expert Witness Manual (State Bar of Texas Family Law Section).
- Orsinger, R. R. & De Lipsey, J. M. (1999). "DSM IV - Eating Disorders" Expert Witness Manual (State Bar of Texas Family Law Section).
- De Lipsey, J. M. (1999). "DSM IV - Impulse Control Disorders" Expert Witness Manual (State Bar of Texas Family Law Section).
- De Lipsey, J. M. & Orsinger, R. R. (1999). "Psychological Syndromes" Expert Witness Manual (State Bar of Texas Family Law Section).
- De Lipsey, J. M., Orsinger, R. R. & Simpson, G. (1999). "Psychological Syndromes: Battered Woman's Syndrome " Expert Witness Manual (State Bar of Texas Family Law Section).
- De Lipsey, J. M., Orsinger, R. R. & Simpson, G. (1999). "Psychological Syndromes: Rape Trauma Syndrome & Child Sexual Abuse Accommodation Syndrome" Expert Witness Manual (State Bar of Texas Family Law Section).
- De Lipsey, J. M. & Crow, D. M. (1999). "Psychological Assessment Procedures: Basic Information" Expert Witness Manual (State Bar of Texas Family Law Section).
- De Lipsey, J. M. (1997). Competency of the Young Child Witness Reporting Sexual Abuse. Voice, vol. 26, No. 8, 22-28
- De Lipsey, J. M., Nickelson, G. L., & Morey, J. (1997). Handling a Custody Case Involving allegations of Sex Abuse. 23rd Annual Advanced Family Law Course, State Bar of Texas.
- De Lipsey, J. M. & Nickelson, G. L. (1996). Sexual Abuse Allegations and Domestic Relation Proceedings: Practice Guidelines for Attorneys. 19th Annual Marriage Dissolution Institute State Bar of Texas.
- De Lipsey, J. M. (1992) "Male Sexual Victimization: A review of current social science literature" Published proceedings of the 2nd Annual Midwestern Mental Health in Corrections Symposium.
- De Lipsey, J. M. & Shuman, D. W. (1991). "The Child Witness" In Costello, R. (Ed). Forensic Psychology for the Journeyman Clinician. San Antonio, Texas: The Texas Psychological Foundation, Inc.
- Morris, G. & De Lipsey, J. M. (1990). "Expert Testimony in Cases Involving Sexual Assaults on Children" Austin, Texas: Texas Criminal Defense Trial Lawyers Association.
- De Lipsey, J. M. & Shuman D. S. (1989) "The Child Witness" Austin, Texas: Texas Criminal Defense Trial Lawyers Association.
- De Lipsey, J. M. & James S. K. (1987). "Videotaping the Child Witness: The Texas Experience" In S. M. Sgroi, (Ed). Vulnerable Populations. Lexington, Mass.: D.C. Heath & Co.
- De Lipsey, J. M. & James, S. K. (1985). "Guidelines for Interviewing Suspected cases of Child Sexual Abuse." Dallas County Rape Crisis Center, Dallas, Texas.
- James, S. K., & De Lipsey, J. M. (1982). "Rape Defense, A Complex Issue" Institute of Criminal Justice Studies, 2.

MISCELLANEOUS

TV appearances:

Law Enforcement Television Network (1990)
Black Horizon (1984)
CBS Morning News (1983)
Perspectives (1982)
Nestro Dia (1982)
Dave Layman Show (1981)
Insights (1981)
Counterpoint (1982)
KXAS, Channel 5 News (1980 - 1985)
WFAA, Channel 8 News (1980 - 1985, 1991 - 1993, 1998)
KDFW, Channel 4 News (1980 - 1984, 1990 - 1992)

Interviews in Magazines and Newspapers:

Houston Chronicle (1996)
Texas Monthly (1990, 1991)
Newsweek - Consultant (1985, 1988 -1991)
Austin, American Statesman (1984)
United Press International (1984)
Houston Post (1984)
Every Woman Newspaper (1981, 1982)
Family Circle (1982)
Dallas Times Herald (1980 - 1995)
Dallas Morning News (1980 - Present)

KEITH M. NELSON
5950 Sherry Lane, Suite 800
Dallas, Texas 75225

EDUCATION: B.A. from Baylor University in December, 1979
J.D. from Baylor University School of Law in May, 1983

LICENSED: State of Texas, 1983; U.S. District Court, Northern District of Texas, 1983; U. S. Court of Appeals, Fifth Circuit, 1984

AREAS OF PRACTICE: Divorce, Child Custody Modification, Support; Board Certified by the Texas Board of Legal Specialization in Family Law(1988 - present)

PROFESSIONAL EMPLOYMENT

McCurley, Orsinger, McCurley, & Nelson, L.L.P., Partner
5950 Sherry Lane, Suite 800
Dallas, Texas 75225
(214) 273-2400
1992 to Present

Koons, Fuller, McCurley & Vanden Eykel
2311 Cedar Springs Road, Suite 300
Dallas, Texas 75201
(214) 871-2727
1988 to 1992

PROFESSIONAL ASSOCIATIONS:

- American Academy of Matrimonial Lawyers (1993 - Present)
- Member of Special Concerns of Children Committee, American Academy of Matrimonial Lawyers (1998 - 2001)
- Texas Academy of Family Law Specialists
- American Bar Association (Family Law Section)
- State Bar of Texas (Family Law Section)
- Dallas Bar Association (Family Law Section)
- Texas Supreme Court Advisory Committee on Child Support and Visitation Guidelines
- Board of Directors, Childrens Hope Chest (1999 - 2002)
- Texas Trial Lawyers Association
- National Association of Counsel for Children (NACC)

PUBLICATIONS AND SPEECHES:

- Co-author of "Strategies in Custody Cases," State Bar of Texas Marriage Dissolution Course, 1990.
- Co-Author and Speaker, "Pre- and Post-Nuptial Agreements", Marriage Dissolution Seminar, 1993.
- Co-Author and Speaker on "Creative Decision Making in Custody Arrangements; Psychological Theories; Their Admissibility After Daubert; And Alternatives to Trial"; American Academy of Matrimonial Lawyers 1998; Annual Meeting, Chicago, November 12, 1998.
- Co-Author and Speaker on "Promoting Infant Mental Health in the Family Court"; The Texas Association for Infant Mental Health, Dallas, Texas, January 7 and 8, 2000.
- Speaker on "Psychological Testing in Child Abuse Custody Cases: Risks and Precautions;" 16th Annual Symposium in Forensic Psychology, April 6-9, 2000, Newport Beach, California.

- Co-Author and Speaker on “Reliability and Validity of Psychological Testing for Custody Determination”; 26th Annual Advanced Family Law Course, San Antonio, Texas, August 21 - 24, 2000.
- Co-Author and Speaker on “Use and Abuse of Mental Health Professionals after Daubert,” 28th Annual Advanced Family Law Course, 2002.
- Speaker on “Attorney Perspectives on Child and Adolescent Forensic Psychiatry”; Presentation to Forensic & Child and Adolescent Psychiatry Residents, UT Southwestern Medical Center at Dallas, Dallas, Texas, January 29, 2002.
- Speaker on “Attorney Perspectives on Family Law - Custody and Divorce”; Forensic Psychiatry Presentation, UT Southwestern Medical Center at Dallas, Dallas, Texas, March 25, 2003.

TABLE OF CONTENTS

I. INTRODUCTION 1

II. ETHICAL GUIDELINES OF MENTAL HEALTH PROFESSIONALS 1

III. LEGAL THRESHOLD: WHAT ARE EXPERTS SUPPOSED TO BE DOING? 2

 A. Admissibility (Validity) & Relevancy..... 2

IV. SOCIAL SCIENCE TERMINOLOGY IN TESTING: BACKGROUND NECESSARY TO ATTACK OR PROTECT TESTING BY THE EXPERT. 3

 Correlation..... 3

 Correlation Coefficient..... 3

 Confidence Level (Intervals)..... 3

 Construct Validity..... 3

 Mean 3

 Normal Curve 3

 Norms..... 4

 Percentile Rankings 4

 Psychological Assessment..... 4

 Reliability 4

 Reliability Coefficient..... 4

 Standard Deviation 4

 Standard Error of Measurement..... 4

 Standardization 5

 Standardization Sample..... 5

 Structured Test..... 5

 T-Score..... 5

 Validity..... 5

 Validity Coefficient 5

V. CURRENT PRACTICES IN PSYCHOLOGICAL TESTING 5

 A. Minnesota Multiphasic Personality Inventory-2 5

 B. Rorschach Inkblot Test..... 7

 C. Millon Clinical Multixal Inventory-III 10

 D. Thematic Apperception Test..... 11

 E. Summing up the Big Four: MMPI, RORSCHACH, MCMI, TAT 12

IX. CONCLUSION 15

ADDENDUM A: 17

 Specialty Guidelines for Forensic Psychologists..... 17

 Guidelines For Child Custody Evaluations In Divorce Proceedings 18

 Ethical Principles of Psychologist and Code of Conduct..... 18

 Rights and Responsibilities of Test Takers: Guidelines and Expectations..... 20

THE GOOD, THE BAD AND THE UGLY: A Complex Review of Psychological Testing in Custody Evaluations¹

I. INTRODUCTION

This paper address facets of psychological testing in custody evaluations at an advanced level. Rather than review the scales of tests and what the test measures, which has been the trend with family law seminar papers addressing testing, this paper covers advanced issues related to testing that will arm the lawyer to attack or defend (if the testing is proper and properly conducted) psychological testing in a custody evaluation.

