

Chapter 13

Review of Psychometrics of Forensic Interview Protocols with Children

Olga Cirlugea and William T. O'Donohue

Review of Psychometrics of Forensic Interview Protocols with Children

There is a misperception that if a child has been sexually abused there usually will be medical evidence corroborating the abuse (Frasier & Makoroff, 2006). If that were the case, there would be less need to subject children to an interview; however, research indicates that only about 4 % of all child sexual abuse (CSA) investigations produce medical evidence such as genital anomalies, bruising and cuts supporting the occurrence of the abuse (e.g., Berenson et al., 2000; Heger, Ticson, Velasquez, & Bernier, 2002). Even with physical evidence key questions remain: Who was the perpetrator? How many times did the abuse occur? In what jurisdiction did the abuse occur? When did the abuse occur? Did any adult know of this abuse and failed to stop it? This key information can only be gathered through an interview with the child. The forensic interview also can provide a safe and supportive environment for disclosure to occur and can facilitate disclosure. Children sometimes do not disclose abuse, at least not immediately. For example, Malloy and colleagues (Malloy, Brubacher, & Lamb, 2011) found that 20 % disclosures occurred within 1 month of the alleged abuse, and an additional 57 % occurred up to several years after the event.

O. Cirlugea, B.A. (✉)
Department of Psychology, University of Nevada, Reno, MS 0296, Reno, NV 89557, USA
e-mail: olgacirlugea@gmail.com

W.T. O'Donohue, Ph.D.
Department of Psychology, Victims of Crime Treatment Center, University of Nevada, Reno,
MS 0296, Reno, NV 89557, USA
e-mail: wto@unr.edu

Egregious Examples of Problems in Interviewing

Even though the American Educational Research Association encourages that “those who select tests and interpret test results should refrain from introducing biases that accommodate individuals or groups with a vested interest in decisions affected by the test interpretation,” (1999, p. 131) CSA interviewers can bring personal biases into the interview and may even have their own agenda for the interview. The purpose of the interview should always be to elicit accurate and complete information (whatever this may be), but interviewers can have affiliations that may lead to biases (e.g., ultimately be employed by prosecutors), or have biases that the child was abused or not abused before the interview has even begun (Ceci & Bruck, 1995). In fact, history is replete with high-profile trials involving very poorly conducted CSA interviews that focused on only one hypothesis—that the child was sexually abused—and that had severe negative consequences for all concerned (e.g., millions of dollars spent, innocent people serving many years in prison, etc.; Rabinowitz, 2003). The infamous McMartin trial, which lasted from 1987 until 1990, is probably the most notorious and was one of the first to expose widespread concerns regarding suggestive techniques used in forensic interviews with children (Ceci & Bruck, 1995). Seven teachers at a Manhattan Beach, California preschool were charged with kidnapping and sexually abusing hundreds of children. Extensive interviews conducted by child advocate Kee MacFarlane led to allegations of satanic rituals, for example, children being forced to drink blood, watching babies being beheaded, flights over the Pacific Ocean where babies were fed to sharks, and thousands of counts of sexual abuse including group sex and sodomy. However, reviews of videotapes of the interviews indicated that MacFarlane had relied heavily on suggestive interview techniques that elicited allegations of sexual abuse (Schreiber et al., 2006). The recognition of this problematic interviewing eventually led to the seven teachers being cleared of all charges, however, not before some had spent years in prison and had lost their homes, their families, and their reputations (Rabinowitz, 2003).

In the early 1990s, Edenton, North Carolina experienced a similar trial involving several preschool workers at the Little Rascals Daycare (Ceci & Bruck, 1995). The owner, Bob Kelly, his wife, and 5 other caregivers were accused of raping and sodomizing 29 preschool children. Initially no child made any allegations of sexual abuse or satanic rituals taking place at the daycare. However, after months of repeated interviews during therapy sessions, persistent questioning at home and attending “court school” in preparation for testimony in court, the children began disclosing details of satanic rituals during which children were allegedly vaginally and anally penetrated with various objects (e.g., pins and markers), thrown into pools of sharks, and beaten. The children’s statements also included fantastical stories of being flown in spaceships and hot air balloons. Unlike the case of the McMartin trial, no videotapes of the interviews were available for review—which in itself is quite problematic—as the therapists conducting the interviews had lost or destroyed them (Anderson, 2007). However, an important part of the problem

seemed to be that interviewers and other officials associated with these cases appeared to be more concerned about false negatives (e.g., acquittal of guilty perpetrators) and showed little or no concerns about false positives (e.g., conviction of innocent defendants; Ceci & Bruck, 1995).

On the other hand, children may fail to be interviewed and their abuse can remain undetected and this can set the stage for the perpetrator to continue to abuse them or to abuse others. For example, in the infamous recent Penn State case, former coach Jerry Sandusky was convicted of 45 counts of abuse he had perpetrated from 1994 to 2009. The investigation was prompted by the first victim's disclosure in 2008 and many of his victims did not come forward until the trial.

The Heterogeneity of Interview Protocols

Poor forensic interviewing techniques like those utilized in the McMartin and Edenton cases created a need for successful interviewing protocols that minimize suggestive questioning, as well as avoiding other mistakes in order to maximize the accuracy of information elicited and subsequently a number of such interview protocols have been attempted. While the American Education Research Association (1999, p. 43) requires that all "tests and testing programs should be developed on a sound scientific basis," additional controversy is created because some forensic protocols still use less than supported techniques (e.g., sexually anatomically correct dolls and anatomical drawings; Elliott, O'Donohue, & Nickerson, 1993). Nevertheless, because these protocols are still widely used in the United States, it is necessary to critically review them based on a set of criteria proposed in a later section.

A thorough search of the literature has identified the following three protocols as the most influential forensic protocols for CSA interviewing:

1. National Institute of Child Health and Human Development (NICHD) Investigative Interview Protocol (Orbach et al., 2000; Lamb, Orbach, Hershkowitz, Esplin, & Horowitz, 2007)
2. RATAC Forensic Protocol (CornerHouse, 1990; 2003; 2007)
3. Step-wise interview (Yuille, Hunter, Joffe, & Zaparniuk, 1993)

The rate at which these protocols, combinations, variants, or wholly idiosyncratic interviews are actually used is currently unknown, although the NICHD Investigative Interviewing Protocol, or some variant of it, seems to be the most often used. The fidelity of the interviewers in the field adhering to these protocols is also unknown. Additionally, because there are a wide variety of professionals interviewing children, in a wide variety of jurisdictions, with varied backgrounds (e.g., police, social workers, psychologists, interview specialists, etc), varied levels of experience in forensic interviewing and levels of training, variability in the content of the interviews will inevitably be produced.

