

Chapter 19

Psychometric Analysis of Forensic Interviews and Post Hoc Interview Evaluations

William T. O'Donohue and Matthew Fanetti

Evaluating Forensic Interviews with Children Who May Have Been Sexually Abused

Adults in many settings sometimes question children with the goal of discovering information about their experiences. When a variety of professionals (e.g., police officers, social workers, clinical psychologists, physicians) engage in this for the purpose of gaining information that is legally relevant this process is generally called “a forensic interview.” When the goal of the forensic interview is to discover from the child some information about whether or not he or she has been sexually abused, and if so, details about this abuse, this process may be called “a forensic interview of a child regarding sexual abuse status.”

These have had a long and some might even say a notorious history (Rabinowitz, 2004). There have been cases—most notably the McMartin case in the 1980s in Manhattan Beach, California—but also others (e.g., a daycare case in Edenton, NC and the Kelly Michaels case in New Jersey) where these forensic interviews were done so shoddily that in all likelihood false allegations were created by these interviews or the interview failed to uncover the truth that the children had not been sexually abused by the adult or adults in question. As a result of these problematic interviews, a variety of harm was done—reputations were ruined, innocent individuals were prosecuted, some innocent individuals were imprisoned for several years, families were stressed and even torn apart, millions of dollars were needlessly spent,

W.T. O'Donohue
Department of Psychology, University of Nevada, Reno, MS 0296, Reno, NV 89557, USA

M. Fanetti (✉)
Department of Psychology, Missouri State University,
901 S. National Ave., Springfield, MO 65897, USA
e-mail: mfanetti@missouristate.edu

and children were given false identities as abuse victims. On the other hand, errors certainly have also occurred in forensic interviews in which children who actually have been abused but the interview failed to elicit this information. Not only is this a failure of the interview to accurately uncover actual abuse status, and not only is there a grievous failure of justice, but these children can be returned to environments where they can be further victimized.

No one disputes that these forensic interviews ought to be done well. What is less clear is how exactly should these interviews be conducted so that they are indeed done well. This clearly involves a meta-question: By what standards ought the quality of a forensic interview with a child who may have been sexually abused be evaluated? A related question is what kind of, or how much, error can one of these forensic interviews contain and still be used to make reasonably accurate conclusions about the child's abuse standards? This chapter will examine these issues. We will examine various criteria for evaluating interviews. We will conclude that there is much missing evaluative information regarding these interviews, that there is reason to be concerned about gaps in what is currently known about the psychometrics of forensic interviews, and there are other issues that are also troubling when evaluating these interviews. We call for a prioritization of corrective action on these problems.

Partly in response to the question of quality of the interview, interview protocols have been developed (see Cirlugea and O'Donohue (2015) for reviews of the major interview protocols). The basic idea behind the development and use of an interview protocol is to provide some standardization—certain stages or moves in the interview will always occur—certain moves will be avoided, and a certain temporal sequencing will occur. Without this standardization, psychometric properties cannot be determined; for example, reliability and validity need to be of some relatively fixed measurement process. Interview protocols attempt to decrease heterogeneity in these interviews. If interviewing were a “free for all”—any interviewer could make up anything as they go along—there would simply be no fixed process that can be evaluated for its strengths and weaknesses. It is important to note that a certain amount of forensic interviews of children do not follow any protocol—they are in fact—“make it up as you go” and a weakness of this is that the accuracy of these is entirely unknown. A final caveat is called for—if an interview protocol is modified then its psychometrics are also unknown. That is, if some entity modifies a known psychometric protocol, the interview becomes different and one cannot assume the psychometrics of the original protocol remain unchanged.

In addition, the basic idea is that these interview protocols would also contain the “best thinking” about how these interviews of children ought to be conducted so that accuracy is maximized—for example, the best thinking suggests that it is important to initially establish rapport with the child so this is a first stage in most interview protocols; and the best thinking generally includes the importance of avoiding suggestive questioning, so interview guidelines define suggestibility and define strategies to avoid this. However, it is also important to note that all these interviews protocols are what is called *semistructured*—these do not contain an invariant script that the interviewer must follow. Rather some flexibility and hence heterogeneity

must be allowed to accommodate the wide range of unique values of key variables encountered in the interview (e.g., the child's prior answers, the developmental level of the child, the nature of the abuse itself, unique difficulties the child may have).

