CHAPTER 13

Forensic Interviews with Children
A Two-Way Street: Supporting Interviewers in Adhering to Best Practice Recommendations and Enhancing Children’s Capabilities in Forensic Interviews

DEIRDRE BROWN
MICHAEL E. LAMB

FORENSIC INVESTIGATORS HAVE long acknowledged how crucial it is to elicit reliable information from the alleged victims of child sexual abuse because information from other sources is seldom available. A large body of research (summarized in previous chapters) has shown, however, that due to immature cognitive, communicative, and social skills, children are often unable to provide detailed accounts without support from interviewers when retrieving and structuring reports of their experiences. Although there are valid reasons to be wary of children’s responses to suggestive questioning techniques, researchers have clearly demonstrated that children are able to provide reliable testimony when questioned appropriately. The quality of their testimony may be influenced by characteristics of the children themselves, the events they have experienced, and, importantly, the ways in which interviewers attempt to elicit information.

The perceived reliability of any testimony elicited from child witnesses may influence the verdict in a trial, or even the likelihood of a case reaching the courts. Research has shown that both the presence of an eyewitness and any testimony that raises doubt about their credibility significantly affect juror decision making (Hatvany & Strack, 1980; Loftus, 1974; Sigler & Couch,
2002). Historically, child witnesses have been viewed as particularly unreliable (see Ceci & Bruck, 1998, for review), and surveys of various legal professionals, psychologists, and laypeople have revealed negative perceptions of children's competency and suggestibility but positive perceptions of their sincerity and honesty (e.g., Leippe, Brigham, Cousins, & Romanczyk, 1989; Leippe & Romanczyk, 1987; Ross, Dunning, Toglia, & Ceci, 1990; Yarmey & Jones, 1983). Furthermore, jurors perceive testimony from alleged victims to be centrally important in child sex abuse cases (Myers, Redlich, Goodman, Prizmich, & Imwinkelried, 1999). As a result, any doubts about the quality of the interviewing techniques may reduce the likelihood of conviction based on the children's allegations. High-quality interviewing techniques are thus essential to minimize challenges to the reliability of children's testimony and thereby maximize the protection of both child victims and innocent suspected perpetrators.

TRAINING INVESTIGATORS: A FAILURE TO TRANSLATE THEORY INTO PRACTICE

Three decades of research exploring factors affecting the reliability of children's eyewitness testimony has yielded broad consensus about ways in which interviewers can help abused children recall and report their experiences without compromising the reliability of those reports (e.g., American Professional Society on the Abuse of Children [APSAC], 1997; Fisher & Geiselman, 1992; Goodman & Melinder, 2007; Home Office, 2002; Jones, 1992; Lamb & Brown, 2006; Lamb, Hershkowitz, Orbach, & Esplin, in press; Lamb, Orbach, Hershkowitz, Esplin, & Horowitz, in press; Orbach et al., 2000; Perona, Bottoms, & Sorenson, 2006; Poole & Lamb, 1998; Yuille, Marxsen, & Cooper, 1999). Although the guidelines differ in some details, common elements can be identified. Most importantly, all approaches place primary emphasis on eliciting as much information as possible using open-ended invitations and questions that give control of the information-sharing process to children. Thereafter, interviewers are advised to pursue additional information or clarify ambiguous statements using the most open form of questioning possible, only using focused questions toward the end of the interview when needed. The importance of establishing rapport and the ground rules are universally highlighted, and practice interviews about neutral topics are widely recommended.