Evaluation Criteria for Forensic Interviews with Children

In order to assess the quality of interview protocols, one needs a reasonable set of evaluative criteria. A set of criteria for evaluating CSA protocols is proposed below. Some of these criteria have been drawn from the Standards for Educational and Psychological Testing developed by the American Educational Research Association (1999) while others have been taken from the extent literature on CSA.

1. *Interrater Reliability*. Interrater reliability indicates the “degree of agreement between scores or ratings obtained from different sources (observers, instruments, and clinicians)” (Haynes, Smith, & Hunsley, 2011). Every interview protocol should be tested for interrater reliability prior to its use outside of a research setting. It is important to know that the interview results would not have varied significantly if another interviewer were to have conducted the interview. Results should indicate that the protocol has high interrater reliability to ensure that two or more raters are able to agree on the inferences made based on the child’s statements, for example that the child was sexually abused. Because there are multiple inferences made in an interview (e.g., whether the child was abused; what the abuse consisted of; where it occurred, etc.) there is actually a series of interrater reliabilities to be examined in an interview.
2. *Component Construct Validity*. Construct validity is the “degree of validity of inferences about unobserved variables (constructs) based on observed indicators” (Haynes et al., 2011). In the context of a forensic interview, one inference made is about whether adequate rapport has been established with the child. Assessing this construct often relies on multiple indicators regarding the child’s behavior such as the child’s general affect, his or her willingness to have a parent leave the room, his or her willingness to discuss details surrounding the abuse, as well as other indicators. However, eventually the interviewer comes to some sort of general conclusion that “sufficient” rapport was or was not attained and the accuracy of this inference must be determined. Forensic protocols are comprised of a number of distinct components constructs (e.g., rapport building; understanding of the meaning of telling the truth versus a lie; lack of threats or bribes; propositional competence) and psychometrically there is an interest in the degree of validity of each inference that is made about each of these constructs.
3. *Predictive (Postdictive) Validity*. Once the construct validity of the individual components of a protocol is established, the inferences made based on the integration of information gathered from a forensic interview become relevant. It is important to determine the accuracy of inferences involved in conclusions that may be drawn from the interview such as, “The interview suggests that this child was anally penetrated on four separate occasions in Sacramento California by her Uncle Joe between March 2011 and August 2011 and no one knew of this abuse, and no other acts or actors were involved.” Because these events are in the past, they fall under the psychometric term of “postdictive validity.” Postdictive validity may be defined as the accuracy of inferences made about historical

events and there are a number of inferences about the past that the forensic interviewer seeks to make, including:

- (a) Abuse Status (that the child was or wasn't sexually abused)
- (b) Who the child has identified as the alleged perpetrator
- (c) What type of sexual abuse it was (contact or noncontact, and whether penetration took place is particularly important)
- (d) Where and when the abuse took place
- (e) How many times the abuse was perpetrated
- (f) If anyone else knew of the abuse and was complicit in it

As mentioned in an earlier section, these details play a large role in the charging and sentencing of the perpetrator and could mean the difference between a lighter and a more severe judgment.

4. *Incremental Validity*. Incremental validity is defined as the degree to which data from one or more measures “increase validity or utility of a judgment beyond what can be accomplished with other sources of data” (Haynes et al., 2011). When conducting forensic interviews it is important, given previous pieces of information gathered during the CSA investigation from the medical examination, interrogation of the alleged perpetrator, collateral contacts, etc., to what extent information elicited by the interviews adds to the judgment facilitated by those data. Interviews are time-consuming and costly, and it is important that these costs be justified.
5. *Sensitivity/Specificity*. Sensitivity is described as the “proportion of positive cases so identified on the basis of a measure from a particular assessment instrument” while specificity refers to “the proportion of negative cases so identified by an assessment instrument” (Haynes et al., 2011). The reason interview protocols were developed in the first place was to increase the likelihood that the child will provide the most accurate and detailed narrative possible in order to most precisely determine whether sexual abuse did in fact occur or not, while at the same time decreasing the likelihood that any personal biases will enter the professional judgment of the interviewer. Therefore, it is extremely important that research reveals that an interview protocol has adequate sensitivity and specificity and is thus able to distinguish between children who have been abused and those who have not. Not doing so can have serious consequences for persons involved in the allegation. In the case of false positives, the alleged perpetrators may be falsely accused of sexually abusing a child and may end up serving time, have to register as a sex offender, pay monetary compensation, etc., because of a crime he did not commit. When a protocol results in false negatives the perpetrator is not correctly identified and brought to justice for his crimes and this increases the likelihood that he will have the opportunity to reoffend and hurt more children before finally being caught.
6. *Developmental Appropriateness*. In attending to a child's age during forensic interviews, we are in fact interested in the child's cognitive development. Two-year-olds have different cognitive capacity from 12-year-olds, and this difference

must be taken into consideration when asking questions and when evaluating the child's statement. For example, investigators interviewing younger children must use simpler words and shorter sentences. Research indicates that younger children provide fewer details when free narratives are elicited (e.g., Ceci & Bruck, 1993). Additionally, they are also more susceptible to suggestibility and the formation of false memories than older children and adults, although research has shown that even adults are capable of falsely accepting events that never happened as true (Loftus & Pickrell, 1995). Having certain disabilities (e.g., developmental disabilities) may also complicate forensic interviews. An autistic child, for example, may have reduced cognitive ability, problems verbalizing, attention issues, etc., and all these would affect the accuracy and completeness of the child's report and the inferences made based on information elicited during the interview. If modifications are made to accommodate disabilities in the protocol, "the validity of inferences made from test scores and the reliability of scores on tests administered to individuals with various disabilities should be investigated and reported by the agency or publisher that makes the modification" (American Educational Research Association, 1999, p. 107).