One other preliminary matter ought to be quickly addressed. First, some might claim that these interviews of children ought not to be evaluated—these are simply “hearing what the child has to say.” This view is false and more than a bit naïve for three reasons: (1) this ignores the vast literature of child suggestibility (see x this volume) as well as the past record of problematic interviews resulting in false conclusions about child abuse status. Interviewing is a complex, protracted, interpersonal process that involves memory, information processing, interpersonal influence, and other psychological processes; and (2) this view ignores the fact that even if the interview is construed as “just hearing what the child has to say” this is still a measurement task—we want to detect “what the child has to say” and we can make errors in doing this. It is a basic principle of psychometrics that all measurement contains error—although certainly some contain more than others—and the goal of psychometrics is to understand the kind and degree of error of any measurement task—even if the measurement task is “just hearing what the child has to say” (Haynes, Smith, & Hunsely, 2011). Finally, the view that there is little concern about the degree of accuracy of these interviews is perplexing because it assumes that no matter what the interviewer does, the interview will invariably and inevitably produce accurate information from the child. This seems implausible particularly because we know that children can contradict themselves across interviews and by logic contradictory statements cannot both be true.

Evaluative Criteria for a Forensic Interview

How should forensic interviews of children who may have been sexually abused be evaluated? Is it sufficient to show that if the interviewer has earned some sort of general professional credentials that the interview was sound—and if so, what credentials—a mental health license in the state, a certificate of some sort of training completion? Is it sufficient for some interviewer to simply show that they have some sort of general past experience in some area relevant to interviewing children who may have been sexually abused, say in childsexualabuse or clinical interviewing? These seem to be the de facto standards of competence in the field—perhaps because these kinds of credentials generally can pass legal muster during a voir dire process. However, psychometrically these seem quite inadequate.

One can raise obvious questions about these sorts of standards:

1. Which credentials exactly? No credentials have been shown to actually assure accuracy of the forensic interview of the child.
2. To what extent does the typical generic training in a variety of professions (social work, clinical psychology, pediatrics, police work) actually make one a competent forensic interviewer of children? Most of this generic training would

give short shrift to this particular domain and thus would be unlikely to produce competence.

3. How much training in the forensic interviewing of children who may have been sexually abused is sufficient? No training has been shown to be necessary or sufficient for assuring accuracy of these interviews.
4. Does this training need to be refreshed—as drift is often the case in faithfully following protocols? Very little information is available on the durability of training on interview fidelity.
5. Ought there be some sort of rigorous test to show that the training was at least initially successful—and if so what is an adequate test of this competence and what should the cut score for competence be? Again, these sorts of questions have been ignored in the literature and in practice. No test score on any test has been shown to assure forensic interview accuracy.
6. Which of the variety of interview protocols ought interviewers be trained in, and why? How psychometrically sound is the protocol? Cirlugea and O'Donohue (2015) point out the vast amounts of missing psychometric information on the protocols used to interview children who may have been sexually abused.
7. Does one become globally competent after this training, that is, competent to interview a child of any age, any child from any culture, and even a child with any special circumstances (e.g., being developmentally delayed)? This question has very little data and is unsettled.

These are tough questions that are generally ignored in the “x was trained in a protocol” or the “x is a licensed professional with some sort of experience in childsexualabuse” views of competency. Thus, we conclude that this standard of evaluation is unsatisfactory.

An Interview Protocol's Sensitivity and Specificity

Typically in medicine a fairly quick and dirty evaluation of a test's overall quality can be given by the metrics of *sensitivity* and *specificity*. A measure's sensitivity is the probability that it will detect x if x is actually present. That is, it is the probability that if x is present (say cancer) that the test will indicate that the cancer is present. Alternatively, a test's sensitivity can be said to be the probability of a false negative—a test has failed to be sensitive if it says that no cancer is present, when in fact cancer is present. Obviously in this case, a forensic interview would fail to be sensitive if it concludes that abuse has NOT occurred, when in fact it has.

Specificity is the converse and equally important. A test can assure perfect sensitivity if the test always indicates “X is present”—after all if the test says x is always present then the measure would never miss an actual incidence of presence. However, the problem with this strategy is that the test would produce a number of false positives—the test would indicate that x is present when it is in fact not. The probability of a false positive is a test's *specificity*—a desirable characteristic

for a test to have is to say *x* is NOT present when in fact *x* is NOT present. Obviously in a forensic interview with children who may have been sexually abused it is important to conclude that the child has not been sexually abused when the child has not been sexually abused.