Readers unfamiliar with psychological testing will find the chapters on basic testing reviewed in The Expert Witness Manual Volume II a helpful aid in understanding concepts in this paper.ⁱ

This paper covers the ethical responsibilities and practice guidelines of the expert, the legal threshold for properly admitted testimony, social science terminology in testing, test construction and research, current practices in psychological testing, a critical review of the four most often given tests (MMPI-2, Rorschach, MCMI-III, TAT), and concludes with questions the attorney can utilize in examining the expert regarding testing and testing procedures.

When Dorothy undertook her journey to Oz, Glenda the good witch of the North told her to start at the beginning. Following Glenda's advice in writing about psychological testing in custody cases, one should begin with the ethical responsibilities and practice guidelines of the expert.

II. ETHICAL GUIDELINES OF MENTAL HEALTH PROFESSIONALS

First, the terms "psychology and psychological" are restricted by statute. Therefore, by law, only psychologists and psychiatrists are permitted to conduct psychological testing and call it such. These professionals are permitted to administer and interpret any test, for which they have the background and experience to administer.

There are five groups of professionals who are licensed to provide mental health care in Texas: 1) Licensed Professional Counselors, 2) Social Workers, 3) Marriage and Family Therapists 4) Psychologists 5) Psychiatrists. The first three groups typically have training in education. Only those trained in psychology can sit for the psychology-licensing exam. Psychiatrists have both medical and psychological (preferred but not required) training.

Licensed professional counselors, social workers, and marriage and family therapists are permitted to administer and interpret objective tests but are not permitted to administer and interpret projective tests. Projective tests elicit a free form narrative or written response. The ethical codes of these three professional groups require licensees engage in testing to have adequate training and experience to do so. Simply put, objective tests are those that require a structured response.

Because this paper targets "psychological testing", it is limited to examining the guidelines of psychologists only. There are five codes of conduct and ethics, which can directly apply to psychologists performing psychological testing in family law cases:ⁱⁱ

1. The Ethical Principles of Psychologists and Code of Conduct
2. Standards for Educational and Psychological Testing
3. Specialty Guidelines for Forensic Psychologists
4. Guidelines for Child Custody Evaluations in Divorce Proceedings.
5. Rights and Responsibilities of Test Takers: Guidelines and Expectations

These codes are guidelines – not rules. Nevertheless, they are intended to protect the welfare of the client. There may be rare circumstances when it is in a child's best interests to violate one or more provisions of these guidelines. If a psychologist consciously decides to violate a guideline provision, he or she should be prepared to explain and support their reasoning.

These five sets of ethical guidelines are not mutually exclusive. Rather, there is a marked consistency and overlap among these documents. A psychologist conducting a child custody evaluation should be guided by the ethical principles in all of these documents, not just one.

¹ Jan Marie De Lipsey is a licensed psychologist in Texas and the president of Trial Resources, a litigation consulting company, 4514 Travis, Suite 211, Dallas, Texas 75205 (214) 599-0141, jmdphd1@airmail.net.

So, when critically examining the custody evaluation of a psychologist, review these ethical principles and then determine if the psychologist followed them. All excerpts from three of the guidelines (Ethical Principles, Forensic and Custody Evaluation) that relate to psychological testing are quoted in Addendum A. The Rights and Responsibilities of Test Takers is summarize in Addendum A. The Standards for Educational and Psychological Testing are too lengthy to summarize in this paper, but a summary of the chapters and information necessary to order these standards are in Addendum A as well.

SUMMARIZE GENERAL TRENDS

III. LEGAL THRESHOLD: WHAT ARE EXPERTS SUPPOSED TO BE DOING?

Daubert, Robinson and the host of other Federal and State affirming cases have only somewhat impacted child custody evaluations in the United States. The impact in criminal law and torts; however, has been significant. The effect in Texas seems to be concentrated in major metropolitan areas: Dallas, Fort Worth, Houston, San Antonio and perhaps El Paso. These courts customarily encounter Daubert challenges, and judicial scrutiny of such cases has resulted in increasing instances of testimony being limited or barred altogether. Still, such events are the exception rather than the rule. Daubert challenges in cases other than child custody evaluations have been qualitatively different than those in the family law area. The areas of expert inquiry and testimony in other matters of law are usually narrower than the broad concept of “best interests of a child”. Areas of inquiry such as risk of dangerousness, recidivism rates of particular crimes, or risks associated with improper interviewing of young witness have substantial empirical research platforms. These types of social science questions are more easily given to operational definitions, and thereby given to more direct empirical investigation. This has not been the case with the question of “best interests of a child”.

The broad concept of “best interests of a child” defies operational definition. Herein lies the problem with direct application of research. How can the all the factors affecting child development be controlled and studied over a period of 10 to 20 years? Moreover, how would “success” or “failure” be defined? How would genetic predisposition and vulnerability be controlled and accounted for? Ultimately, “best interests” is a judgment call.

Texas public policy directs that both parents be as involved with a child (ren) as possible by the presumption of joint custody, “(1) The public policy of the state is to assure that the child will have frequent and continuing contact with parents who have shown the ability to act in the best interest of the child, provides a sage, stable, and non-violent environment for the child; and (2) encourage parents to share in the rights, powers, and duties of raising the child after the parents have separated or dissolved their marriage (b) A court may not render an order that conditions the right of a conservatorship to possession of or access to a child on the payment of child support.”ⁱⁱⁱ

On the other end of the spectrum, Texas law clearly does not tolerate child abuse. In between these divergent ends of the spectrum looms a colossal universe of other human factors that might negatively or positively impact a child’s development. Common sense should guide the public (jurors) and courts in many decisions about this universe of factors. So, where does the role of the expert and psychological testing come into play?

A. Admissibility (Validity) & Relevancy

The last decade of Texas case law has revamped the standards for admissibility of expert testimony. Common threads through these cases are concerns regarding the **validity** and **relevancy** of the expert’s testimony. A three-step model was developed for this paper in order to explain the complex relationship between testing and ultimate recommendations issued by the expert.

Step One: Is the testing valid? Does the test really measure what it purports to measure? (validity in this context would necessarily include reliability-consistency in measurement, see endnote for further explanation).^{iv}

Step Two: If a test finding is established as valid, how does it relate to the referral question, i.e. the question that necessitated the involvement of a custody evaluator? The relationship between the testing and the referral question is usually defined (or bridged) by scholarly research. Therefore, the validity test must again be applied to the scholarly research. Is it valid?

Step Three: If the research is valid and the testing is valid, then the third level of inquiry is whether the research directly relates to the referral question (relevancy).

The lawyer must move through each of these steps in critically evaluating the expert’s recommendations.

IV. SOCIAL SCIENCE TERMINOLOGY IN TESTING: BACKGROUND NECESSARY TO ATTACK OR PROTECT TESTING BY THE EXPERT.

Attorneys will need to know specialized language used in testing and research in order to effectively challenge or defend results from particular psychological tests. If the attorney possesses a good working knowledge of social science testing terms he or she will be well armed to ask appropriate questions of the expert. Do not allow an expert to get away with a general claim that a particular test or research to which test findings are applied is reliable or valid. Ask for coefficients, means, standard deviations, and probability values in research. The following list will arm to you make them “show you the money”.

These terms are listed in alphabetical order.

Correlation

A correlation represents the degree to which two or more variables are related to each other. A correlation refers to the direction that the variables move and does not represent cause and effect. (Example: height and weight are positively correlated. As one increases, the other tends to increase as well. Age and youthful appearance are negatively correlated, as one increases, the other tends to decrease).

Correlation Coefficient

A correlation coefficient is a statistic or number representing the degree to which two or more variables are correlated, i.e. related. Often abbreviated 'r'.

Confidence Level (Intervals)

The confidence level is expressed (most often) in terms of probability “*p*”. Probability expresses the level of confidence one can have in a particular score. A “*p*” value of .05 indicates that 5 times out of 100, the obtained score (or result) is in error – a finding that occurs by chance alone. Higher probability values indicate more risk for error. Lower probability values indicate less risk for error. In most cases, probability values higher than .05 are not acceptable in social science research.

Probability values are applied to research study results indicating the strength (or lack thereof) of the study findings.

Construct Validity

Construct validity is a form of validity. Construct validity answers the question of whether or not the measuring device actually measures the construct under question.

Mean

The mean, a measure of central tendency, indicates the average score on a test.

Each test will have a different mean for the test results, and most tests have a separate mean for specific categories of measurement. For example, the mean on the Wechsler Intelligence tests is 100 and the mean for the MMPI-2 is 50. (See Addendum B, MMPI-2 scores are expressed in terms of a “T”-Scores)

Normal Curve

The normal curve is an assumption in research of a graphical interpretation of a population that is ‘bell shaped’ with the highest frequency in the middle. The frequency diminishes in the tails of both sides of the curve. The normal curve is used to visually represent a normal distribution.