Unfortunately, many studies have demonstrated discrepancies between recommended best practice and the conduct of forensic interviews with children (Cederborg, Orbach, Sternberg, & Lamb, 2000; Craig, Scheibe, Kircher, Raskin, & Dodd, 1999; Cyr, Lamb, Pelletier, Leduc, & Perron, 2006; Davies, Westcott, & Horan, 2000; Korkman, Santila, & Sandnabba,
2005; Lamb, Hershkowitz, Sternberg, Esplin, et al., 1996; Lamb, Sternberg, & Esplin, 2000; Lamb, Sternberg, Orbach, Aldridge, et al., 2006; Powell, Fisher, & Wright, 2005; Sternberg, Lamb, Davies, & Westcott, 2001; Sternberg et al., 1996; Sternberg et al., 1997; Thoresen, Lønnum, Melinder, Stridbeck, & Magnusson, 2006; A. G. Walker & Warren, 1995; N. Walker & Hunt, 1998; Westcott & Kynan, 2006). Studies of training programs for forensic interviewers have consistently shown few improvements in recommended practice and little avoidance of inappropriate techniques (e.g., closed or focused questions, option-posing prompts, suggestive questions, anatomical dolls) that may compromise the reliability of the children’s evidence (Aldridge, 1992; Aldridge & Cameron, 1999; Cederborg & Lamb, 2006; Craig et al., 1999; Freeman & Morris, 1999; Lamb, Sternberg, Orbach, Esplin, et al., 2002; Lamb, Sternberg, Orbach, Hershkowitz, 2002; Stevenson, Leung, & Cheung, 1992; Warren et al., 1999; Westcott, Kynan, & Few, 2006). Thus, although interviewers may know how to interview children appropriately (particularly after training), they may not actually interview children in that fashion.

Expert professional groups agree on the following key points:

- Interviewers should develop rapport, discuss ground rules, use open-ended prompting as a primary strategy for eliciting information, use specific or closed prompts for eliciting critical details (i.e., particularization) or addressing ambiguities only later in the interview, and avoid suggestive questioning.
- Training programs appear to have little impact on the way interviewers behave in interviews, although participants learn more about developmentally sensitive interviewing techniques.
- Bridging the gap between theory and practice: The development of interview techniques and protocols is critical.

Several approaches to conducting forensic interviews with children have been developed, with considerable recent interest in structured interview protocols. The benefits of more structured techniques or protocols are twofold: First, they operationalize guidelines, thus helping to bridge the gap between theory and practice. Second, they help children recall and report their experiences by addressing their memorial and communicative skill deficits. We next briefly review each of these approaches, mentioning any research on their effectiveness that has been published.

Cognitive Interview

The Cognitive Interview (CI) was originally developed for use in police investigations with adult eyewitnesses (Fisher, Geiselman, Raymond,
Jurkevich, & Warhaftig, 1987), and was further revised to incorporate instructions designed to facilitate rapport, encourage open-ended questioning, hand control of the information-reporting process to the witness, and encourage complete reporting of all information recalled. The revision also incorporated mnemonic techniques for aiding retrieval of information from memory (Fisher & Geiselman, 1992). The revised CI includes five stages, (1) introduction to the purpose of the interview, (2) request for open-ended recall, (3) probed recall, (4) review, and (5) closing. During the probed recall stage, four techniques are used to encourage complete recall and reporting:

1. Witnesses are encouraged to report everything they can remember, even small details they may feel are unimportant.
2. Witnesses are asked to mentally reinstate the event and report any details that they can recall, including descriptions of the environment, the people, smells, feelings, and reactions to events.
3. Witnesses are asked to recall the event in different temporal sequences (e.g., reverse order, starting from the most salient aspect and moving forward and then backward in time from that aspect).
4. Witnesses are asked to recall the event from different perspectives (e.g., from the perspective of someone else who was present).

In studies concerned with staged events, use of the CI yielded significantly more information than standard police interviews, without adverse effects on accuracy (Fisher et al., 1987).

In the wake of such promising findings with adults, several researchers explored the CI's effectiveness with children. Laboratory studies involving children ranging in age from 4 (e.g., Holliday & Albon, 2004) to 11/12 (e.g., Akehurst, Milne, & Kohnken, 2003) years yielded mixed findings. Several researchers reported improved recall when children were interviewed using the CI (Akehurst et al., 2003; Geiselman & Padilla, 1988; Holliday, 2003; Holliday & Albon, 2004; Milne & Bull, 2003; Saywitz, Geiselman, & Bornstein, 1992), and some showed increased resistance to suggestion and/or misinformation (Holliday, 2003; Holliday & Albon, 2003; Milne & Bull, 2003). However, some researchers reported increased numbers of errors (Memon, Wark, Bull, & Köhnen, 1997), although the proportion of incorrect to correct information appeared to remain constant (McCauley & Fisher, 1995), and still others reported no improvement when children were interviewed using the CI (Memon, Cronin, Eaves, & Bull, 1995).