As Cirlugea and O'Donohue (2015) have pointed out as a field we do not know the sensitivity or the specificity of the protocols we use to interview children who may have been sexually abused. This is a serious gap in our knowledge. As previously stated this gap becomes even more of a concern when interviews are made up on the fly—that is, that the interviewer fails to follow any protocol—because in principle this *ad hoc* process can have no known sensitivity or specificity—it is not sufficiently constant or fixed to become an object of study. Thus, currently, if an interviewer is questioned along the following lines:

1. What is the rate of false negatives of the interview you administered?
2. What is the rate of false positives of the interview you administered?

The interviewer unfortunately would simply have to say, “These are unknown.” This is quite problematic as we do not know if one or both of these are high—and perhaps so high as to render serious concerns about the interview’s error rates. Moreover, as a field we don’t know which is higher for a particular protocol—false negatives or false positives and this for a variety of reasons seems important to know. If for example, when false negatives of an interview protocol are high—we ought to be more cautious about concluding a child has not been abused even though this is the conclusion of the interview.

In addition, it must also be recognized that there actually might be a series of sensitivities and specificities of a forensic interview protocol that ought to be known. For example, in oncology there is not simply a single sensitivity of biopsies, the sensitivity can vary to the type of cancer, the size of the mass examined, etc. There is a parallel situation here—there may be a range of sensitivities and specificities depending on variables such as the age of the child being interviewed, the experience of the interviewer, the severity of the abuse, the relationship between the abuser and the child, etc.

Finally, it must be recognized that the use of a particular protocol actually involves two separate issues: (1) the sensitivity and specificity of the interview protocol but also (2) evidence that the interviewer faithfully followed the interview protocol. Thus, the actual interview must be examined to determine the extent to which it faithfully followed the interview protocol. It is a difficult question—and again unsettled in the field—of how much deviation is permissible? More basically there is little known about how to assess fidelity to a protocol. However, it should not simply be assumed that a particular interviewer faithfully followed a protocol.

What is probably most concerning currently is that there has not been and there currently seems to be little urgency in the field for addressing this critical knowledge gap. Forensic practice seems to be relatively complacent with the ignorance of the sensitivity and specificity of forensic interviews of children who may have been sexually abused. Admittedly, this research can be difficult to conduct but the question raised in a clinical science approach to practice is, “Are these research difficulties

sufficient to excuse practice that may contain a level of error that is injurious to our consumers?" This is particularly perplexing because the Daubert criterion of legal admissibility is that the instrument or theory must have a known error rate and these interview protocols do not meet this criterion, yet surprisingly these interviews are routinely admitted in court.

Other Important Psychometric Properties

It is fair to say that the Standard for Educational and Psychological Testing is an important but underutilized document in the field of interviewing children who may have been sexually abused. This document lists several other important evaluative criteria that ought to be known about any measure including:

1. Interrater reliability (the extent to which two different interviewers will arrive at the same conclusions)
2. Test–retest reliability (the extent to which two interviews given at two different times will have the same results)
3. Split half reliabilities (the extent some half of a test agrees or is consistent with some other half)
4. Construct validity (the extent to which the score of this measure agrees with the score of another measure of the same construct—there can be a variety of constructs involved in a forensic interview of a child—rapport, truth knowledge, prepositional competence, etc.)
5. Postdictive validity (the extent to which the score of the measure agrees with some criterion in the past, e.g., actual abuse history)
6. Incremental validity (the extent to which some measure adds or subtracts to the accuracy of some criterion, e.g., the extent to which the forensic interview adds or subtracts accuracy from other information such as medical tests)

Cirlugea and O'Donohue (2015) reviewed psychometric information on several of the protocols and find much missing information although clearly the NICHD protocol has the most known psychometrics known at this point in time. Again, it must be emphasized that "adaptations" of some known protocol or free-form interviews would be missing this key evaluative information.

In addition, the suggests other key information ought to be present for a measure. For example:

1. A manual that will aid the administrator in proper administration procedures, offer information on limitations of the test, and provide information on correct scoring and interpretation.
2. The provision of information related to fair testing and limitations of this (e.g., perhaps testing with some cultural groups results in decreased validity).