Test scores are often expressed in terms of a normal distribution – a normal curve of all scores. The normal curve assumes that about 68% of test scores fall between (+) 1.00 and (-) 1.00 *standard deviations* from the mean. Ninety-five percent (95%) of scores would fall between (+) 2.00 and (-) 2.00 and 99.8 % of the scores falls between (+) 3.00 and (-) 3.00 standard deviations from the mean. It is important for the lawyer to understand these percentages when inquiring about a client’s test score because it is meaningful to know whether the client’s test score is substantially similar or dissimilar to normal population scores or clinical population scores.

Different tests have different *means*. For example, a mean score on a Wechsler intelligence test would be 100 with 68% of the population scoring between 85 and 115 (the S.D. on the Wechsler is 15). It is helpful to translate scores into percentiles since this is a concept that most people understand. A score on a Wechsler IQ test of 130 would be two standard deviations above the mean – translated in to percentiles – this score would be at the ninety-eighth

percentile (98thtile) indicating that 98% of the population scored lower than 130. The mean on the MMPI-2 is 50, so the standard deviations would fall at different scores than on that of the WAIS-R.

Norms

A norm is an expectation based on multiple observations. Test designers define norms on populations when the test is constructed. When the test is initially given to a selected sample of persons, this group is the *standardization sample* and yields the test norms. Norms can also be the parameters encompassing the proper persons to whom the test is properly given and to whom the results can be properly applied and interpreted. For example, a test “normed” on people ages 18 to 70 cannot properly be given and interpreted to a child age 14. The person taking the test must be in the norm group in order to compare the results of that person’s testing with other individuals.

Norms are particularly important in understanding achievement scores or intelligence testing scores. With these types of tests, norms are usually expressed in terms or grade equivalents, age equivalents, or percentile rankings.

Percentile Rankings

Percentile Rankings can be computed for many tests. A score expressed in terms of a concept is easy to understand. The percentile ranking indicates the percentage of scores equal to or below a particular score. For example, a score at the 75th percentile indicates that seventy-five percent (75%) of the scores of others were at or below this score and that twenty-five percent (25%) of scores were at or above the score.

Psychological Assessment

Psychological Assessment refers to an evaluation of an individual’s functioning at a particular point in time. In most instances, psychological assessment refers to a broader process than just testing.

Reliability

As referenced earlier in this paper, reliability only refers to consistency in measurement or test scores. Higher reliability will be associated with more consistent test or measurement results. In psychological testing, if a test is not first proven to be reliable, then the question of validity is moot. Reliability is the foundation of validity.

Estimates of reliability come in many shapes and forms during test construction and subsequent research on a test. Different kinds of tests have different thresholds of acceptable reliability. For example, on intelligence tests a reliability coefficient should be at least .80 or above. The Rorschach reliability coefficients often fall in the .30 to .40 ranges – too low to be considered useful.

With every test there will be error. Therefore, it is important to know what part of a test score is related to the true score and what may be due to other factors, i.e. error.

Reliability is established through calculations and the resulting figure is called the *reliability coefficient*, “*r*”.

There are many ways to establish reliability: Test-Retest reliability, Equivalent Forms reliability, Internal Consistency reliability, Split-half reliability, Cronbach’s Alpha. For a more thorough review, see Chapter 3-35 in the Expert Witness Manual, “Psychological Assessment Procedures: Basic Information”.

Reliability Coefficient

The value ranges from “0” to “1.00”. 1.00 is perfect reliability. There is never a negative value for the reliability coefficient (but there is a negative value for a *validity coefficient*).

Standard Deviation

The standard deviation is a measure of spread within a distribution. It is the most popular and most reliable measure of *variability* but the more skewed a distribution (when the bell shape curve is distorted), the more error there will be in the *standard deviation* because of its reliance on the *mean*. The *shape of a distribution affects the mean*.

Standard Error of Measurement

A statistical procedure used to determine the amount of error of any measurement device.

Standardization

Standardization refers to two aspects of testing and test construction: Administration & Scoring procedures and *Norms*.

Standardization Sample

This is the selected sample of persons to whom a test is given when it is being developed. It is also called a “reference group”.

Structured Test

Also called an “*objective test*”. An objective test is differentiated from an unstructured test (*projective test*) primarily by the scoring procedures. A structured test has clearly defined scoring procedures.

T-Score

A T score is a standard score that sets the *mean* to fifty and *standard deviation* to ten. MMPI-2 scores are converted to “T-scores”. The commonly used cutoff score signaling pathology on the MMPI-2 is sixty-five (65). This score is 1.5 standard deviations above the mean indicating that approximately 95% of the normal population scores fall below it (on the MMPI-I, the cutoff was 2 S.D. above the mean, a T-score of 70 which would be around the 98th percentile).

Validity

This is the most important concept to understand in test construction and testing. Validity refers to how well a test measures what it purports to measure. There are many shapes and forms of validity but *criterion related validity* and *predictive validity* are probably more relevant to the concerns of attorneys. An example of *criterion related validity* would be the comparison of achievement scores to a student’s actual grades. One would expect that a valid test would show higher achievement scores correlated with higher grades and vice versa. *Predictive validity* would refer to the test’s ability to forecast future performance or manifestation of a problem. An example would be a test’s ability to forecast future dangerousness.

For additional information on validity, see Chapter 3-35 in the Expert Witness Manual, “psychological Assessment Procedures: Basic Information”.

Validity Coefficient

Validity is expressed in terms of a *correlation coefficient* indicating the extent to which things fit or are associated together. Remember that correlation coefficients range from -1.00 to $+1.00$ and that the relationship can be either negative or positive. The relationship of two variables (things) does not indicate causation.

V. CURRENT PRACTICES IN PSYCHOLOGICAL TESTING

There have only been a few mainstream studies surveying testing procedures of mental health professionals published in the past decade. A 2001 study found that evaluators ranked psychological testing as the fourth (4th) most important procedure used in child custody evaluations.^v Testing was preceded in ranking by parent-child observations, (3rd) clinical interview with the child (2nd), and clinical interview and history with the parents (1st).^{vi} Approximately three hours were devoted to psychological testing of adults with about two hours spent with testing of children.

A 1997 study found that ninety percent (90%) of the experts who conduct child custody evaluations employ testing instruments. But in keeping with the 2001 study, they also used clinical interviews.^{vii} The 1997 study found that the most commonly employed tests are the Minnesota Multiphasic Personality Inventory, the Rorschach, the Wechsler Adult Intelligence Scales-Revised, the Millon Clinical Multiaxial Inventory, and the Thematic Apperception Test.^{viii} Four of these tests will be briefly reviewed in this section. The WAIS-R, an intelligence test, is not reviewed. Note that no test that has been specifically designed for custody evaluation was in the top five.

A. Minnesota Multiphasic Personality Inventory-2

Test Description: The MMPI –2 is an objective test of 567 T-F questions.

Frequency of Use: The MMPI/MMPI-2 is thought to be the most commonly administered objective test (91% of the respondents in the 1997 study reported giving this test but did not discriminate which version they were using).

Research: The MMPI/MMPI-2 has generated more research, over 15,000 books and articles, than any other testing instrument.

Use in Training/Education for Psychologists: Use of the MMPI-2 is currently taught in virtually every educational program for psychologists. It is probably the most common topic of continuing education programs in testing for psychologists.

Standardized Testing Requirements:

Age: The MMPI-2 is designed for use with adults aged 18 or older. (The MMPI-A is recommended for individuals between the ages of 14 and 18)

Reading Abilities: Some researchers hold that the MMPI-2 is geared toward a fifth-grade reading level to understand content. However, recent research indicates that twenty-five percent (25%) of the items on the MMPI-2 require a reading level beyond the 8th grade. It is not appropriate to use the written version with persons with poor reading skills.^{ix} *Normative Sample:* The normative group for MMPI-2 was 2,600 people (1,462 women and 1,138 men) randomly drawn from seven regions of the across the United States). The sample included Blacks, Asian Americans, Hispanics, American Indians, and Anglos. The sample is said to be generally reflective of the U.S. population in terms of age, gender, and minority representation.

Scoring:

The MMPI-2 can be hand scored or computer scored. Because scoring is laborious and complicated, errors are common in hand scoring. Computer generated scores and interpretative profiles should be from mainstream sources such as Psychosocial Assessment Resources or National Computer Systems. These services regularly update their database with current valid research.

Scales

MMPI-2 scores are clustered into three groups: *Validity scales, Clinical scales and Content Scales.*

Validity scales are comprised of nine (9) measures: Scale (Cannot Say), L Scale (Lie Scale), F Scale (Confusion Scale), F (B) (Back Page F), VRIN (Variable Response Inconsistency), TRIN (True Response Inconsistency), Mp (Malingering), Sd (Social Desirability) and the K Scale (Correction Scale).