7. *Cultural Sensitivity*. The APA (2003) stresses that due to a growing population that is increasingly multicultural, psychologists should demonstrate cultural competence in their practice. A number of issues, ranging from language barriers to different attitudes toward authorities, could emerge when conducting forensic interviews with such populations. In fact, American Educational Research Association (1999) affirms that "testing practice should be designed to reduce threats to the reliability and validity of test score inferences that may arise from language differences" (p. 97). If a forensic protocol has been translated into other languages, it is important to outline "the methods used in establishing the adequacy of the translation," and "empirical and logical evidence should be provided for score reliability and the validity of the translated test's score inferences for the uses intended in their linguistic groups to be tested" (American Educational Research Association 1999, p. 99). Cultural differences may pose additional barriers when interviewing a child. Talking about sexual abuse is difficult for any child, but children from certain cultures may be less likely to disclose abuse to an interviewer because such events are usually kept "in the family" and are not discussed with authorities (Fontes & Plummer, 2010). It is also useful to know if a protocol has evaluated with other populations (e.g., people with disabilities, people from a different culture) and if the studies indicate that the protocol is appropriate for use with those populations.
8. *Trainable Successfully (Implementation Fidelity)*. The American Educational Research Association (1999) advises that "those who use psychological tests should confine their testing and related assessment activities to their areas of competence, as demonstrated through education, supervised training, experience, and appropriate credentialing" (p. 131). Any successful protocol must include a successful training that ensures the desired level of competence. Because a variety of professionals, from psychologists to law enforcement personnel, are trained to conduct forensic interviews, protocol developers must keep

the users and their differences in mind. One cannot assume that a psychologist with a background in child development will have the same knowledge about memory, suggestibility, behavioral principles, etc., as a police officer who may have never taken relevant courses and read relevant research. Therefore, it is probably more prudent to err on the side of caution and provide sufficient background knowledge to any training course in CSA interviewing. Finally, drift and supervision issues must be also addressed. Research shows that even though professionals spend valuable time and money getting trained in CSA interviewing, over time some just fail to adhere to the protocol (Lamb, Sternberg, Orbach, Esplin, & Mitchell, 2002; Lamb, Sternberg, Orbach, Hershkowitz et al., 2002). Drift and nonadherence to the protocol may demonstrate a need for continued supervision and help with difficult cases. Therefore, it is imperative that field studies be conducted to evaluate the effectiveness of training by assessing fidelity of protocol adherence.

Description and Evaluation of Major Protocols

National Institute of Child Health and Human Development Investigative Interview Protocol

The NICHD Investigative Interview Protocol (Orbach et al., 2000; Lamb et al., 2007) is the best-researched and most widely used forensic protocol for CSA interviewing. This structured protocol is divided into two stages, the presubstantive and substantive portions of the interview. The interviewer introduces him/herself, discusses the child's duty during the interview (i.e., tell the truth), and covers the rules and expectations (e.g., use of "I don't know" responses) during the introductory phase. During the rapport phase, the interviewer seeks to build rapport with the child in a comfortable environment. The narrative training phase helps the child get accustomed to responding to open-ended questions about a neutral event. A transitional phase occurs between the presubstantive and the substantive phase of the interview in which the interviewer orients the child to the target event/s under investigation through the use of prompts. If the transitional phase elicits a disclosure, the interviewer moves on to the free-recall phase and, once the interviewer has gathered as many possible details through free-recall prompting, the transition is made to directive questioning about information previously provided by the child. At this time, the child may take a break. After the break, the interviewer continues to ask direct questions about the disclosure. When the required information has been elicited, the interviewer may go on to the closing phase, and a neutral topic (for example, asking the child about his/her plans for the day) may also be discussed with the child.

Two memory enhancing techniques, Physical Context Reinstatement (child is interviewed at the scene of the alleged crime) and Mental Context Reinstatement (guided mental reconstruction of the setting of the alleged crime) have both been

used in conjunction with the NICHD protocol (Orbach et al., 2000; Hershkowitz et al., 2001; Hershkowitz, 2002). Both of these techniques appear to have elicited additional details from the children. Studies conducted in Israel, United States, United Kingdom, and Canada all have demonstrated that interviewers using the NICHD Investigative Interview Protocol as opposed to those using non-protocol methods used more open-ended and free-recall prompts, and used fewer focused, directive, and option-posing questions (Orbach et al., 2000; Sternberg et al., 2001; Lamb et al., 2006; Cyr & Lamb, 2009). However, results relevant to amount of information provided by the children in response to these questions revealed no differences between conditions in the number of informative information given by the child, although children in the protocol condition did provide most of their information in response to open-ended and free-recall prompts (Lamb et al., 2009).

Interrater Reliability. Hershkowitz et al. (2007) evaluated the interrater reliability of the judgments of 42 Israeli youth investigators. Twenty-four forensic interviews were selected, of which half were classified as plausible and half as implausible based on the Horowitz et al. (1995) "ground truth" scale that utilized independent evidence to corroborate allegations made during an interview. Half of the interviews used the NICHD Investigative Interview Protocol, while the other half did not follow a protocol (non-protocol condition). In order to elicit an interrater reliability coefficient, "seven child investigators independently judged the credibility of each of the transcribed interviews" using a 4-point scale to indicate how likely it was that each alleged incident had really taken place (p 103). Results indicated that there was a difference between the interrater reliability of investigators rating non-protocol interviews ($a=.764$) and the interrater reliability of those rating the NICHD Investigative Interview Protocol interviews ($a=.874$). Additionally, a significant difference emerged when rating cases involving implausible allegations ($a=.338$ versus $a=.642$ for non-protocol and NICHD Investigative Interview Protocol interviews respectively).

Component Construct Validity. While the protocol was developed by experts in the field of child interviewing, there is no evidence that it has undergone subsequent content validation. Several studies reveal that interviewers using the NICHD Investigative Interview Protocol were more likely to engage in the recommended techniques (e.g., to explain the ground rules and utilize rapport building techniques) than those using a non-protocol interview (Sternberg et al., 2001). Additionally, the use of the protocol increased the number of open-ended utterances posed by the interviewers (Orbach et al., 2000; Sternberg et al., 2001; Lamb et al., 2006; Cyr & Lamb, 2009).