The CI has not yet been evaluated either in the field or in relation to traumatic or stressful events, but it may help children provide more complete accounts of past experiences. The technique may not work with children younger than 8 years, however, because they cannot understand
some of the component techniques, such as changing perspectives or temporal sequences (Geiselman, 1999). Most studies have also involved short delays between the event and the interview (Poole & Lamb, 1998), although Larsson, Granhag, and Spjut (2003) showed increases in the amounts of correct information reported by 10- to 11-year-old children after a 6-month delay. Neither the extent to which the CI may be useful over forensically relevant delays of 6 to 18 months nor the interaction between child characteristics (e.g., developmental level, motivation), event characteristics (e.g., experienced versus witnessed, single vs. repeated, negative vs. neutral or positive), and interview characteristics (e.g., single vs. repeated) have yet been established, however.

Two studies (Milne & Bull, 1996; Robinson & McGuire, 2006) have explored the effectiveness of the CI when the children interviewed have mild mental retardation. As with many studies involving typically developing children, both studies found that more correct information was reported when the CI was used, but Robinson and McGuire also found increases in the numbers of incorrect details reported. Neither study found that susceptibility to suggestion was affected.

In summary:

- Originally developed for use with adult eyewitnesses, the Cognitive Interview (CI) shows promise as a technique for increasing the amount of information reported by children.
- Evidence is mixed with respect to its impact on accuracy and suggestibility.
- More research is needed to explore its usefulness when more forensically relevant events are recalled after longer delays, or when interviews are repeated.

Narrative Elaboration Training

Saywitz and Snyder (1993, 1996) developed an innovative interviewing technique to explore the effectiveness of preinterview training and practice talking about experienced events on children's later reports of target events. Narrative Elaboration Training (NET) addresses metalinguistic deficits by teaching children about categories of information and the level of detail that makes reports of experienced events more complete, and it addresses retrieval deficits by providing pictorial cue cards to prompt retrieval of different types of information (specifically about people, setting, actions, conversations, and affects). Children are first trained to talk about experienced events using the cards, and are then asked about the to-be-remembered event. In the laboratory, the NET helps children, including
preschoolers and children with mental retardation, to report events more completely, without compromising accuracy (Bowen & Howie, 2002; Brown & Pipe, 2003a, 2003b; Dorado & Saywitz, 2001; Nathanson, Crank, & Saywitz, in press; Saywitz & Snyder, 1996; Saywitz, Snyder, & Lamphear, 1996), and does not elicit reports of false events (Camparo, Wagner, & Saywitz, 2001). On the other hand, researchers have shown that use of component strategies of the NET (such as verbal prompting for categories of information, or presenting pictorial cards) without the training that is an integral part of the NET, can be just as effective as the NET (Bowen & Howie, 2002; Brown & Pipe, 2003a; Elischberger & Roebers, 2001; Roebers & Elischberger, 2002).

The effectiveness of NET over a long delay has been examined in only one study. Brown and Pipe (2003b) showed that more information was reported, without adverse effects on accuracy, when 8- to 9-year-old children were interviewed after a 9-month delay. The NET has yet to be evaluated in the context of more ecologically valid events, in field studies, and across repeated interviews.

In summary:

- The NET may help children of all ages, including those with cognitive limitations, report more forensically relevant details about experienced events.
- Components of the technique can be just as helpful as the full package.
- Further research is needed in situations likely to be encountered in forensic contexts (e.g., in field studies exploring ecologically valid events, with repeated interviews after varying delays).