The *validity scales* are interpreted in context of one another, rather than independently. A validity profile will indicate the client's general response set, i.e. normal, defensive, faking good, faking bad, extreme confusion, or lack of cooperation. The validity profile will indicate whether or not the resulting clinical profile is valid and if so, how to interpret the clinical scales – i.e. do they properly estimate, over estimate or under estimate the client's true functioning. Invalid profiles and the reasons for invalid profiles can, and in some instances, are as meaningful as a valid profile.

Clinical Scales: There are ten (10) main clinical scales, which are viewed as hypotheses regarding the individual's functioning. These hypotheses are either confirmed or ruled out by other data (other testing, collateral records, history). The clinical scales are hypochondriasis (1), depression (2), hysteria (3), psychopathic deviate (4), masculinity- femininity (5), paranoia (6), psychasthenia (7), schizophrenia (8), hypomania (9), and social introversion (0).

Content Scales: The content scales on the MMPI-2 assess four general areas: Internal Symptomatic Behaviors, External Aggressive Tendencies, Negative Self-view and General Problem Areas.

Administration: The MMPI-2 can be administered via booklet with paper and pencil or can be administered with the client taking the test on the computer. The latter method of administration has the advantage of fewer scoring errors. There are short (370 items) and long versions, but the long version is the only acceptable one for custody evaluation, unless it is being used as a screening instrument. As in all forensic testing, the client should be monitored during the test. The test booklet instructs the client to answer the statements as it applies to their own opinion of him or herself. The client is encouraged to answer every statement.

Interpretation: A valid MMPI-2 is interpreted by high point scales and combinations of high point scales (over a T score of 65). High scales are indicative of psychopathology. Valid scores on scales under a T-score of 65 are

thought to be in a “normal” or subclinical range. Subclinical scores are typically should not be interpreted. Content scales are used to provide additional data in light of the clinical profile.

Code types and high point codes are then interpreted in light of other research studies in the domain of the trait or problem measured. The reliability and validity of these studies and their correlation to code types varies from study to study. And, the reliability and validity of the scales themselves will vary.

What does the MMPI-2 Test? The MMPI-2 tests the presence of psychopathology and/or symptomatic behavior that correlates to nine of ten clinical scales. The masculinity-femininity scale (5), which assesses traditional gender interests, is not considered to be a measure of psychopathology. If the expert attempts to attach some pathological hypotheses with the scale, it should be vigorously challenged. In simplistic terms, the test performance of populations with known problems is compared to the client’s profile. The logical assumption is that if the client’s score is in a similar range to a clinical sample (T-score over 65), then the client is suspected to have similar difficulties as the know population. The key word here is *suspected*. Elevated scales on the MMPI-2 must be interpreted in light of other data. Clinical scale elevations are supposed to be viewed as hypotheses regarding a client’s functioning. For example, a very disturbed individual with reality testing problems will score similarly to a person suffering from post trauma stress. A person who molests children and commits no other crimes will score similarly to a normal symptom free individual. The MMPI-2 does not directly measure parenting abilities but some of the data could be used to assess parenting capacity. For example, a parent who is severely depressed would have limited parenting capacity.

Reliability: Each scale as its own reliability data. Remember that reliability has to do with consistency of measurement, i.e. if the test was taken next week, would the resulting scores be the same? Research indicates that test-retest reliability on clinical samples ranges from .92 (social introversion scale) to .67 (paranoia scale). Depending on the high point or two-point code, the scale on which a client is measured may or may not be reliable. It is important to pinpoint which scales are elevated (if any) and if so, learn the reliability data for that or those particular scales.

Asks the expert for this very particular information. If the reliability is poor, then the scale findings are without merit.

Validity: Apart from the validity index of the profile (a different animal than the validity of the scales), each one of the scales has its on validity data and research. As with reliability, it is important to pinpoint exactly which scales are elevated and then to learn the validity data with that particular scale. Nailing down the expert with this kind of information will take some effort on the part of the attorney. The attorney will need to ask specific questions and sometimes be willing to wait (if in deposition) for exact answers. For example, scale 6 (paranoia) is the least reliable scale and therefore has less validity support than other scales. On top of that, in the context of a custody evaluation, the elevation of Scale 6 might be expected, depending on the circumstances.

B. Rorschach Inkblot Test

Test Description: The Rorschach is a projective test that consists of ten (10) cards, some in black and white cards and others in color. The client is shown the cards in numerical order one at a time.

Frequency of Use: According to the Ackerman 1997 study, the Rorschach was used by forty-seven percent (47%) of the examiners conducting custody evaluations. The Rorschach is probably the most well known projective test.

Research: There are thousands of research studies regarding the Rorschach, many of the modern ones relate to prediction of psychopathology, dangerousness, reality testing, and schizophrenia. Appropriately applying Rorschach research findings is nearly impossible because there are enormous problems with consistent administration and scoring. If administration and scoring procedures used by the examiner are exactly like the procedures used in a research study, then findings of the study might be appropriately applied. Unfortunately, there are dozens of administration and scoring systems of the Rorschach. Exner, Klopfer, and Beck systems are the most popular. The only studies that approach usefulness in custody evaluations are those involving the use of the Exner scoring system because the gap between the test given by the examiner and the tests used in a study is narrower.

The Exner system was first published in the late 1950’s. The Exner system was designed to develop structured administration, scoring, and interpretation with a hope to develop reliability and validity research. Today, younger

psychologists use the Exner method. Despite some fifty years of research with the Exner system, there are still serious concerns regarding the soundness of these studies. Most of the research by the Exner group has published in-house. Much of this work has not undergone the scrutiny of outside, referred journal review.^x

Some of the leading researchers in testing and psychology refer to the Rorschach as a “method” rather than a “test” meaning that it generates data that can be interpreted from many different theoretical views. Such an approach would clearly rely on subjective interpretation.

Use in Training/Education for Psychologists Some post-graduate programs teach the Rorschach and some do not. Of those that do, most teach the Exner scoring system. There are several continuing education programs conducted by the Exner group which teaching they’re scoring and interpretation system.

Standardized Testing Requirements: Because there are many scoring and testing systems, comments in this section are limited to the Exner system.

Age: The Rorschach can be administered to persons from age five (5) and up.

Reading Level: No reading is required. The subject must be able to communicate either verbally or by written means.

Normative Sample: There have been only limited numbers of studies with Blacks and virtually no studies of substance with other minorities save for prison studies and prediction of dangerousness. The few studies with Black populations indicate that fewer responses are given than with Anglos which could impact the validity of the resulting protocol. None of the studies related to custody evaluations.

Scoring: Some examiners do not even score the protocol, some use their own scoring system, and some use more structured scoring systems. As in the administration section, scoring comments are limited to the Exner system. Other scoring systems have no reliable or valid research base to support their admissibility in a forensic case.

Exner’s system is comprehensive and taps four areas of data interpretation: Quantitative data (data that can be compared to data in other Rorschach protocols of particular populations such as depressed persons or schizophrenic persons), internal data (data within the protocol itself such as response style), sequencing of scores (attitude toward testing or changes in responses as the test progress), and qualitative data (focusing on the exact words of the client).

There are nine broad categories for scoring and within each category there are numerous subcategories. The possible combinations of scores for each response are huge and will vary from examiner to examiner depending on the training and experience of the examiner and the expertise with which the test was administered. Scoring a Rorschach using the Exner system is extremely complicated and time consuming. It is also technically difficult and requires extensive training and education.

One of the main problems in scoring is reliability. Two examiners will likely score a Rorschach protocol differently. Even the same examiner will likely score it differently if weeks pass between scoring it. Therefore, inter-rater reliability is a significant issue in scoring. Taking a cold protocol to an Exner expert will invariably result in different scoring than that of most custody examiners because of the infrequency with which custody evaluators use the system and because of their lack of experience compared to a specialist in Exner scoring.

Scales

There are no scales on the Rorschach. The Exner scoring and interpretation system yields numerous indexes and score ratios claiming to tap into varying areas of functioning ranging from a measure of ego-centricity to schizophrenia, depression and so forth. The main index is known as the four squares. This index is the foundation for interpretation and incorporates the basic scores and ratios thought to be characteristic of one's problem-solving style. The four indices of the 4-square are 1) Erlebnistypus (EB, the ratio of human movement to weighted color responses); 2) Experience Actual (EA, the sum of human movement and weighted color responses); 3) Experience Base (eb, the ratio of nonhuman movement to shading and gray-black responses); and 4) Experience Potential (ep the sum of non-human movement, shading, and gray-black responses).

Administration: The Exner administration method is detailed and lengthy. There are two stages of administration, “free association” and “inquiry”. The subject is allowed to sit in any position other than face-to-face with the examiner. The preferred seating is side-by-side. The seating of examiner and client is essential in avoiding visual cues from the examiner and distracting actions by the examiner, such as the verbatim response writing. A verbatim

statement regarding introduction of the test and presentation of the inkblots is provided by the Exner system. On the presentation of the inkblot the examiner is permitted to only say, “What might this be?” Exner stresses that this phrase not be altered at all. If the phrase is altered, then the examiner has improperly administered the test via the Exner procedures. There are numerous do’s and don’ts in the Exner administration procedures such as the proper response to a turned card or question from the client. Unless the examiner is completely familiar with this system, there will be errors, some of which can invalidate the entire test.

