Chapter 1
Introduction: Outlining the Problem

The phrase "valid custody evaluations" largely remains a paradox. There is no clear evidence that custody evaluators can conduct evaluations that are, in fact, in the "best interests of the child." This is a serious and important legal and societal problem. In dissolution of relationships, courts often rely on mental health professionals to aid them in making decisions so that harm to children is minimized and their best interests are served. But do mental health professionals know how to achieve these goals? (O'Donohue & Bradley, 1999). What is the scientific evidence that mental health professionals can actually meet the expectations and needs of the court, and produce an accurate account of arrangements that will serve the best interests of the child? What is the evidence that two evaluators will make similar recommendations in the same case (i.e., the inter-rater reliability of custody evaluations)? Is this a systemic problem, i.e., the knowledge of valid assessment protocols and decisions about custody dispositions based on these is simply not available? That is, has the science not progressed to this degree? Or, is it also in part that there is a range of competence of individual custody evaluators and that some can achieve this end better than others? If there is a quality crisis in custody evaluations, what should be done? Should there be a moratorium so that at least mental health professionals are following the Hippocratic ethical dictum of "at least doing no harm"? (O'Donohue & Bradley, 1999). How should professionals conducting custody evaluations be held accountable by the courts, by parents, and by professional ethical boards?

Children are affected in many ways, and for many years, by the recommendations that mental health professionals make in custody evaluations. These evaluations influence the amount of time children spend with each parent and under what conditions (i.e., supervised visitation, weeknights, weekends, overnights, etc.). Recommendations can influence whom the child spends important holiday and birthdays with; whether relocation on the part of one of the parents is allowed or disallowed and how this affects time with the remaining parent. These recommendations can last for a long period of time—in some cases 17 years, i.e., until the child comes of age (which of course makes the prediction task of the evaluator all that more difficult). Of course, these recommendations also affect parents, grandparents, and other relatives and friends.

L.W. Tolle and W.T. O'Donohue, *Improving the Quality of Child Custody Evaluations*,
DOI 10.1007/978-1-4614-3405-4_1, © Springer Science+Business Media, LLC 2012

Mistakes or errors in judgment made by an evaluator can be detrimental to the child's well-being in a number of ways. If child abuse or neglect is alleged by one parent and the evaluator erroneously determines that there is no sufficient evidence to support the allegation [e.g., no reports to Child Protective Services (CPS); child is not alleging this], the recommendations made by this evaluator might place a child in a harmful environment where further abuse could occur. In relocation cases, where the child moves with the primary parent, the child could suffer relational problems with the parent who does not move. Given the number of crucially important factors associated with a custody evaluation, one would hope that all precautionary measures are taken to ensure that the best interest of the child (BIC) is upheld. Disturbingly, however, there is little science guiding evaluators so that they can make valid, empirically driven recommendations for child custody (Emery, Otto, & O'Donohue, 2005; O'Donohue & Bradley, 1999).

The prominent philosopher of science Laudan (1977) has suggested that science has both conceptual and empirical problems. All too often psychologists have acted as if there are only empirical problems. Laudan points out that scientific progress also is made when progress is made in a science's conceptual problems. For example, psychologists have done more work trying to detect possible differences between ethnic groups in intelligence (an empirical problem) than in solving the conceptual problem of what is intelligence? This has led the eminent analytic philosopher Wittgenstein (1940) to comment.

> The confusion and barrenness of psychology are not to be explained by calling it a "young science"; its state is not comparable with that of physics, for instance, in its beginnings. (Rather with that of certain branches of mathematics. Set theory.) For in psychology there are experimental methods and *conceptual confusion*. (As in the other case, conceptual confusion and methods of proof.) The existence of the experimental method makes us think we have the means of solving the problems that trouble us; though problem and method pass one another by (PI p. 232).

We think, similarly, custody evaluations and the child best interest standard "pass one another by" to use Wittgenstein's felicitous phrase. There is conceptual confusion regarding child custody evaluations and important conceptual problems to be solved. A key cause of this confusion, in our view, is the lack of clarity and explication of the key standard underlying these—the best interests of the child.