Postdictive Validity. There is no research available evaluating the accuracy of inference made about the alleged CSA and details surrounding it.

Incremental Validity. While no studies specifically examined the incremental validity of the NICHD protocol, Darvish et al. (2005, as described in Lamb et al., 2008) evaluated the amount of investigative leads provided by NICHD Investigative Interview Protocol interviews versus non-protocol interviews.

Investigative leads were categorized as information about the suspect, witnesses, medical leads, material leads, and “miscellaneous” and as “very strong” to “very weak” on a 6-point scale. Details elicited were classified as either central or peripheral, and the verifiability of entire statement of the child was rated from “very low” to “very high” on a 4-point scale. Results indicated that the NICHD Investigative Interview Protocol interviews yielded significantly more leads categorized as “very strong,” and statements that were more highly verifiable than the non-protocol interviews.

Sensitivity/Specificity. The Hershkowitz and colleagues (2007) study described in the Interrater Reliability section above examined the accuracy of judgments made by investigators in addition to the reliability of their judgments. Results revealed that 59.5 % of the judgments of the NICHD Investigative Interview Protocol interviews were accurate (95.2 % of judgments about plausible statements and 23.8 % of judgments about implausible statements), while only 29.6 % of the judgments of non-protocol interviews were accurate (38.1 % of judgments about plausible statements and 11.9 % of judgments about implausible statements). These findings indicated that, while the NICHD Investigative Interview Protocol interviews had better outcomes when interviewers rated plausible statements, interviewers rating statements elicited by both NICHD Investigative Interview Protocol and non-protocol interviews failed to accurately rate those when the judgments were made about statements that were implausible.

Developmental Appropriateness. Multiple studies have been conducted examining the ability of interviewers using the NICHD Investigative Interview Protocol to elicit accurate and detailed information from children of different ages. The typical study compared the effects of NICHD Investigative Interview Protocol interviews to non-protocol interviews on interviewer utterances (invitations, directive, option-posing, and suggestive) and on amount and accuracy of details given by the children. Some of the studies (e.g., Sternberg et al., 2001; Hershkowitz, 2001) have failed to identify any differences among age groups. However, Orbach et al. (2000) found that older children gave more details than younger children in both the NICHD Investigative Interview Protocol and non-protocol conditions. Additionally, Lamb et al. (2003) found that 8-year-old children provided a greater amount of details than 4-year-old children, although there were no differences in the amount of information elicited by each type of utterance. Alridge et al. (2004) noted that when Human Figure Drawings were added to the protocol, younger children (ages 4–7) provided 27 % more details after having allegedly exhausted their memories, versus 19 % for 8–10-year-olds and 12 % for 11–13-year-olds. The authors caution that these additional details may have come at the expense of less accurate information. When Mental Context Reinstatement was added to the protocol (Hershkowitz et al., 2001), all children provided proportionally more details in response to invitations than to other prompts, with children ages 4–6 reporting more free-recall information (41 %) than children ages 7–9 (15 %) and 10–13 (17 %), although the overall number of details did not increase.

Additionally, the protocol has been evaluated in children with developmental disabilities. Dion and Cyr (2008) examined 34 forensic interviews of children with low verbal abilities (LVA) as indicated by low scores on the Vocabulary subtest of the WISC III. Half of the interviews were conducted with the NICHD Investigative Interview Protocol and half without a protocol. Findings indicate that interviewers using the protocol provided significantly more invitations and significantly less suggestive utterances than those not using the protocol, and there was no significant difference in amount of directive and option-posing utterances. Furthermore, when compared to children of average verbal ability (AVA), children with LVA interviewed with the NICHD Investigative Interview Protocol gave more details than children with AVA interviewed without the use of a protocol. When both sets of children were interviewed using the protocol, children with AVA provided more details than their LVA counterparts. Brown et al. (2012) assessed the ability of intellectually disabled children (mild-IQ below 80, and moderate-IQ 40-55) to provide reliable accounts of an experienced event. The children witnessed a classroom event and were subsequently interviewed in a supportive manner 1 week or 6 months after the event using the NICHD Investigative Interview Protocol. Suggestive questions were added at the end of each interview. Results revealed that the mildly intellectually disabled children were able to provide highly accurate information about the experienced event, particularly to open-ended prompts. However, moderately intellectually disabled children required more specific prompting and more focused questions, and had poorer performance overall. All children provided more inaccurate information in response to the suggestive questions.

Cultural Sensitivity. The protocol has been tested in four countries, Israel, United States, United Kingdom, and Canada. Cyr and Lamb (2009) found that Canadian interviewers using the NICHD Investigative Interview Protocol utilized significantly more open-ended prompts and significantly less suggestive and option-posing questions than interviewers conducting a non-protocol interview with French-speaking children. Additionally, the children provided more details per prompt when the NICHD Investigative Interview Protocol was used, and these results were replicated by Lamb et al. (2009) in a British sample.

Trainable Successfully. Several studies have been conducted on the effects of training on the quality of forensic interviews. Examined 192 interviews conducted by 21 Israeli youth investigators. The authors tested the following four conditions: validation; rapport building; "victims" protocol in which the interviewers were trained in the NICHD protocol; and "suspects" protocol condition. The validation and rapport building trainings consisted of brief workshops while the "victims" and "suspects" protocol conditions consisted more intensive training followed by continued supervision and case reviews in the "victims" protocol condition. Interviews conducted in one of the four conditions were compared to interviews in baseline conditions (that is, interviews previously conducted by the same interviewers). Results demonstrated that interviewers in the "victims" protocol condition performed significantly better as evidenced by using more open-ended prompts and fewer focused prompts. This indicates that the more intensive training and subsequent supervision increased the quality of the forensic interviews. Lamb, Sternberg, Orbach, Esplin et al. (2002) and

Lamb, Sternberg, Orbach, Hershkowitz et al. (2002) conducted a similar study in which the interviews conducted by eight experienced forensic investigators while they were receiving ongoing supervision were compared to the interviews conducted by the same group of investigators after supervision had ended. Results indicated that the termination of supervision had an adverse effect on the interviewers' behavior, as interviewers used significantly fewer invitations and more option-posing and suggestive prompts after supervision had ended. In light of these findings, the authors suggested that continued supervision may be required to ensure that investigators maintain a high quality of forensic interviews.