Throughout the test, the examiner is supposed to record the client’s verbatim statements, the client’s behavior, response time to each card, card holding time, and other information- a nearly impossible task. All cards are presented in ascending order. This stage of administration is called the “free association “ stage. Protocols with less than 14 answers are considered invalid and Exner allows for this stage to be repeated once in order to obtain more responses.

The next administration phase is the “Inquiry”. The Inquiry phase is used to confirm the examiner’s understanding of how the client drew perceptions from the cards at the time the verbal response was given. No new verbal responses are allowed. In this phase, the examiner reads the previous response verbatim, out loud and then prompts, “What made it look like ”. Giving only to this prompt is essential in avoiding contamination of the subject’s inquiry. Again, if this prompt is essentially changed, it is not appropriate to apply Exner scoring and interpretation to the protocol.

Interpretation:

As referenced earlier, the Exner four square is the basis for interpretation. From the four square other ratios and indices are reviewed for comment by the examiner. The interpretive conclusions of the Exner system are believed to be supported by research studies. As noted early, a common criticism of the Exner system is its heavy reliance on in-house research that has not been subjected to mainstream journal scrutiny. But making the leap (one too large to qualify as admissible evidence) and assuming the research, as a whole is sound, the research itself is skewed toward maladjustment and psychopathology. In fairness to the Exner group, within the last decade they have provided data of “normal” adults and children to broaden analysis to a more normal population. Nevertheless, the number of subjects has been low (700 adults and about 1400 children and adolescents).^{xi}

What does the Rorschach test?

It is difficult to say what the Rorschach tests (if anything) because of poor reliability and validity support. Looking only at the Exner system and the collection of research, with this system, Exner indicates that the Rorschach should not be used to generate a specific diagnosis (referring to the Diagnostic and Statistical Manual for Disorders).

Exner allows it could be used in concert with other data but certainly should never be the keystone of a diagnosis.

Given that these areas of functioning can be assessed through collateral data and other forms of testing, and that the Rorschach seems to rely heavily on subjective interpretation and clinical opinion, the down side of using it outweighs any potential gains. In the rare circumstances, its use might be justified with Anglo clients who cannot read well enough to take other tests or if there are some questions regarding schizophrenia or some sort of thought disorder and collateral data is limited. Such circumstances would be rarely encountered in a custody evaluation

Reliability

Overall, reliability studies show poor support for the Rorschach. There are even more significant reliability problems with testing children. Tests only six (6) months old with children under fourteen (14) are considered invalid by the Exner group. This cautionary warning makes sense given that as a young child develops there are substantial changes in perception, verbal expression, cognitive abilities and so forth. If an expert makes some claim that the Rorschach is reliable, then the attorney has a right to expect that the expert provide full support for the claim in terms of research studies and so forth because the vast amount of research available does not support the claim. Unfortunately, as several Daubert hearings have (wrongly) shifted the burden to the opponent of the admissibility of the evidence, successful defeat of such testimony might require employment of yet another expert to “prove” the lack of support to the court.

Validity

Given that the reliability of the Rorschach, even with the Exner scoring system, is overall poor, there is little reason to even address validity. Nevertheless, in a 2000 review of the scientific reliability and validity of projective tests,

researchers commented that the Rorschach test had aroused more controversy than any other projective tests.^{xii} It is doubtful that the Rorschach Inkblot test would pass muster on the question of general acceptance by the relevant scientific community. Plus, an expert attempting to put forth Rorschach findings as valid will encounter the same problems as with reliability. The majority of the research shows significant problems with validity. Because the lack of validity support is generally accepted by the scientific community, this places a heavier burden on the expert taking a contrary posture to supply direct support of his or her opinion by way of research articles and studies.

C. Millon Clinical Multixal Inventory-III

Test Description: The MCMI-III is an objective test of 175 T-F questions.

Frequency of Use: The MCMI-III is believed to be the second most commonly administered objective test in child custody evaluations. It should be noted that several researchers has asserted that the MCMI-III should not be used routinely in child custody evaluations as it will reflect inflated disturbance in normal individuals.^{xiii} If the individual tested has evidenced signs of psychopathology, then the test would be appropriate. Even the test manual warns against using the MCMI-III as a general measure of personality for normal populations.

Research: The MCMI-III has been the subject of hundreds of research studies. The research overall has had very mixed results. Current assertions are that the MCMI-III should be used cautiously in forensic settings because of the lack of research supporting its validity.^{xiv} Ackerman and Ackerman in their book, Psychological Experts in Divorce Proceedings recommend against using the MCMI-III in custody evaluations.

Use in Training/Education for Psychologists: Use of the MCMI-III is taught in most educational program for psychologists. The MCMI-III, though, was not published until 1994 so most psychologists have probably been trained with earlier versions of the MCMI which, unlike the MMPI and MMPI-2, have undergone substantial changes (90 of the 175 items were changed from the 2nd to the 3rd version). Therefore, it is important to ensure the examiner has had training in interpreting the MCMI-III.

Standardized Testing Requirements:

Age: Subjects should be eighteen (18) years and older.

Reading Abilities: Eighth grade minimum reading level is required.

Normative Sample: The normative data is based entirely on samples that evidence emotional problems, interpersonal symptoms, who are undergoing therapy or who are undergoing an evaluation. This type of sample is called “clinical” meaning the absence of symptom-free or normals.

Scoring:

The raw scores are translated into standardized base rate scores that reflect the percentages of patients who evidence the disorder across different settings. This score is expressed in terms of “BR” (base rate). The BR was designed to anchor varying cutoff points with a view to maximize correct classification in a given diagnostic category (that fits with the DSM). Scores at different threshold represent hypothetical presence or prevalence of particular disorders.

Scales

MCMI-III scales are clustered into five groups: *Modifying Indices, Clinical Personality Pattern scales, Severe Personality Patterns scales, Severe Clinical Syndrome scales, Clinical Syndrome Scales.*

Modifying scales are comprised of four (4) measures: Validity, Disclosure, Desirability, and Debasement. These scales (similar to the validity index profile of the MMPI-2) reflect the test-taking approach of the client, i.e. oppositional, random responding, openness, defensiveness, faking good, faking bad.

Clinical Personality Patterns are the basic personality scales that tap stable and chronic characteristics. They are Schizoid, Avoidant, Depressive, Dependent, Histrionic, Narcissistic, Antisocial, Aggressive, Compulsive, Passive-Aggressive, and Self-Defeating.

Severe Personality Patterns tap into severe problems (as compared to the clinical personality patterns) in emotional control, rationality and/or relationships. These three (3) scales are Schizotypal, Borderline, and Paranoid.

Clinical Syndrome Scales measure transient symptoms similar to DMS Axis I disorders: Anxiety Disorder, Somatoform Disorder, Bipolar: Manic Disorder, Dysthymic Disorder, Alcohol Dependence, Drug Dependence, Post-Traumatic Stress.

Severe Clinical Syndrome Scales measure pathological symptoms associated with marked impairment: Thought Disorder, Major Depression, Delusional Disorder.

Administration.

Similar to the MMPI-2, the MCMI-III is a paper and pencil test filled out by the client. The testing should be monitored to ensure that the client fills out the answers without assistance or interference from others.

Interpretation:

Scales above a BR score of 75 are considered meaningful. If several scales are elevated, then the highest two or three personality scales are given more weight in the interpretation. However, the greater the number of scales above 75 usually indicates greater pathology. There is a limited pocket of research which indicates that a few of the scales are associated with positive traits such as coping mechanisms or self confidence rather than just pathology (histrionic, narcissistic, compulsive). The significant elevation of either the clinical syndromes or severe clinical syndrome scales would support a hypothesis of marked problems such as a personality disorder. Finally, BR elevations on any of the Severe Personality Pathology scales reflect significant pathology

Reliability: Most of the scales of the MCMI-III have an internal consistency coefficient above .80 which is acceptable (all of the scales on the MCMI-II had an internal consistency of .80 or above). The few which do not should be interpreted with caution: Narcissistic, Antisocial, Sadistic (aggressive), Compulsive, Bipolar: Manic and Delusional Disorder.

Validity:

Although the MCMI-II validation studies were impressive, the MCMI-III validation studies have been a disappointment. The original validation study published in the MCMI-III manual received a great deal of well-founded criticism. ^{xv}

D. Thematic Apperception Test

Test Description: The TAT (3rd revision) is a projective test consisting of thirty one (31) cards of black and white drawings of people in the context of everyday life. One of the cards is blank. The drawings are supposed to be neutral in nature; however, some of the drawings clearly have a negative or disturbing connotation.

Frequency of Use: The TAT is the 2nd most widely used projective test in custody evaluations.

Research: The TAT has been the subject of hundreds of research studies.

Use in Training/Education for Psychologists: Use of the TAT is often included in a general testing course for postgraduates. It is rare that more than 3 hours of instruction would be devoted to the TAT. Continuing education courses in the use of the TAT are also few and far between.

Standardized Testing Requirements:

Age: Subjects should be at least four (4) years of age. There is no upper age limit.

Reading Abilities: No reading is required.