In general, there are two other issues raised by the Standards for Educational and Psychological testing. First, there is little information about proper scoring and

interpretation of the results of the forensic interview with children who may have been sexually abused. For example, consider the following situations:

- (a) A child in the initial part of an interview says that she has not been sexually abused but in a later part of the interview says that she has been abused. How is the interviewer to interpret this inconsistency? What if the sequencing were reversed—initially the child says she was abused but in a subsequent part of the interview she says she was not? Does this order affect the conclusions? How is the interviewer to interpret these inconsistencies?
- (b) How is the interviewer to interpret a child's report that contains fantastical details, for example, a teacher touched them inside their underwear and while this was occurring real witches were flying around the room?
- (c) What kind of conclusions is the interviewer to make in an interview in which the child gives very little detail about the abuse, for example, "My stepfather touched my chest" but can elaborate no further—cannot recall the number of times, who was present in the home when this allegedly occurred, how many times this occurred, how long it occurred, and cannot give a time of day or date, etc. How much does this lack of detail affect the conclusions of the interview? Of course it would seem that this would partly depend on the child's age but exactly how? Exactly how much detail can we expect from an average 5 year old vs. an average 9 year old?
- (d) What kind of conclusions is the interviewer to make in an interview in which implausible details are given, for example, when a 4-year-old child reports that an adult anally penetrated them but that he or she felt no pain?
- (e) How is an interviewer to interpret an interview when a parent engaged in suggestive interviewing practices before the interview took place (e.g., repeatedly asking several dozen times if an uncle touched them before the child made an outcry)?
- (f) What kind of conclusions is the interviewer to make in an interview in which a child claims that they completely forgot the abuse for several years but then suddenly remembered all of the abuse—a so-called recovered memory?
- (g) How is the interviewer to interpret a child's denial of abuse when several other witnesses say they saw the child being abused?
- (h) What kind of conclusions is the interviewer to make in an interview that contains a child's responses when English is not their first language and they displayed some difficulties either comprehending or expressing themselves in English—yet there was no interpreter in their native language available?
- (i) Finally, how does an interviewer interpret a child's statements in the interview when the child has a rather extensive history of lying—perhaps even displayed in the interview about topics other than abuse (e.g., prizes the child has won)?

These are difficult questions—and no doubt depend on a number of other details surrounding the particular case. However, it is too infrequently recognized that there is in fact an interpretation task at the end of a forensic interview of a child who may have been sexually abused—and these interpretations have a potential for error. Rarely do these interview data "speak for themselves." These conclusions can even

be more problematic if the interviewer has a bias—is prone to interpret data in one way or another (exculpatory vs. incriminating)—this issue will be discussed more below.

The second issue to be recognized is there are actually multiple components of an interview and each of these can have their own psychometric properties. Psychometrics refer to the accuracy of inferences being made and a typical interview actually results in a number of inferences. This of course adds considerably to the complexity of an evaluation of the forensic interview with a child who may have been sexually abused. For example, a forensic interview protocol of a child who may have been sexually abused has several components and inferences can be made about each of these—and these inferences can be correct or incorrect.

- (a) Rapport building phase of the interview—can result in the inference, “Adequate report was established”
- (b) Knowledge of the truth—can result in the inference, “Adequate knowledge was displayed”
- (c) Propositional competence phase—can result in the inference, “The child knew key propositions such as ‘inside’ and ‘underneath’”
- (d) “I don’t know” responses are permissible—can result in the inference, “The child knew that it was permissible and important to say that they did not know an answer to a question, when in fact they don’t know”

Each of these inferences can be correct or incorrect. Note there are several other key elements in most protocols (e.g., the special importance of saying the truth in this context, the importance of correcting the interviewer if they say something wrong, etc.). It is important to note that inferences concerning each of these components can have their own psychometrics.

Was the Interviewer Unbiased?

Another way a forensic interview can be evaluated is to ask, “Was the interviewer objective and unbiased?” This question has been too infrequently asked and may be a core reason why some cases have gone so awry—the interviewer was working for a side that had a vested interest and may have been intentionally or unintentionally attempting to please their employer. Ideally, an interviewer ought to have no allegiance to any side—say the prosecution or the defense—they ought to be fully committed to finding out the truth and what the child has to say. However, not all interviewers in all situations may meet this standard and the degree to which this standard of objectivity is met needs to be assessed.