Rates of divorce have been reported to range from 40% to 60% in the USA (divorcerate.org). This is a slightly misleading statistic, however, as it combines first, second, and third marriages and divorce rates have been found to be higher in second and third marriages. Specifically, approximately 41% of first marriages, 60% of second marriages, and 73% of third marriages end in divorce (Baker, 2003).

Divorce rates have skyrocketed over the last 30 years. Many reasons have been purported to explain the dramatic increase seen in the late 1960s and 1970s, including a greater number of women entering the workforce, an increase in feminism and feministic attitudes, and the adoption of the "no-fault" divorce (divorce granted without having to establish wrongdoing by either party), making a divorce much easier to obtain (Powell, 2003). An estimated 660,000 divorces occurred between 2006 and 2008 in the USA, though this is likely an underestimate as not all states

provide or keep track of divorce counts (i.e., California, Georgia, Hawaii, Indiana, Louisiana, and Minnesota) and monthly counts might be underreported (National Center for Health Statistics, 2008).

Divorce Rates in Families with Children

Along with the overall dramatic increase in divorce in the 1970s, the number of couples with children who divorced increased by 700% from 1900 to 1970 (Davis, 1977). Similarly, the 1998 Census Bureau reported that children under the age of 18 living with one parent increased from 12% in 1970 to 28% in 1996, and children living with both parents decreased from 85% to 68% during the same time span (U.S. Census Bureau, 1998). Despite these statistics, divorce rates have been found to be lower in couples who have children than in childless couples, with estimates of 40% of couples with children divorcing and 66% of childless couples divorcing (Heaton, 1990). In attempting to account for this discrepancy, researchers have reported that the attitude to "stay together for the sake of the family" (Thornton, 1985) may be an important reason that parents do not divorce. Another potential deterrent that has been reported is the financial expense related to both the cost of a divorce and the decrease from a dual-parent income to a single-parent income (Albrecht, Bahr, & Goodman, 1983). Finally, infertility in couples has been associated with an increase in risk for divorce (Myers, 1997). Despite the seemingly prophylactic effect children can have on marriage stability, there are still many cases for which this is an insufficient deterrent (Willats, 1993).

Annually, approximately one million children in America are involved in a divorce (American Academy for Child and Adolescent Psychiatry, 1997). In addition, close to one-half of children from divorced homes will also witness the dissolution of a parent's second marriage (Furstenberg, Peterson, Nord, & Zill, 1983). Additionally, approximately 100,000 custody evaluations occur annually in the USA. Despite the number of child custody evaluations completed every year, there is little research guiding how these evaluations should be conducted so that they actually describe the best interests of the child. Standards for establishing that child custody evaluation models are empirically supported should be comparable to standards for establishing that psychological treatments are empirically supported. In order for a treatment to be considered empirically supported (or validated) per the Division 12 Task Force (Chambless et al., 1998):

1. It must be supported by at least two randomized, controlled trials showing their superiority to placebo control conditions or another established treatment with appropriate sample sizes to detect significant differences.
   Or
   It must be supported by a number of single-case designs that involved good experimental design and comparison of one treatment to another.

2. The studies must be conducted with treatment (or in this case assessment) manuals or some equivalent.
3. Characteristics of the client samples must be clearly specified.

When applying these standards to models of custody evaluations, it is found that not only they meet none of these criteria—there is no outcome testing of the consequences of custody evaluations—but also no manualized models for conducting evaluations have been established. To be clear, currently there is no evidence that custody evaluations bring about superior outcomes for children involved in these. Put more pessimistically, there is also no evidence that custody evaluations do more good than harm. Longitudinal research is urgently needed which shows the predictive validity of custody evaluations.