Ethnicity: There are few studies using the TAT with minority groups. Therefore, it should be limited in use to Anglos.

Administration: The manual instructs the examiner to administer the cards in two separate sessions on two separate days. Ten cards are administered in each session. Cards are chosen by the gender and age of the client. Not all cards are given.

The examiner is to sit behind the client and record the verbatim statements of the client. ^{xvi} The test manual provides instructions the examiner should give to the client. The instructions basically instruct the client to “make up as dramatic a story as you can for each...” The client is then presented with the card and tells a story for each card.

After the first story the examiner should ask questions to bring out key elements that might be missing. The elements are listed in the manual. After this comment, the examiner should refrain from additional comments other than noting the time and noting whether the client is ahead or behind the average time of five minutes in their story telling.

The average story is between 100 and 300 words long. It is doubtful that any examiner could record the stories verbatim. Some examiners therefore, tape record the stories.

Scoring: Like the Rorschach there are several scoring systems. Unlike the Rorschach there is no general system of scoring that is widely used. The test constructor, Murray, indicated that after the examiner has basic facts about the client, that the examiner use client background information to look for significant elements. To Murray, such elements were related to philosophical mythology of the hero, forces of nature, triumphs, defeat and so forth.

Scales There are no scales on the TAT

Interpretation: As with the scoring problems, there is no general system of interpretation of the TAT that is widely used. After the examiner has looked for significant elements, the examiner then looks for themes in the stories. From the themes in the stories, the examiner draws conclusions about the client’s personality.

What does the TAT measure? There is no meaningful reliability of validity data to support the TAT. As would be expected, the test administration procedures run the gamut and the interpretation is highly vulnerable to clinical impression. Under such circumstances, results from this kind of test usually say more about the examiner than what is examine. Most researchers in the forensic realm recommend against using the TAT in custody evaluations^{xvii}

E. Summing up the Big Four: MMPI, RORSCHACH, MCMI, TAT

Critically examining the research support for the most commonly used tests, the Rorschach and the TAT fall miserably short of the threshold for evidentiary admissibility. The use of these tests has been extremely controversial in the social science scientific community. These projective test properties fail the legal standards for admissibility. Use of the tests in a custody evaluation likely violates sections of each of the five canons of ethics and practice of psychologists. Nevertheless, they are frequently used and infrequently challenged.

The research support for the MMPI-2 is sufficient to warrant its use in custody evaluations under most conditions. However, the MMPI-2 findings must be extrapolated and linked to research relevant to the referral questions as described in the 3-Step process earlier in this paper. For example, what good does it do to know that an individual is depressed unless the depression is linked to research about the effects of depression on parenting or child development?

The research support for the MCMI-III is shaky. If it is used, it should be under conditions where there is adequate support that psychopathology exists to the degree to warrant using the test. As with the MMPI-2, findings must be extrapolated and linked to research relevant to the referral question.

Attorneys’ Checklist: Questions about Testing and Testing Procedures

A list of potential question areas is provided to guide the attorney through a general inquiry regarding whether testing was appropriately and properly used in a custody evaluation. This list incorporates the ethical principles, social science terms, and correct procedures for testing and will get the attorney off to a good start in ascertaining whether the testing is reliable, valid, or relevant.

It is unfair to expect any psychologist to be able to answer the following questions without the proper books and testing manuals. Attorneys are advised to communicate the intention of asking reliability and validity questions prior to deposition or trial. Ambush in these areas of inquiry will not yield the needed results. Notice of a Daubert challenge is sufficient to warn the expert to be armed with this material.

If the psychologist cannot answer these questions with their resources and materials available, then the attorney has enough ammunition to seriously challenge the admissibility of the testing and probably the admissibility of the entire evaluation and opinions. Whether or not these types of questions are posed at the expert’s deposition, a Daubert hearing or trial is a strategy decision. Nevertheless, if the questions are posed at deposition or a Daubert hearing, it

will give the opposing lawyer time to investigate the expert's claims or review them with their own expert and prepare evidence or testimony to attack the claims if they are unfounded.

Potential Question Areas:

Expert's Background & testing:

- Exactly what courses prepared you to administer the _____ test (ask for each test)
- Exactly what continuing education prepared you to administer the _____ tests (ask for each test)
- What supervision, if any, have you had in administering the _____ test?
 - Explain the supervision and testing process

Pre-testing tasks:

- What are the referral questions, i.e. the psychological and legal questions?
- What, if anything, did you tell the client about the purpose of the evaluation?
[ethical guideline]
- What, if anything, did you tell the client about the nature of the procedures you employed?
[ethical guideline]
- What did you tell the client about the instruments and techniques employed in the testing?
[ethical guideline]
 - What did you tell the client about the kinds of tests that would be used?
 - What did you tell the client about the planned use of the results?
- Did you inform the client of the limits of confidentiality?
[ethical guideline]
- Did you allow an opportunity for the client to ask questions about the testing?
 - If so, how so?
 - Where is this event documented?
[ethical guideline]
- Did you obtain her or his consent to undergo testing (the evaluation)?
[ethical guideline]

Identifying Tests

- What tests were administered?
[Was each test given commonly known? ,Were obscure tests given?, If so, why?]
- What was the order of the tests given?

Psychometric Properties of Tests

- What are the psychometric properties of each test given?
 - What is the norm group of the _____ test?
 - Does the client fit the norm group?
 - What are the reliability coefficients of the _____ test?
[Were reliability coefficients at least .80?]
 - Exactly where is the reliability research published, i.e. names, dates, journals, books? (Some of this will be published in the test manual)
 - What are the validity coefficients for the _____ test?
[Were the validity coefficients within an acceptable range?]
 - Exactly where is the validity research published, i.e. names, dates, journals, books? (Some of this will be published in the test manual)
 - Is the test reviewed in Mental Measurements Yearbook?
 - What are the opinions by test reviewers in the MMY?

[ethical guideline prohibits use of obsolete or outdated tests or test results]

Test Administration

- Did you or someone else administer all of the tests?
 - If someone else administered a test, what did you do to ensure the data were gathered in the proper manner?
[ethical guideline]
- Who, if anyone, monitored the client during the testing?

- Were all tests completed under monitoring in the office setting?
- What administration procedures did you use for the _____ test?
 - Exactly what are the standard administration procedures for the _____ test?
 - Did your procedures match the administration procedures that were published for the _____ test?
 - If not, why not?
 - If not, on what do you base your opinion that the test is reliable and valid?
 - If someone else administered, how did you ensure that they followed proper procedures?

Test Scoring

- What scoring procedures did you follow for the _____ test?
 - Exactly what are the standard scoring procedures for the _____ test?
 - How was each test scored? [hand or computer?]
 - Who scored each test?
 - Did you personally review the scoring?
 - Did you double-check the scores?
 - If the test was computer scored, what computer scoring system was used?
 - Where are the computer scoring procedures published?

Test Interpretation & Psychometric Properties of Scales

- What significant findings, if any, were evidenced on the _____ test?
 - What are the psychometric properties of the scale that represented these findings?
 - Reliability?
 - Validity?
- What interpretation procedures did you use for the _____ test?
 - Exactly what research supports these interpretation procedures? [names, dates, journals, findings]
 - Was a computerized interpretation system used to interpret the _____ test?
 - If so, which one?

[ethical guideline: even if computer interpretation is used, the expert is still responsible to know the reliability and validity of that program and procedures]
- What are your conclusions from the test data?

Research Support and Conclusions from Test Data

- What research supports your conclusions from the test data?
 - Reliability of research study?
 - Validity of the research study?
 - Author's names, journals, etc.?
 - Have the research findings been duplicated in other studies?
 - Does the study group include individuals similar to the client?
 - (cannot apply the study results if they are markedly dissimilar to the Client)
 - How do the study findings relate to the testing results?
 - How do the study findings related to parenting or the referral questions?
 - How does Texas public policy square with the your conclusions and recommendations?
- What other data is consistent (or inconsistent) with the test findings?
 - [ethical guidelines – gather data from multiple sources]
 - [Confront the expert with inconsistent fact data and require a proper explanation.]
- What, if any, are the limitations in the methods used?
 - [ethical guideline]
- What, if any are the limitations of the data used?
 - [ethical guideline]

Client Feedback

- Did you give the client an explanation of the test results?
 - [ethical guideline]
 - If not, did you obtain informed consent from the client to waive her or his right to feedback?

- If not, why not
- Did you give the client an explanation of the bases for your conclusions?
[ethical guideline]
 - If not, did you obtain informed consent from the client to waive her or his right to an explanation of the bases for your conclusions?
 - If not, why not?
[ethical guideline also notes that such feedback should be timely]

Documentation of Work

- Did you provide all the raw data that was produced as a result of your testing?
[ethical guideline]
- Did you document all of your significant findings and conclusions?
[ethical guideline]
- Did you maintain written records of your work?
[ethical guideline]

IX. CONCLUSION

No psychological test can perfectly measure a theoretical construct. Consequently, test results are inferential and subject to their own reliability and validity strengths or weaknesses. Psychological tests generate “hypotheses” about a person related to the purpose of the test. Test results must be fleshed out with additional information and, if such results are able to meet the appropriate reliability and validity thresholds AND are relevant to the referral question, then the test results likely are properly admissible. Nevertheless, the research indicates that it is common for psychologists conducting custody evaluations to administer tests which have poor reliability and validity properties and then rely upon such data in making recommendations without challenge from the opposing attorney. Hopefully, this paper will assist the opposing attorney make proper challenges to poor testing and will guide the proposing attorney in establishing a strong defense to a Daubert challenge if proper tests and procedures are used.