The Step-Wise Interview

Developed by Yuille and his colleagues (Yuille, Hunter, Joffe, & Zaparniuk, 1993), the Step-Wise Interview is designed to: (a) increase interviewer skills thereby reducing the need for multiple interviews and reducing the stress placed on interviewees; (b) guide interviewers to use nonsuggestive questions in order to avoid contaminating children’s reports and to maximize the amount of information obtained; (c) aid recall using techniques drawn from the CI (see earlier discussion) if appropriate; and (d) allow for flexibility so that, for example, interviewers can adapt their approach depending on the function of the interview (e.g., child protection as opposed to criminal prosecution). The Step-Wise Interview begins with rapport building, after which the children describe two specific past experiences to practice talking about the past and responding to questions while interviewers evaluate their verbal skills. Interviewers then explain the need to tell the truth and
introduce the event of concern in a step-wise fashion, encouraging a free narrative about the event using nonsuggestive and open-ended questioning, followed by more specific questions (if necessary for clarification). Interview aids such as drawings and dolls are used only after children have reported an abusive event, if at all. Children are then thanked, given opportunities to ask any questions that they may have, and told how to contact the interviewers if they later want to do so.

This protocol has been used by some police and child protection services in Canada, the United Kingdom, and the United States, but little effort has been made to evaluate the relative strengths and weaknesses of Step-Wise Interviews in the field. In one field study, however, interviewers who were trained in the Step-Wise Interview elicited fewer statements that were considered by the interviewers themselves to be contaminated than did untrained interviewers (Marxsen, Yuille, & Nisbet, 1995). No researchers have evaluated the impact of the Step-Wise Interview on variables known to influence children’s reliability (e.g., event characteristics, child characteristics, interview characteristics).

In summary:

• Although widely used, the Step-Wise Interview has not been systematically evaluated in laboratory or field studies.

**CORNERHOUSE RATAc PROTOCOL**

In the United States, the Finding Words technique, otherwise known as CornerHouse’s RATAc (Rapport, Anatomy Identification, Touch Inquiry, Abuse, and Closure) protocol (Walters, Holmes, Bauer, & Vieth, 2003), has been officially adopted by many jurisdictions and states. Unfortunately, however, there has been no research on what RATAc-trained interviewers actually do when interviewing children. There is an urgent need for evaluation of this popular technique particularly because, as noted, researchers have repeatedly found that training programs do not generalize to practice in the field. Because the Finding Words approach includes such aids as drawings and anatomical dolls that can adversely affect the reliability of the information elicited, researchers need to evaluate the quality of information elicited by interviewers following the Finding Words guidelines.

In summary:

• Although gaining popularity in the United States, the effectiveness and safety of RATAc protocol for eliciting reliable testimony from children has yet to be systematically evaluated.
Achieving the Best Evidence in Criminal Proceedings

Achieving the Best Evidence (ABE) in Criminal Proceedings (Home Office, 2002) was published by the U.K. Home Office and Department of Health to provide standards of practice for video-recorded interviews of children, following legislative changes that allowed video recordings to be presented as evidence in court provided the child was available for cross- and reexamination. Like its predecessor, the Memorandum of Good Practice (Home Office, 1992), ABE advocates careful planning so that interviews are conducted in a thorough and safe manner, reducing the need for subsequent interviews. According to ABE, planning should take into account the child’s level of development, language ability, state of mind, culture, and any disabilities. The interview itself is divided into four phases: (1) rapport-building (the purpose of the interview, the ground rules, and the importance of telling the truth); (2) the free narrative account of the alleged abuse with no reference made by interviewers to the alleged actions or the perpetrator; (3) questioning following a hierarchical progression, beginning with open-ended questions, followed by specific nonleading questions (avoiding yes/no questions), and finally closed questions (which provide multiple options for responding); (4) leading questions which may have limited value in court because they may be ruled inadmissible; and (5) a closing phase, in which interviewers check their understanding, discuss a neutral event, and provide contact details in case the child later has further information to discuss.