Indicated how a biased interviewer can unintentionally influence children to provide false statements. Subjects were 120 preschool children, 90 of whom attended a birthday party with a visitor. The remaining 30 children did not attend the party but instead spent time coloring with a visitor. Interviewers were graduate students from

social work and counseling programs who knew that the children had participated in an activity with a visitor but not what it was. Each interviewer individually questioned four children to discover what the child had done with the visitor. Unknown to the interviewer, the first three children that he or she interviewed had been at the birthday party but the fourth had not.

This study found that interviewers after questioning the first three children who had attended the party wrongly assumed that the fourth child had also attended the party. The interviewers then (unintentionally) engaged in biased questioning with the fourth child in an apparent attempt to confirm their faulty preconceptions. What is particularly interesting is that in response to these suggestive interviews, 60 % of children who had not actually attended the birthday party made false claims to have been there, and 85 % of interviewers wrongly concluded that all four of the children they questioned had attended the party. Thus, even well-intentioned child interviewers can become biased based on their expectations and background beliefs and then use suggestive techniques to extract false statements from children. A key question is to what extent does this sort of phenomena occur in centers that routinely interview children who may have been sexually abused? What steps can occur to assure that it does not?

Another study by can depict that these sorts of problems can occur even before the forensic interview takes place. It is often the case that other professionals have contact with the child—perhaps to hear their initial outcry before the forensic interview takes place. In this study, two professionals, a teacher and social worker, were given a list of activities that had supposedly occurred during a play session in a group of preschoolers. Unknown to these professionals, half of the activities had not really occurred. These professionals then questioned the children to learn what had happened during the play session. Data from the study indicated that interviewers repeatedly used suggestive questions to ask the children about the bogus activities. In response, these children falsely agreed that they had engaged in about 30 % of these bogus activities, some of which involved bodily touch. Further, some children who initially denied that the bogus event occurred later changed their accounts and provided false details about it. This study shows both accounting for the possible biases of any adult who interviews the child can be critical and again shows the inculcating of false memories by biased professionals.

One other study shows the importance of bias in adults. Had preschool-aged children witness four science demonstrations in a university laboratory. Four months later parents were mailed stories that contained descriptions of their children's visit to the lab. Two of the stories were true and two were false (i.e., described experiments that the children had not seen). Each story finished with a fabricated account of what happened when it was time to leave the lab: "Mr. Science wiped (child's name) hands and face with a wet-wipe. The cloth got close to (child's name) mouth and tasted really yucky." Parents read the story to their children three times. Later, children told the experimenters that they had participated in demonstrations that they had not (i.e., the false stories read by their parents). More than half of the participants said that Mr. Science had wiped their

mouths and many elaborated on their “yes” answers. When asked if Mr. Science had actually wiped their mouths or did their mother just read the story 71 % of the children maintained that it really happened. This study was replicated using children from a wider age range (3–8-year-olds). Findings were similar except they found that when asked if Mr. Science wiped their mouths or if their mother just read the story the older children tended to recant their claims and said that their mother told them.

Thus, ascertaining potential biases of anyone who has questioned the child about their possible abuse seems important. Duke, Uhl, Wood, and Price (2015) recommend that the forensic interview be expanded so that the individual or individuals who heard the child's initial outcry be interviewed to understand if their questions could have been suggestive and thus biasing. In addition, it should be asked, who is the interviewer working for? Who is paying the interviewer?

McMartin Mistakes vs. Suggestive Pathways Before the Forensic Interview

Another incorrect view currently held in the field seems to be along the lines of “if no McMartin type mistakes—repeated questions, conformity pres, suggestive questions are made then the interview is good.” We shall argue that this view is incorrect—it sets too low of a bar. We argue that this is a necessary criterion of adequacy but not a sufficient one. The interview also needs to meet two other criteria.