RATAC Forensic Protocol

The RATAC forensic protocol (CornerHouse, 1990, 2003, 2007, described in Anderson et al., 2007) is a semi-structured interview protocol comprised of five stages: Rapport; Anatomy Identification; Touch Inquiry; Abuse Scenario; and Closure. The first stage, Rapport, seeks to establish the child's comfort, communication, and competence. The second stage, Anatomy Identification, utilizes anatomical drawings for a number of different purposes depending on the child's age. The drawings are used with young children to assess whether they can identify their own gender as well as to capture the child's idiosyncratic language for different body parts. The protocol also allows the use of drawings as memory cues. Stage three, Touch Inquiry, assesses the child's understanding of good touches and unwanted touches. Children are asked to define a touch, "identify who gives the touch, and to indicate" what body part has been touched (Anderson et al., 2007, p. 297). If the child has made a disclosure, the interviewer proceeds to the Abuse Scenario phase in which information is gathered about the child's experience including who the perpetrator was and how many times the abuse took place. During this phase, the use of interview aids such as drawings, anatomical drawings, and anatomical dolls is allowed, the latter that are introduced after disclosure has occurred in order to clarify details or get a visual demonstration of the child's experience. The protocol recommends that interviewers take into account the child's developmental level when employing such aids. The last stage of the protocol, Closure, is a time for the child to share any other information he/she may have about the alleged abuse; to validate the child's emotions surrounding the disclosure; to address any questions the child may have about the interview; and to thank the child for his/her participation in the interview. This stage also incorporates education about personal safety, about reporting future experiences, and exploration of safety options should abuse occur in the future. Interviewers may modify or eliminate any one of these stages to better address the child's developmental level.

Interrater Reliability. No research has examined the interrater reliability of the RATAC forensic protocol.

Component Construct Validity. RATAC components include Rapport, Anatomy Identification, Touch Inquiry, Abuse Scenario, and Closure. However, none of these

components have been evaluated to ensure that they have each been adequately addressed during the forensic interview and they have not been validated for content by experts in the field. Additionally, some of the stages utilize questionable techniques (for example, the multitude of interviewing aids) that have not been validated for use with potential victims of CSA.

Postdictive Validity. No studies are available examining the predictive validity of the RATAc forensic protocol.

Incremental Validity. There are no studies assessing the incremental validity of the RATAc forensic protocol.

Sensitivity/Specificity. No research has been conducted on the sensitivity and specificity of the RATAc forensic protocol.

Developmental Appropriateness. There are no studies examining the use of the protocol with children of different ages. However, the protocol aims to take a developmentally appropriate approach to interviewing children that takes into consideration differences in children's memory functions, attention span, comprehension, simple versus complex language, and concrete versus abstract concepts. Additionally, it provides general guidelines for age-appropriate questions (e.g., using only "who" and "what" questions with 3-year-olds, adding "where" questions with 4-year-olds, and omitting the use of "why" questions with all children). The protocol also discusses question type (e.g., open-ended, focused, etc.) in the context of child development and recommends that more direct questions be used with younger children.

Cultural Sensitivity. No studies have been conducted examining the validity of the protocol with individuals from different cultures. However, the protocol does indicate that culture plays a role in how children disclose given the cultural differences in narrative models (e.g., children from Western cultures may discuss their feelings, thoughts, and preferences more than those from Eastern cultures).

Trainable Successfully. There is no empirical evidence that the protocol can be trained successfully and that interviewers who have undergone the RATAc training conduct superior interviews to those who have not. Nevertheless, the protocol cites case after case in which expert testimony has been admitted in court because the expert witness was trained in this protocol (Anderson et al., 2007). Additionally, Vieth (2009) notes that interviewers trained in the RATAc forensic protocol receive continued supervision, technical assistance, etc., although none of these claims have been evaluated.

Step-Wise Interview

The Step-wise interview (Yuille et al., 1993) was developed in order to attain the following goals: minimize trauma experienced by the child during the interview; maximize the information provided by the child about the alleged abuse; minimize

contamination of the child's information; and "maintain the integrity of the investigative process" (Yuille et al., 1993, p. 100). This protocol proceeds in nine phases: rapport building, requesting recall of two specific events, telling the truth, introducing the topic of concern, free narrative, general questions, specific questions (if necessary), interview aids (if necessary), and concluding the interview. The interview begins with a rapport building phase in which the investigator discusses neutral topics with the child in order to develop rapport. During this phase, the child is asked to describe two past experiences, the goal being to assess how much detail the child can be expected to provide as well as to model the form of the interview for the child. The next phase assesses the child's ability to define truth and lies, to identify whether specific statements are truth or lies, and to determine the child's understanding of the consequences for lying. Next, the topic of concern is introduced in a step-wise manner. Open-ended questions are first used to elicit a disclosure, then more specific prompts are utilized such as "Has anyone done something to you" and "Has anything happened to you which you would like to tell me about?" However, the authors advise against using the name of the alleged perpetrator or suggesting what happened during the alleged abuse. Drawings of both genders may also be used to determine if the child can name and describe the functions of all body parts from head to toe, and to assess if the child has seen any of the private parts (genitals and anus) on another person or if anyone has touched those parts on the child. After the child is oriented to the topic of concern, prompts such as "tell me what happened" are used to elicit a free narrative from the child. General questions based on the information provided by the child can be used to elicit additional details about the event. The authors advise against using leading or suggestive questions. The specific questions phase should only be covered if the free-narrative and open-questions phase have not extracted sufficient details and there is a need for further clarification or extension of the child's answers. This is also a time for resolving any inconsistencies in the child's statement. The authors suggest that interview aids may be used with young children or children with language or emotional difficulties. Although they allow the use of anatomical dolls, they do so with a cautionary statement that they only be used after the child has made a disclosure in order to clarify what sexual act has taken place. In the case that the child appears to acquiesce to suggestion, the authors also recommend asking a few leading questions not related to the event to determine the child's suggestibility. The final phase is the conclusion of the interview. The child's questions are answered, and he/she is thanked for his/her participation. The protocol strongly advises against making any promises to the child, for example, that the abuse will not happen again.