ADDENDUM A:
Sections of Ethical and Practice Guidelines which Apply to
Psychological Testing by Experts Conducting Child Custody Emulation

SPECIALTY GUIDELINES FOR FORENSIC PSYCHOLOGISTS

<http://www.unl.edu/ap-ls/foren.pdf>

- **Relationships (IV.)**
 - During initial consultation with the legal representative of the party seeking services, forensic psychologists have an obligation to inform the party of factors that might reasonably affect the decision to contract with forensic psychologist. These factors include, but are not limited to: (IV.A.)
 - The known scientific bases and limitations of the methods and procedures that they employ and their qualifications to employ such methods and procedures. (IV. A. 4)
- Forensic psychologists have an obligation to ensure that prospective clients are informed of their legal rights with respect to the anticipated forensic services, of the purposes of any evaluation, of the nature of the procedures to be employed, of the intended uses of any products of their services...(IV. E.)
 - In situations where the client or party may not have the capacity to provide informed consent to services or the evaluation is pursuant to court order, the forensic psychologist provides reasonable notice to the client's legal representative of the nature of the anticipate forensic service before proceeding. (IV. E. 2.)
- **Methods and Procedures (VI)**
 - Because of their special status. ...(experts) are obligated to use that knowledge, consistent with accepted clinical and scientific standards, in selecting data collection methods and procedures for an evaluation, treatment, consultation, or scholarly/empirical investigation. (VI.A.)
 - Forensic psychologists...When forensic psychologists conduct an examination or engage in the treatment of a party to a legal proceedings, with foreknowledge that their professional services will be used in an adjudicative forum, they incur a special responsibility to provide the best documentation possible under the circumstances. (VI.B.)
 - Documentation of the data upon which one's evidence is based is subject to the normal results of discovery, disclosure, confidentiality and privilege that operate in the jurisdiction in which the data were obtained. (VI. B.1)
 - The duties and obligations of forensic psychologists with respect to documentation of data that form the basis for their evidence apply from the moment they know or have a reasonable basis for knowing that their data and evidence derived from it are likely to enter into legally relevant decisions. (VI B.2)
 - When a forensic psychologist relies upon data or information gathered by others, the origins of those data are clarified in any professional product. In addition, the forensic psychologist bears a special responsibility to ensure that such data, if relied upon, were gathered in a manner standard for the profession. (VI. F.3)
- **Public and Professional Communications**
 - Forensic psychologists provide information about professional work to client in a manner consistent with professional and legal standards for the disclosure of test results, interpretations of data, and the factual bases for conclusions. A full explanation of the results of test and the bases for conclusions should be given in language that the client can understand. (VII.A.2)
 - When disclosing information about a client to third parties who are not qualified to interpret test results and data, the forensic psychologist complies with *Principle 16 of the Standards for Education and Psychological Testing*. When required to disclose results to a nonpsychologist, every attempt is made to ensure that test security is maintained and access to information is restricted to individuals with a legitimate and professional interest in the data. Other qualified mental health professionals who make a request for information pursuant to lawful order are, by definition, "individuals with a legitimate and professional interest". (VII.A.2.a)
 - In providing records and raw data, the forensic psychologist takes reasonable steps to ensure that the receiving party is informed that raw scores must be interpreted by a qualified professional in order to provide reliable and valid information.

GUIDELINES FOR CHILD CUSTODY EVALUATIONS IN DIVORCE PROCEEDINGS

www.apa.org/practice/childcustody.html

- The focus of the evaluation is on parenting capacity; the psychological and developmental needs of the child, and the resulting fit. (I. 3.)
- The psychologist uses current knowledge of scientific and professional developments, consistent with accepted clinical and scientific standards in selecting data collection methods and procedures. The *Standards for Educational and Psychological Testing* (APA 1985) are adhered to in the use of psychological test and other assessment tools. (III.5.B)
- The psychologist obtains informed consent from all adults participants... The psychologist informs adult participants about the nature of the assessment instruments and techniques and informs those participants about the possible disposition of the data collected. (III.9)
- The psychologist informs participants about the limits of confidentiality and the disclosure of information (III.10)
- The psychologist uses multiple methods of data gathering. (III.11)
The psychologist strives to use the most appropriate methods available for addressing the questions raised in a specific child custody evaluation and generally uses multiple methods of data gathering, including but not limited to, clinical interviews, observation, and/or psychological assessments. Important facts and opinions are documented from at least two sources whenever their reliability is questionable... psychologists corroborate it by at least one other source wherever possible and appropriate and document this in the report. (III.11)
- The psychologist neither over interprets nor inappropriately interprets clinical or assessment data. The psychologist refrains from drawing conclusions not adequately supported by the data. The psychologist interprets any data from interviews or tests, as well as any questions of data reliability and validity, cautiously and conservatively, seeking convergent validity. The psychologist strives to acknowledge to the court any limitations in methods or data used. (III.12)
- Recommendations, if any, are based on what is in the best interests of the child. ...Recommendations are based on articulated assumptions, data, interpretations, and inferences based upon established professional and scientific standards. Psychologist guard against relying on their own biases or unsupported beliefs in rendering opinions in particular cases. (III. 15.)
- The psychologist maintains written records. All records in the process of conducting a child custody evaluation are properly maintained and filed in accord with the APA Record Keeping Guidelines (APA 1993) and relevant statutory guidelines. All raw data and interview information are recorded with an eye toward their possible review by other psychologists or the court, where legally permitted. Upon request, appropriate reports are made available to the court.

ETHICAL PRINCIPLES OF PSYCHOLOGIST AND CODE OF CONDUCT.

www.aldea.com/guides/99/b315ud.html

- Psychologists provide services, teach, and conduct research with populations and in areas only within the boundaries of their competence, based on their education, training, supervised experience, consultation, study, or professional experiences (2.01(a))
- Where scientific or professional knowledge in the discipline of psychology establishes that an understanding of factors associated with age, gender, gender identity, race, ethnicity, culture, national origin, religion, sexual orientation, disability, language or socioeconomic status essential for effective implementation of their services ... (2.01 (b))
- When assuming forensic roles, psychologists are or become reasonably familiar with the judicial or administrative rules governing their roles. (2.01(f))
- Maintaining Competence: Psychologist undertake ongoing efforts to develop and maintain their competence (2.03)
- Bases for Scientific and Professional Judgments. Psychologists' work is based upon established scientific and professional knowledge of the discipline (2.04)
- Informed Consent. When psychologist conduct research or provide assessment, ...they obtain the informed consent of the individual or individuals using language that is reasonably understandable...(3.10(a))

- For persons who are legally incapable of giving informed consent, psychologist nevertheless (1) provide an appropriate explanation, (2) seek the individuals asset, (3) consider such persons' preferences and best interests, and (4) obtain appropriate permission from a legally authorize person, if such substitute consent is permitted or required by law. (3.10 (b))
- When psychological services are court ordered or otherwise mandated, psychologists inform he individual of the nature of the anticipated services, including whether the services are court ordered or mandated and any limits of confidentiality, before proceeding (3.10 ©)
- Psychologists discuss with person and organizations with whom they establish a scientific or professional relationship (1) the relevant limits of confidentiality and (2) the foreseeable uses of the information generated through their psychological activities (4.01 (a))
- Psychologists create, and to the extent the records are under their control, maintain, disseminate, store, retain, and dispose of records and data relating to their professional and scientific work in order to (1) facilitate provision of services later by them or by other professionals,...(5) ensure compliance with the law. (6.02 (a))
- Bases for Assessments. 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. (9.01(a)).
- Except as noted in 9.01c, psychologists provide opinions of the psychological characteristics of individuals only after they have conducted an examination of the individuals adequate to support their statements or conclusions. When, despite reasonable efforts, such an examination is not practical, psychologists documents the efforts they made and the result of those effort, clarify the probable impact of their limited information the reliability and validity of their opinions, and appropriately limit the nature and extent of their conclusions or recommendations. (9.01(b)).
- When psychologists conduct a record review or provide consultation or supervision and an individual examination is not warranted or necessary for the opinion, psychologists explain this and the sources of information on which they based their conclusions and recommendations. (9.01(c)).
- Use of assessments. 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 research on or evidence of the usefulness and proper application of the techniques. (9.02(a)).
- Psychologists use assessment instruments whose validity and reliability have been established for use with member of the population tested. When such validity or reliability has not be established, psychologists describe the strengths and limitations of test results and interpretation. (9.02(b))
- Psychologists use assessment methods that are appropriate to an individual's language preference and competence, unless the use of an alternative language is relevant to the assessment issues (9.02 (c)).
- Psychologists obtain informed consent for assessments, evaluations...Informed consent includes an explanation of the nature and purpose of the assessment, fees, involvements of third parties, and limits of confidentiality and sufficient opportunity for the client/patient to ask questions and receive answers. (9.03(a))
- Psychologists inform persons with questionable capacity to consent for whom testing is mandated by low or governmental regulations about the nature and purpose of the proposed assessment services, using language that is reasonably understandable to the person being assessed. (9.03(b))
- Release of Test Data. The term test data refers to raw and scaled scores, client/patient responses to test questions or stimuli, and psychologists' notes and recordings concerning client/patient statements and behavior during an examination. Those portions of test materials that include client/patient responses are included in the definition of test data. Pursuant to a client/patient release, psychologists provide test data to the client/patient or other person identified in the release. Psychologists may refrain from releasing test data to protect a client/patient or others from substantial harm or misuse or misrepresentation of the data or the test, recognizing that in many instances release of confidential information under these circumstances in regulated by law. (/9.04 (b)).
- Interpreting test results. 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 interpretation. (9.06)
- Assessment by unqualified persons. Psychologist do not base their assessment or intervention decisions or recommendations on data or test results that are outdated for the current purpose (9.08)