First, it must attempt to understand and resolve any problems in the key dimensions of the allegation. Suggested that a forensic interview with a child who may have been sexually abused in order to be comprehensive attempt to understand the following dimensions:

1. *Outcry analysis*—the general circumstances of the child's initial accusations should be determined and analyzed for possible bias
2. *Stake analysis*—whether or not anyone who had significant contact with the child has a hidden agenda relevant to a guilty or not guilty verdict toward the accused should be established
3. *Parental/Significant Other suggestion*—whether or not a caregiver or parent has made leading statements or engaged in leading questioning with the child and thus the child has developed a false memory should be evaluated
4. *Forensic Interview analysis*—whether or not biased interviewing techniques were practiced should be evaluated
5. *Memory analysis*—whether or not memory errors (e.g., errors of omission or commission) may have occurred should be determined
6. *Sufficiency of details provided by the child*—whether the child can describe in an age-appropriate manner events that occurred before, during, and after in a way that makes a coherent, understandable, narrative should be assessed

7. *Inconsistencies analysis*—a contradictory statement by logic contains falsehoods, thus whether or not the child has provided inconsistent core details between or within statements should be assessed
8. *Logistical detail analysis*—whether or not the allegation contains logistical implausibilities should be assessed. For example, claims that the child was anally raped but did not experience pain would make the report logistically problematic
9. *Fantastical details analysis*—the presence or absence of fantastical details in the allegations should be examined
10. *Personological analysis*—whether the child suffers from any mental health problems or history that may indicate an increased probability of either truth telling or problematic reports should be assessed

The basic idea is that these dimensions of a sexual abuse allegation are central to understanding what the child is indeed saying or attempting to say. Children due to a variety of factors may not be articulate clearly what happened and the interviewer needs to be mindful and probe key dimensions of a possible sexual abuse allegation so that the interview provides as much clarity and as much detail as possible. Moreover problems with the child's statements need to be identified and disclosed and not ignored or swept under the rug—doing this is not consistent with objectivity. Thus, the degree to which the interview actually addresses these dimensions and attempts to resolve any problems with these, and objectively admits any of these in the conclusions is part of a forensic interview's being comprehensive and objective.

The Protocol for Evaluating Forensic Interviews of Children

Rising general awareness of potentially problematic interviewing practices necessitates the post hoc evaluation of forensic interviews of children for the presences or absence of these practices. However, because these evaluations are yet another form of assessment, they too are bound by the need to establish at least minimal reliability and validity. Unfortunately, very few methods currently exist to identify the presence of problematic interviewing practices and necessary assessments. Furthermore, only one method has been published in peer-reviewed journals with explication of the manner in which it established reliability and validity. That method is the Protocol for Evaluation of Forensic Interviews of Children (PEFIC; Fanetti, O'Donohue, & Bradley, 2006; O'Donohue, Benuto, & Fanetti, 2010).

The PEFIC is an observational behavioral rating system. The PEFIC lists 17 ways, referred to as "biasing factors," that a child's event recall or verbal report may be skewed away from accuracy. Some of these are intra-interview factors—those specifically exhibited by the interviewer in the course of the interview. Others are extra-interview and represent other ways (e.g., including child beliefs) that a child's report or event recall may be biased. For this second set, the interviewer is not

responsible for the presence of the potential problem, but rather is responsible for the *assessment of its presence or absence*. Below is a list of each of the factors specifically identified in the PEFIC.

Intra-interview Factors

1. *Difficulties in establishing rapport*: Does the interviewer maintain sufficient rapport to facilitate a successful interview?
2. *The presence of leading questions*: Does the interviewer, at any point, provide forensically relevant details to the child before the child has iterated said details?
3. *Interviewer disconfirmations of child responses*: Does the interviewer provide either direct or functional disconfirmations of the child's responses?
4. *Inappropriate styles of reinforcement*: Does the interviewer provide a differential form of responding to the child's answers (e.g., accusatory vs. exculpatory)?
5. *Repetitive questions/questioning style*. Does the interviewer repeat questions directly or functionally without providing a justification?
6. *Communication modality inconsistencies*. Does the interviewer fail to observe or note potential disagreements between the child's verbal, paraverbal, and non-verbal responses?
7. *Encouraging the child to speculate*. Does the interviewer encourage or endorse a child's stated speculation while answering questions?
8. *Conformity pressure*. Does the interviewer mention prior or existing reports by the child or other individuals during the context of the interview?
9. *Response class focus*. Did the interviewer redirect the child to a specific individual or class of behavior?