In order to create some sort of floor for the quality of custody evaluations, there are some very vague standards regarding custody evaluations such as those promulgated by the American Psychological Association—and although these may decrease the likelihood of truly egregious evaluations (e.g., the standard's stipulation that multiple sources of information must be utilized would rule out a single source evaluation) but there is still too much room for unwanted variance. In addition, and partly because there are no existing manualized models, none have been tested or compared against anything else in order to determine their accuracy or error rates (i.e., no one model for conducting evaluations has been compared to another or any control group, mediation, etc.). What appears to exist instead of this more systematic process is: (1) custody evaluations with unwanted variability (i.e., evaluators use "clinical judgment" to interpret disparate collected data and use various measurement tools, each with its own error term, that they, for various reasons, have come to believe are relevant perhaps only because the tests seem face valid); (2) custody evaluations conducted in the context in which there is no research evaluating predictive validity of these (i.e., what is the long-term accuracy of what evaluators are predicting—to what extent are the predictions of best interests accurate?); (3) custody evaluations conducted in the context of no evidence of inter-rater reliability (i.e., if two evaluators were to conduct evaluations with the same family—would they come out with the same recommendations?); and (4) no evidence of construct validity (what data are important, what are not, and for what reasons?).

Thus, it might be said that there is a "quality crisis" with regard to custody evaluations. For decades these have been conducted in idiosyncratic ways, with substantial variance in the methods, and with no known accuracy. It is unclear, but quite discouraging to contemplate, how many lives have been affected by this quality problem. It might not be too much to say that the poor quality standards in this area have been a form of systematic child abuse.

There are also several important conceptual issues that O'Donohue and Bradley (1999) have argued need to be settled before higher quality custody evaluations can take place. First, there needs to be a model for what constructs ought to be measured. Ought parent–child relationships be measured, and if so, why? Ought parental psychopathology be measured, and if so, why? It is critical to enumerate the entire set of constructs that are relevant to the Best Interests of the Child. To date, this important—even foundational—task has not been done. Instead, to date evaluators have

based their evaluations on inchoate, informal models of the child's best interests. For example, Stahl (2010) recently approvingly quoted the legal standard of one state—Michigan's—as if this state had adequately captured the standard. This unsystematic practice has several important negative consequences: (1) there will be unwanted variability in the models relied upon across clinicians resulting in poor inter-rater reliability, and (2) the full logic of the custody evaluation will remain hidden and thus there will be a kind of unaccountability. When the constructs are fully explicated, their measurement operations explicated and the logic of the decisions clearly stated, all parties can more clearly understand and critique the custody evaluations. Such feedback is important for quality improvement and for custody evaluations to become a more rational process. Without this they have a kind of Wizard of Oz aura in which a clinician picks constructs out of the air, measures these in unique ways, and synthesizes these mysteriously to produce custody evaluations. This is puzzling, and perhaps harmful, for the adults and children involved.

This absence of a firm conceptual foundation for custody evaluations is a distressing reality and has been the case for decades. As courts look to mental health professionals to conduct child custody evaluations, it is incumbent on clinical scientists to develop models and tools to begin to demonstrate consistency of process, and the reliability and validity of the inferences made from these evaluation processes. It is important to conceptualize what a solution to this complex problem would look like.

Purpose of This Book

A model of custody evaluations that is construct valid (i.e., the extent to which operationalization of the constructs actually measures what it purports to measure), that serves as a format to result in improved inter-rater reliability of custody evaluations and that has known and acceptable predictive validity would be an important first step. In looking at such a model, it would be necessary to first define predictive criteria. This book attempts to develop such a model, which would ultimately guide the assessment process in custody evaluations. The model will be informed by empirical literature that predicts best outcomes for children.

This book has four main aims: (1) to discuss the past and current state of science regarding child custody evaluations with a focus on the construct of the BIC, (2) to propose a preliminary model (the egregious/promotive factors model, EPFM) that is based on an extensive review of the empirical research related to factors that have been found to be predictive of poorer or positive outcomes in children in order to conduct improved child custody evaluations, (3) to identify and review existing psychological assessments that can reliably measure risk and promotive factors to support the EPFM, and (4) to gain feedback and preliminary support for the EPFM, a pilot study involving family court judges was conducted examining child custody decision making when provided either a EPFM-guided report or an unspecified constructs report.