At any stage of the interview, interviewers are counseled to shift away from difficult topics to reestablish rapport whenever children become anxious, reticent, or have difficulty before returning to substantive issues. ABE suggests that dolls and props may be used after the substantive portion of the interview is finished, if necessary to clarify terms the child may have used. ABE does not support the use of dolls and props as a major part of the evidence presented to court, especially if the actions demonstrated with the props have not earlier been verbalized and therefore require interpretation. It also advocates returning to an earlier phase of the interview to substantiate what the child is attempting to communicate with the doll or prop in a verbal form. ABE synthesizes recommendations derived from a vast body of research and presents them as a set of guidelines for the conduct of forensic interviews with children, but does not provide an interview protocol or explain how to the guidelines might be implemented. Research has yet to explore how well untrained interviewers put these guidelines into practice and the extent to which the ABE helpfully guides interviews with children of different ages, abilities (e.g., children with mental retardation), and across varying delay intervals. However, interviewers trained to follow the earlier
Memorandum failed to follow many of the guidelines in practice (Sternberg et al., 2001).

In summary:

- ABE provides guidelines for the conduct of forensic interviews with children in the United Kingdom.
- There has been no systematic research on either the quality of interviews conducted according to ABE or the effects of factors relating to the child, the event, and the interview (e.g., delay, single vs. repeated).

**National Institutes of Child Health and Human Development Investigative Interview Protocol**

Researchers at the National Institutes of Child Health and Human Development (NICHD) created an interview protocol that operationalized the recommendations from research to help forensic investigators conduct developmentally appropriate interviews with children (Lamb, Hershkowitz, Orbach, et al., in press; Lamb, Orbach, Hershkowitz, et al., in press; Orbach et al., 2000; Sternberg, Lamb, Orbach, Esplin, & Mitchell, 2001).

The NICHD protocol begins with an introductory phase that promotes rapport, includes a “truth and lie ceremony” to communicate to children the importance of truth-telling, and establishes the ground rules for the interview, such as the acceptability of responding “I don’t know,” or “I don’t understand,” and correcting the interviewer. Following this, children are asked to describe two recent experiences (e.g., Christmas, Passover, or a recent birthday) and are prompted for further elaboration using open-ended prompts, followed by more focused prompts such as time segmentation (e.g., “Tell me everything that happened from the time you went home until you went to the store”; Sternberg, Lamb, Esplin, Orbach, & Hershkowitz, 2002, p. 420). Providing children with opportunities to talk about recent events provides practice retrieving and reporting detailed information, exposure to the types of prompts the interviewers will use later in the interview, and an indication of the level of detail required to understand the children’s experiences. The substantive portion of the interview is then introduced, with a series of open-ended prompts used to encourage children to discuss the alleged abuse (e.g., “Tell me why you came here today.”). None of these prompts refer to the actions, perpetrator, or location of the alleged abuse. The NICHD protocol includes specific question stems to be used if a child has not reported (the use of the term “disclosed” presumes that abuse has occurred) abuse in response to more open-ended prompts. The questions were formulated to be minimally
leading, and to be used only when necessary (More than 80% who report do so in response to the most general open prompts; e.g., Sternberg et al., 2001.) The NICHD protocol recommends providing open-ended prompts for further information such as “Tell me everything about that” following responses to any direct questions so as to minimize the risk of contamination by leading or suggestive questioning. When children indicate that there were multiple instances of abuse, the NICHD protocol recommends eliciting a description of the most recent incident first, followed by the first incident, and then any other specific well-remembered incidents (e.g., “the time in the bathroom,” “the time when you were camping”). Again, prompting for further information involves open-ended prompts, followed by more focused questions if necessary. The interview concludes with discussion of neutral events. One of the distinctive features of the protocol is that it uses children’s responses as cues to elicit further information, resulting in a child-directed rather than an interviewer-directed approach throughout the interview.

Field studies have compared interviews conducted by investigators before and after they have been trained to use the protocol in the United States, Israel, the United Kingdom, and Quebec (Cyr et al., 2006; Lamb et al., 2006; Orbach et al., 2000; Sternberg et al., 2001, 2002). In all sites, the majority of children who reported abuse when questioned using the protocol did so in response to open-ended prompts like “Tell me why you came here today” while protocol-guided interviewers adhered to recommended practices better than those not using the protocol—offering more open-ended invitations and fewer problematic prompts, such as option-posing or suggestive questions. Similarly impressive findings are obtained even with very young (4- to 8-year-old) children (Lamb et al., 2003).