This protocol was developed in conjunction with the Statement Validity Analysis (SVA; Raskin & Yuille, 1989), a technique for the assessment of the credibility of children's statements. SVA is made up of two sections, the criteria for content-based criterion analysis (CBCA) that assumes that certain elements are present in a true disclosure and a validity checklist. The CBCA assesses the following 19 elements of a child's statement: coherence, spontaneous reproduction, sufficient detail, contextual embedding, description of interactions, reproduction of conversation, unexpected complications during the interview, unusual details, peripheral details, accurately reported details not understood, related external associations, accounts

of subjective mental state, attribution of perpetrator's mental state, spontaneous corrections, admitting lack of memory, raising doubts about one's testimony, self-deprecation, pardoning the perpetrator, and reports of other's action. In addition, the validity checklist addresses the following factors: statement-related factors, psychological characteristics, appropriateness of language and knowledge, presence of affect, spontaneous gestures, susceptibility to suggestion, interview characteristics, and adequacy of the interview. There is little evidence for the validity of SVA in evaluating the veracity of children's statements; however, because it is meant to be used simultaneously with the Step-wise interview, we will be at times referring to it when evaluating the protocol for the proposed criteria.

Interrater Reliability. No research has been conducted on interrater reliability of the Step-wise interview.

Component Construct Validity. Step-wise interview phases include rapport building, requesting recall of two specific events, telling the truth, introducing the topic of concern, free narrative, general questions, specific questions (if necessary), interview aids (if necessary), and concluding the interview. These components have not been validated for content by experts in the field. Additionally, there is no research examining whether these phases are appropriately addressed by interviewers trained in this protocol.

Postdictive Validity. Zaparniuk and colleagues (1995) evaluated the ability of trained coders to accurately identify statements elicited from interviews guided by the Step-wise protocol as true or false utilizing the CBCA portion of the SVA. Coders followed set decision rules that would help them differentiate true from false statements, for example, having criteria 1 to 5 present, as well as any other 2 criteria from the CBCA. Results indicated that the coders only performed slightly better than chance at distinguishing true from false statements, demonstrating the difficulties in making accurate inferences about historical events.

Incremental Validity. There are no studies evaluating the incremental validity of the protocol.

Sensitivity/Specificity. No research has been conducted examining the sensitivity and specificity of the Step-wise interview.

Developmental Appropriateness. The Step-wise interview has a few factors build in that directly address developmental appropriateness. The phase in which the child is asked to describe two neutral events was developed to obtain a baseline of the child's memory and language skills which can then be compared to the details provided during the disclosure of the sexual abuse. There is also an optional phase in which the interviewer may test the child's prepositional understanding. A set of interview rules are also provided, but are not recommended for use with preschool-aged children. Several studies analyzed the developmental appropriateness of the protocol. Hardy and Van Leeuwen (2004) examined four variations of the Step-wise interview with children ages 3–8. The children watched performances of *The Beast with a Thousand Teeth* given by undergraduate students in their classrooms and preschools. The children were subsequently interviewed in one of four interview

conditions: “a. direct probes with past event talk; b. direct probes with general event rapport talk; c. indirect probes with past even talk; and d. indirect probes with general event talk” (p. 159). Some children were also given four suggestive and ambiguous probes to test their ability to resist suggestion. Results indicated that older children provided more information than the younger children. These results were significant in the indirect probes conditions. Additionally, younger children provided fewer accurate details when questioned about specific past events. No age differences were found in children’s ability to resist suggestive probes except for in the condition using indirect probes, in which older children were less suggestible. Porter, Yuille, and Bent (1995) compared the eyewitness accounts of deaf and hearing children using a procedure based on the Step-wise interview. The children were shown a set of slides that depicted a story in which a man wearing a cowboy hat stole a woman’s wallet after bumping into her. The participants were subsequently interviewed using free recall and direct questions, and accuracy scores were collected. Results revealed no significant difference between amount of detail recalled by deaf and hearing children. Additionally, both deaf and hearing children recalled details with similar accuracies during the free-recall phase. However, when direct questions were used, the details of hearing children were significantly more accurate than those of deaf children.

Cultural Sensitivity. No studies have been conducted that examine the validity of the protocol with different ethnic groups. Additionally, the Step-wise interview has not been translated in any other languages.

Trainable Successfully. Yuille et al. (1983) conducted a field study examining three aspects of the training: the trainee’s satisfaction with the training at the end of the 4-day workshop; a follow-up session 6 months after the training in which trainees rated how often they used the protocol; and “ratings of the quality of the taped interviews of trained and untrained workers” (p. 111). Child Protective Services workers, law enforcement personnel, and prosecutors from two districts attended a 4-day workshop on the Step-wise interview. Professionals from a third district served as the control group and did not receive training in the protocol. Results revealed that participants reported they had a positive view of the training and adequate information was provided. At the 6-month follow up, most participants indicated that they used the protocol “sometimes to always” when conducting CSA interviews. When the control and experimental groups were compared in regard to adequacy of interviews, 30 % of the interviews in the control condition were deemed inadequate due to scant or contaminated information versus 5 % in the experimental groups, illustrating problems in training and implementation of forensic protocols. Additionally, the manuscript did not mention whether the raters were blind, posing additional problems regarding the interpretation of the results.

Conclusions

There are several major protocols for forensic interviews of children who may have been sexually abused. Although these protocols share some key similarities (e.g., the importance of rapport building), they also demonstrate significant divergences. We have proposed criteria of adequacy for the content of these protocols and although no interview currently meets all criteria, future research needs to be conducted to evaluate the importance of each of these domains in impacting the reliability and validity of a protocol.

Of particular importance is the missing psychometric information on each of these protocols. For example, very little is known about the interrater reliability of these protocols—a key question because this sets a limit on validity but also because the field would like the results not to be interviewer dependent (i.e., that another interview would have produced very different information and have come to different conclusions.) Of even greater concern is that there is limited information on the postdictive validity of these protocols (e.g., what are the error rates of these interviews?). Knowing error rates is a key piece of information in rendering a technique admissible in court proceedings. Finally, the extent to which training in these protocols is effective is unknown as there are few data showing fidelity to any protocol in actual practice in the field. The NICHD Investigative Interview Protocol has the most psychometric data but also appears to have significant gaps in this psychometric information as well as content.