Extra-interview Factors

10. *Outside Contamination*. Did the interviewer assess for the nature of external or outside communication that the child has experienced regarding the allegations?
11. *Understanding of role and purpose*. Did the interviewer assess for the child's understanding of the purpose of the interview and the child's role in the interview?
12. *External threats or bribes*. Did the interviewer assess for possibility that the child's report may have been influenced by threats and or bribes?
13. *Concept of truth*. Did the interviewer assess for the child's understanding of the meaning of truth?
14. *Importance of truth*. Did the interviewer assess for the child's knowledge that the truth is especially important in this forensic context?
15. *"I don't know" responses*. Did the interviewer assess for the child's understanding of how and when to answer questions with a phrase such as, "I don't know"?
16. *Authority pleasing*. Did the interviewer assess for the possibility that the child may be answering questions in a way thought to be pleasing to important individuals?
17. *Discomfort*. Did the interviewer assess for the possibility that the child may not feel sufficiently comfortable or free to discuss forensic details in this setting? This is separate from general rapport.

Those trained to reliably identify these interview problems and assessments are then able to state whether they existed in a specific interview, but not whether they actually influenced the child's responses in that interview. Conclusions are limited to the presence or absence of these factors. If absent, it can be argued that they could not have influenced the child. If present, it can be argued that they may have influenced the child or may not. That influence could simply be not ruled out. Without knowing how the child *would have responded in the absence* of those influences, we cannot know the extent of their effect.

Reports related to this form of interview evaluation should contain a rationale for the way the assessment is done, a complete observational explication of the positive indicators related to each factor (e.g., including quoted interactions from the interview), detailed analyses of why each passage is thought to be an indicator of a problem (or of a sufficient assessment), and conclusions about the breadth and extent of identified problems and strengths. Important in the discussion is the idea that the conclusions to not indicate any level (or lack) of veracity related to the child's report. Decisions of veracity of allegations are primarily the responsibility of the jury and or judge. Evaluations of the forensic interview only serve to highlight the context(s) in which those allegations arose.

Conclusions

Forensic interviews of children are assessment procedures. As such, they are bound by the same principles that guide the development of other psychological assessments. In that way, they will be prone to yield results with some degree of error, even if well-controlled interviews are administered with a high degree of adherence. However, we know that humans that use assessment tools sometimes do so incorrectly. Thus, it is our view that forensic science should focus more closely on the psychometric properties of the assessments utilized, so that the errors rates and adherence problems become known quantities, rather than remaining unknown—and thus a source of debate or contention in applied settings. Without psychometric evaluation, descriptions of the utility, adherence, and quality of interviews are often not much more than an appeal to authority. Authority is not a more convincing replacement for scientific psychometric analysis.

Even without the daunting task of measuring adherence to forensic child interview protocols (which are designed to assess events of nearly unlimited variance), the field has reached some agreement on basic problems that occur in interviews that may harm a child's accuracy and likely do not help it. We suggest as a minimal first step that we ensure an ability to identify when these things have happened in an interview. After all, if we cannot agree that certain problems have or have not occurred, how are we then to make the argument that they are or are not a problem in the first place? We think that methods for such observational agreement have existed in the literature for decades, but only as hidden components of other studies—as functional “procedures” used in studies of memory and suggestibility. We argue that these procedures

have merit in their own space and should be developed and psychometrically evaluated as stand-alone assessment methods. This is the goal and application of the PEFIC mentioned above.

References

- Cirlugea, O., & O'Donohue, W. (2015). Review of psychometrics of forensic interview protocols with children. In W. O'Donohue & M. Fanetti (Eds.), *Forensic interviews regarding child sexual abuse: A guide to evidence-based practice*. New York, NY: Springer.
- Duke, M. C., Uhl, E. R., Wood, J. M., & Price, H. L. (2015). Avoiding problems in child abuse interviews and investigations. In W. O'Donohue & M. Fanetti (Eds.), *Forensic interviews regarding child sexual abuse: A guide to evidence-based practice*. New York, NY: Springer.
- Fanetti, M., O'Donohue, W., & Bradley, A. (2006). A method for evaluating child forensic interviews. *American Journal of Forensic Psychology*, *24*(5), 5–27.
- Haynes, S., Smith, G. T., & Hunsely, J. (2011). *Scientific foundations of assessment*. New York, NY: Academic Books.
- O'Donohue, W., Benuto, L., & Fanetti, M. (2010). Children's allegations of sexual abuse: A model for forensic assessment. *Psychological Injury and Law*, *3*, 148–154.
- Rabinowitz, D. (2004). *No crueler tyrannies*. New York, NY: Free Press.