- Obsolete Tests and Outdate Test Results. Psychologists do not base their assessment or intervention decisions or recommendations on data or test results that are outdated for the current purpose. (9.08 [a])
- Psychologists do not base such decisions or recommendations on tests and measures that are obsolete and not useful for the current purpose. (9.08[b])
- Test Scoring and Interpretation Services. Psychologists who offer assessment or scoring services to other professionals accurately describe the purpose, norms, validity reliability, and application and the procedures and any special qualifications applicable to their use. (9.09[a])
- Psychologists select scoring and interpretation services (including g automated services) on the basis of evidence of the validity of the program and procedures as well as on other appropriate considerations (9.09 [b])
- Psychologists retain responsibility for the appropriate application, interpretation, and use of assessment instruments, whether they score and interpret such test themselves or use automated or other services. (9.09[c])
- Explaining Assessment Results. Regardless of whether the scoring and interpretation are done by psychologists, by employees or assistants, or by automated or other outside services, psychologists take reasonable steps to ensure that explanations of results are given to the individual or designated representative unless the nature of the relationship precludes provision of an explanation of results (such as in some organizational consulting, pre employment or security screenings, and forensic evaluation), and this fact has been clearly explained to the person being assessed in advance.(9.10)
- Maintaining Test Security. The term test materials refers to manuals, instruments, protocols, and test questions or stimuli and does not include test data as defined in Standard 9.04, Release of Test Data. Psychologists make reasonable efforts to maintain the integrity and security of test materials and other assessment techniques consistent with law and contractual obligations, and in a manner that permits adherence to this Ethics Code. (9.11)

RIGHTS AND RESPONSIBILITIES OF TEST TAKERS: GUIDELINES AND EXPECTATIONS

<http://www.apa.org/science/ttrr.html> [Summary]

As a test taker, you have the right to:

1. Be informed of your rights and responsibilities as a test taker.
2. Be treated with courtesy, respect, and impartiality, regardless of your age, disability, ethnicity, gender, national origin, religion, sexual orientation or other personal characteristics.
3. Be tested with measures that meet professional standards and that are appropriate, given the manner in which the test results will be used.
4. Receive a brief oral or written explanation prior to testing about the purpose(s) for testing, the kind(s) of tests to be used, if the results will be reported to you or to others, and the planned use(s) of the results. If you have a disability, you have the right to inquire and receive information about testing accommodations. If you have difficulty in comprehending the language of the test, you have a right to know in advance of testing whether any accommodations may be available to you.
5. Know in advance of testing when the test will be administered, if and when test results will be available to you, and if there is a fee for testing services that you are expected to pay.
6. Have your test administered and your test results interpreted by appropriately trained individuals who follow professional codes of ethics.
7. Know if a test is optional and learn of the consequences of taking or not taking the test, fully completing the test, or canceling the scores. You may need to ask questions to learn these consequences.
8. Receive a written or oral explanation of your test results within a reasonable amount of time after testing and in commonly understood terms.
9. Have your test results kept confidential to the extent allowed by law.
10. Present concerns about the testing process or your results and receive information about procedures that will be used to address such concerns.

ORDERING INFORMATION: STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING (1999) 1-800-628-4094, List price 31.95. Website order form: <http://www.apa.org/science/standards.html>.

Sections and Chapters of the Standards:

Part I: **Test Construction, Evaluation, and Documentation** (1. Validity, 2. Reliability and Errors of Measurement, 3. Test Development and Revision, 4. Scales, Norms, and Score Comparability, 5. Tests Administration Scoring and Reporting 6. Supporting Documentation for Tests)

Part II: **Fairness in Testing** (1. Fairness in Testing and Test Use, 8. The Rights and Responsibilities of Test Takers, 9. Testing Individuals of Diverse Linguistic Backgrounds, 10. Testing Individuals with Disabilities)

Part III: **Testing Applications** (11. The Responsibilities of Test Users, 12. Psychological Testing and Assessment, 13. Educational Testing and Assessment, 14. Testing in Employment and Credentialing, 15. Testing in Program Evaluation and Public Policy).

ⁱ The Expert Witness Manual (1999) Family Law Section, State Bar of Texas

ⁱⁱ Other ethical codes can apply but these five are the most important for psychologists testing in the context of custody evaluations.

ⁱⁱⁱ Texas Public Policy

^{iv} Note that legal cases use the term reliability as synonymous with validity. In social science research, reliability and validity have different meanings. In social science research, reliability refers to consistency in measurement only. So there would be instances where a test might be highly reliable, i.e. the test yields highly consistent scores but not valid, i.e. the test is not measuring what it is supposed to measure. Validity refers to the test measuring what it is supposed to measure. Validity is built upon reliability. A test cannot be valid if it has poor reliability.

^v Bow, J. M. & Quinnell, F. A. (2001). *Psychologists' Current Practices and Procedures in Child Custody Evaluations Five Years After American Psychological Association Guidelines*, Journal of Professional Psychology: Research and Practice, Vol 32, No. 3, 261-268.

^{vi} Bow, J. M. & Quinnell, F. A. (2001). *Psychologists' Current Practices and Procedures in Child Custody Evaluations Five Years After American Psychological Association Guidelines*, Journal of Professional Psychology: Research and Practice, Vol 32, No. 3, 261-268. [<http://spider.apa.org/ftdocs/pro/2001/June/pro323261.html>, p 11]

^{vii} Ackerman, M. J. & Ackerman, M. C. *Custody Evaluation Practices: A Survey of Experienced Professionals (Revisited)* 2 Professional Psychological Researches and Practice, 137,140 (1997).

^{viii} Ackerman, M & Ackerman M. (1997) *Child Custody Evaluation Practice: A Survey of Experienced Professionals (Revisited)* 28 Professional Psychology: Research & Practice, No.2, 137-145. Note that this study assessed 800 examiners but only 201 of the returned questionnaires were used in the study. This study assessed the practices of all custody evaluators, not just psychologists. However, the findings are in keeping with an earlier study (1986) by Keilin and Bloom [Kelin, W.B. & Bloom, L.:J. (1986) *Child Custody Evaluation Practices: A Survey of Experienced Professionals*, 17, Professional Psychology: Research & Practice, 338-46.], which surveyed 82 examiners (not just psychologists).

^{ix} Paolo, Ryan & Smith (1991)

^x Research studies which Ares considered to be more reliable and valid are found in referred journals published by the American Psychological Association. These journals use hundreds of social scientists who critically review and evaluate the soundness of the studies. The Exner group has been criticized a great deal because they have issued most of their studies from "in-house" publications.

^{xi} Exner, J. (1990). *Rorschach Workbook*, p 158.

^{xii} Lilienfeld, S. O. et al (2000). *The Scientific Status of Projective tests*, 1 Psychological Science in the Public Interests, 27,31.

^{xiii} Will, T. (1994) *The MCMI-III and Personality Disorders*. Wisconsin Psychological Association. Originally cited in Ackerman & Ackerman (1998). Anastasi, A. & Urbina, S. (1997) *Psychological Testing* p 369.

^{xiv} Ratliff, P. (1996). *MCMI-III Diagnostic Validity: Bad Test of Bad Validity Study*, 66 Journal of Personality Assessment 431-37. ; Rogers, R. et al. (1997). *Convergent and Discriminate Validity of the Millon Clinical Multiaxial Inventory for Axis II Disorders: Time of a Moratorium?* Presented at the 105th Annual meeting of the American Psychological Association (Chicago). ; McCann, J.R. & Dyer F.J. (1996). *Forensic Assessment with the Millon Inventories*, 3.

^{xv} Retzlaff, P. (19967). *MCMI-III Diagnostic Validity: Bad Test or Bad Validity*, 66 Journal of Personality Assessment, 43.

^{xvi} Murray, H. (1943). *Thematic Apperception Test Manual*, 2

^{xvii} Weithorn, L.A. & Grisso ,T. (1987). *Psychological Evaluations in Divorce Custody: Problems, Principles, and Procedures, in Psychology and Child Custody Determinations: Knowledge, Roles, Expertise.*