There are other major pieces of missing information: the incremental validity of these interviews; how to adapt to the developmental variability of children; the cultural appropriateness of these protocols; and the extent to which component domains are validly executed (e.g., rapport, truth/lie distinction, prepositional competence). Clearly much more research is needed to further understand the abilities of protocols to achieve these ends. For example, a common procedure to establish prepositional competence is to have the child demonstrate that they know prepositions like “in” and “on top of” with objects such as a marker and a Kleenex box. However, questions can be raised regarding the extent to which generalizations can be made from this demonstration to whether a child knows whether a finger went “in” his or her vagina or anus.

A key issue is that these protocols can at best be “semi-structured.” Because each child and each potential abuse situation is unique, the interviewer must be given leeway to adapt general principles to the individual situation. For example, there is no mechanical process that can be followed to develop rapport and thus, there will also be an “art” of interviewing. Research will be needed to understand what interviewer characteristics seem relevant to making these decisions on the fly in actual interviews as it may reveal that some individuals are better suited than others to conduct these interviews.

It should be noted that some of this psychometric research is extremely difficult to conduct. There are important ethical constraints that will limit the research that can be done. For example, conducting multiple interviews with actual cases to determine interrater reliability may be both forensically and ethically complicated.

It may also be difficult to conduct this research in analog settings as asking children the kind of questions required in a sexual abuse investigation will raise legitimate ethical concerns. However, without finding a way to address these questions, it is difficult for the field to claim that its practice is evidence based and difficult for investigators to demonstrate adequate psychometrics of their interviews.

These protocols are being asked to accomplish a lot including to be applicable to a wide range of developmental stages, to explore very sensitive information, with a wide range of child characteristics (e.g., withdrawn to hyperactive; Caucasian versus Hispanic), in a wide range of jurisdictions (some even internationally), in a wide range of individual contexts (e.g., a non supportive, poorly functioning mother), and to achieve a wide range of objectives (e.g., establish rapport, not be leading, be sensitive to the presence of threats or bribes, and most importantly to gather complete and accurate information about acts that may have occurred years earlier in a developmentally not fully developed individual). These complexities are important and illustrate the major issues in forensic interviewing. It might be that multiple protocols may need to be developed or that the evaluative questions regarding these interviews need to be more nuanced, i.e., more along the lines of Gordon Paul's (1967) "ultimate question" regarding psychotherapy, "What protocol, by whom, is most effective for this individual, with this specific situation, and why?"

Finally, it may be best practice to place both these protocols and the field interviews into a quality improvement system. Since there is so much to be known about the quality of the interviews themselves as well as the quality of a particular interview protocol, it may be best practice for data to be continuously gathered on several quality dimensions. Fidelity to the protocol can be measured in each interview and interviewers can be given feedback on problems or stuck points. This feedback should be provided, as in all quality improvement procedures, in a supportive manner. Conducting forensic interviews well is an extremely difficult task given the idiosyncratic nature of each child and case, the complexity of the protocols, as well as functioning in a rather complex legal and even clinical context. In addition, the protocols themselves need to be constantly evaluated and improved. Interviewer feedback can be gained regarding issues such as ambiguities or areas where more support is needed. Feedback from other stakeholders can also be systematically gathered, e.g., from parents, prosecuting attorneys, and defense attorneys. In addition, this quality improvement system ought to gather some of the key psychometric data that are missing, benchmark these numbers, and constantly try to improve them.

References

- Aldridge, J., Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., & Bowler, L. (2004). Using a human figure drawing to elicit information from alleged victims of child sexual abuse. *Journal of Consulting and Clinical Psychology, 72*, 304–316.
- American Educational Research Association, W. C., American Psychological Association, W. C., & National Council on Measurement in Education, W. C. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

- American Psychological Association. (2003). Guidelines on multicultural education, training, research, practice, and organizational change for psychologists. *American Psychologist*, 58(5), 377–402.
- Anderson, W. L. (2007). *The trials of the Edenton Seven, or how Joseph Goebbels became the standard for North Carolina justice*. In LewRockwell.com. Retrieved June 26, 2012.
- Anderson, J., Ellefson, J., Lashley, J., Miller, A. L., Olinger, S., Russell, A., ... Weigman, J. (2007). *The CornerHouse forensic interview protocol: RATAC*. Retrieved from <http://www.ncptc.org/vertical/Sites/%7B8634A6E1-FAD2-4381-9C0D-5DC7E93C9410%7D/uploads/%7B0CB5FDDE-6496-40B7-8D70-E4EE6498903D%7D.PDF>
- Brown, D. A., Lewis, C., Lamb, M. E., & Stephens, E. (2012). The influences of delay and severity of intellectual disability on event memory in children. *Journal of Consulting and Clinical Psychology*, 80, 829–841. doi:10.1037/a0029388.
- Berenson, A. B., Chacko, M. R., Wiemann, C. M., Mishaw, C. O., Friedrich, W. N., & Grady, J. J. (2000). A case-control study of anatomic changes resulting from sexual abuse. *American Journal of Obstetrics and Gynecology*, 182(4), 820–834.
- Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, 113(3), 403–439.
- Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony*. Washington, DC: American Psychological Association.
- Cyr, M., & Lamb, M. E. (2009). Assessing the effectiveness of the NICHD investigative interview protocol when interviewing French-speaking alleged victims of child sexual abuse in Quebec. *Child Abuse & Neglect*, 33(5), 257–268.
- Darvish, T., Hershkowitz, I., Lamb, M. E., & Orbach, Y. (2005, Jan). The production of investigative leads in child sexual abuse interviews using the NICHD protocol. Paper presented to the Society for Applied Research on Memory and Cognition; Victoria, New Zealand.
- Dion, J., & Cyr, M. (2008). The use of the NICHD Protocol to enhance the quantity of details obtained from children with low verbal abilities in investigative interviews: A pilot study. *Journal of Child Sexual Abuse*, 17(2), 144–162.
- Elliott, A. N., O'Donohue, W. T., & Nickerson, M. A. (1993). The use of sexually anatomically detailed dolls in the assessment of sexual abuse. *Clinical Psychology Review*, 13(3), 207–221.
- Fontes, L., & Plummer, C. (2010). Cultural issues in disclosures of child sexual abuse. *Journal of Child Sexual Abuse*, 19(5), 491–518.
- Frasier, L. D., & Makoroff, K. L. (2006). Medical evidence and expert testimony in child sexual abuse. *Juvenile and Family Court Journal*, 57(1), 41–50.
- Hardy, C. L., & Van Leeuwen, S. A. (2004). Interviewing young children: Effects of probe structures and focus of rapport-building talk on the qualities of young children's eyewitness statements. *Canadian Journal of Behavioural Science*, 36(2), 155–165.
- Haynes, S. N., Smith, G. T., & Hunsley, J. D. (2011). *Scientific foundations of clinical assessment*. New York, NY: Routledge.
- Heger, A., Ticson, L., Velasquez, O., & Bernier, R. (2002). Children referred for possible sexual abuse: Medical findings in 2384 children. *Child Abuse & Neglect*, 26(6/7), 645.
- Hershkowitz, I. (2001). Children's responses to open-ended utterances in investigative interviews. *Legal and Criminological Psychology*, 6(1), 49.
- Hershkowitz, I. (2002). The role of facilitative prompts in interviews of alleged sex and abuse victims. *Legal and Criminological Psychology*, 7(1), 63.
- Hershkowitz, I., Fisher, S., Lamb, M. E., & Horowitz, D. (2007). Improving credibility assessment in child sexual abuse allegations: The role of the NICHD investigative interview protocol. *Child Abuse & Neglect*, 31(2), 99–110.
- Hershkowitz, I., Orbach, Y., Lamb, M. E., Sternberg, K. J., & Horowitz, D. (2001). The effects of mental context reinstatement on children's accounts of sexual abuse. *Applied Cognitive Psychology*, 15(3), 235–248.
- Horowitz, S. W., Lamb, M. E., Esplin, P. W., Reiter-Lavery, L., & Krispin, C. (1995). Establishing ground truth in studies of child sexual abuse. *Expert Evidence*, 4, 42–51.
- Lamb, M. E., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2008). *Tell me what happened: Structured investigative interviews of child victims and witnesses*. Hoboken, NJ: Wiley.

- Lamb, M., Orbach, Y., Hershkowitz, I., Esplin, P. W., & Horowitz, D. (2007). A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: A review of research using the NICHD investigative interview protocol. *Child Abuse & Neglect*, *31*(11-12), 1201–1231.
- Lamb, M. E., Orbach, Y., Sternberg, K. J., Aldridge, J., Pearson, S., Stewart, H. L., ... Bowler, L. (2009). Use of a structured investigative protocol enhances the quality of investigative interviews with alleged victims of child sexual abuse in Britain. *Applied Cognitive Psychology*, *23*(4), 449–467.
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Aldridge, J., Bowler, L., Pearson, S., & Esplin, P. W. (2006, July). *Enhancing the quality of investigative interviews by British police officers*. Paper presented at the Second International Investigative Interviewing conference, Portsmouth, England.
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., Stewart, H., & Mitchell, S. (2003). Age differences in young children's responses to open-ended invitations in the course of forensic interviews. *Journal of Consulting and Clinical Psychology*, *71*(5), 926–934.
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., & Mitchell, S. (2002). Is ongoing feedback necessary to maintain the quality of investigative interviews with allegedly abused children? *Applied Developmental Science*, *6*(1), 35–41.
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Hershkowitz, I., Horowitz, D., & Esplin, P. W. (2002). The effects of intensive training and ongoing supervision on the quality of investigative interviews with alleged sex abuse victims. *Applied Developmental Science*, *6*(3), 114–125.
- Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, *25*(12), 720–725.
- Malloy, L. C., Brubacher, S. P., & Lamb, M. E. (2011). Expected consequences of disclosure revealed in investigative interviews with suspected victims of child sexual abuse. *Applied Developmental Science*, *15*(1), 8–19.
- Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse & Neglect*, *24*(6), 733–752.
- Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, *31*, 109–118.
- Porter, S., Yuille, J. C., & Bent, A. (1995). A comparison of the eyewitness accounts of deaf and hearing children. *Child Abuse & Neglect*, *19*(1), 51–61.
- Rabinowitz, D. (2003). *No crueller tyrannies*. New York, NY: Free Press.
- Raskin, D. C., & Yuille, J. C. (1989). Problems in evaluating interviews of children in sexual abuse cases. In M. P. Toglia, S. J. Ceci, & D. F. Ross (Eds.), *Adult perceptions of child testimony*. New York, NY: Springer.
- Schreiber, N., Bellah, L. D., Martinez, Y., McLaurin, K. A., Strok, R., Garven, S., & Wood, J. M. (2006). Suggestive interviewing in the McMartin Preschool and Kelly Michaels Daycare abuse cases: A case study. *Social Influence*, *1*(1), 16–47.
- Sternberg, K. J., Lamb, M. E., Orbach, Y., Esplin, P. W., & Mitchell, S. (2001). Use of a structured investigative protocol enhances young children's responses to free-recall prompts in the course of forensic interviews. *Journal of Applied Psychology*, *86*, 997–1005.
- Vieth, V. I. (2009). Forensic interviewer at trial: Guidelines for the admission and scope of expert witness testimony concerning an investigative interview in a case of child abuse. *William Mitchell Law Review*, *36*, 186.
- Yuille, J. C., Hunter, R., Joffe, R., & Zaparniuk, J. (1983). Interviewing children in sexual abuse cases. In G. S. Goodman, B. L. Bottoms (Eds.), *Child victims, child witnesses: Understanding and improving testimony* (pp. 95–115). New York, NY US: Guilford Press.
- Yuille, J. C., Hunter, R., Joffe, R., & Zaparniuk, J. (1993). Interviewing children in sexual abuse cases. In G. S. Goodman & B. L. Bottoms (Eds.), *Child victims, child witnesses: Understanding and improving testimony* (pp. 95–115). New York, NY: Guilford Press.
- Zaparniuk, J., Yuille, J. C., & Taylor, S. (1995). Assessing the credibility of true and false statements. *International Journal of Law and Psychiatry*, *18*(3), 343–352.