While the above studies demonstrate dramatic changes in the techniques used by interviewers in actual investigations of child abuse, however, the quality or accuracy of that information is usually unknown, although lab studies consistently show that open-ended prompts elicit more accurate information than recognition-memory prompts. The accuracy of information elicited when children are interviewed about known but neutral events with the NICHD protocol was recently examined in the United Kingdom (Brown, Lewis, Lamb, Pipe, & Orbach, 2007). Children (5 to 7 years) participated in a staged event at school and were interviewed 4 to 6 weeks and again 6 months later using the NCNIHD protocol. Preliminary analyses indicated that, as in field studies, interviewers made use of recommended interviewing techniques, with open prompts (invitation and cued invitations) making up 75% of the prompts at 1 month and 53% at 6 months. Most of the information provided by the children came in response to open-ended prompting (71% at 1 month, compared to 16% in response to direct
prompts and 5% to suggestive, and 56% at 6 months compared to 27% to direct prompts, and 8% to suggestive prompts). Furthermore, the quality of information varied depending on the type of prompt by which it was elicited. Information reported in response to invitations was most accurate (84%), and the accuracy rates remained stable over time, except that the accuracy of information reported in response to suggestive prompts decreased from 74% to 53%. Analysis of the average amount of information elicited per prompt indicated that invitations, cued invitations and direct prompts were similarly productive at 1 month (three to four pieces of information per prompt), whereas suggestive questions elicited an average of only two pieces of information. After a 6-month delay, the average productivity of the open prompts remained stable (four pieces of information) but the productivity of direct prompts decreased to two pieces of information, and that of suggestive prompts remained at two pieces. In another recent study, Lamb, Orbach, Hershkowitz, Horowitz, and Abbott (in press) compared the accounts provided by young victims with those provided by the alleged perpetrators. As predicted, details provided by the children in response to open-ended prompts were more likely to be confirmed by the perpetrators than those provided in response to recognition prompts. All of these studies thus document how effectively the NICHD protocol helps children provide detailed and accurate accounts of their experiences, even after a long delay. They highlight how reliance on open prompting can elicit both more as well as more accurate information than closed or suggestive questioning.

Recent research with the NICHD protocol has also explored its use with special groups of witnesses—those who are reluctant or unwilling to report abuse that has occurred and those with learning disabilities. Although the dangers of eliciting false reports of abuse from children have been discussed widely, little attention has been paid to an equally serious issue—children who have experienced abuse but do not disclose. In Israel, for example, approximately a third of the suspected child victims interviewed do not disclose abuse during forensic interviews (Hershkowitz, Horowitz, & Lamb, 2005, 2007) while 23% of the 294 American children studied by Pipe, Lamb, Orbach, Stewart, et al. (2007) did not make a disclosure when interviewed, despite sound reasons (e.g., suspect confession) to suspect that abuse had occurred. The likelihood of disclosure increases with age, and decreases if the suspected perpetrator is a close family member, especially when the children are younger (Hershkowitz et al., 2005, 2007; Pipe, Lamb, Orbach, Stewart, et al., 2007). Similarly, delayed (as opposed to non-) disclosure is also more likely when the alleged perpetrator is a close family member. Modifications to the protocol for use with reluctant witnesses are currently being developed and evaluated.
On another front, lawyers, forensic interviewers, child psychiatrists and psychologists are increasingly calling for research exploring the abilities and vulnerabilities of witnesses with mental retardation and other learning disabilities (e.g., Jones, 2003; Kebbell & Davies, 2003; O’Kelly, Kebbell, Hatton, & Johnson, 2003; Stobbs & Kebbell, 2003). Children with mental retardation are much more likely than their typically developing counterparts either to witness or experience abuse, but are less likely to have their complaints investigated and/or prosecuted because their capacity to provide reliable evidence is often doubted (Cederborg & Lamb, 2006; Hershkovitz, Horowitz, & Lamb, in press; Westcott & Jones, 1999). The few studies available suggest that these children are as capable of providing reliable information about their experiences as children matched for developmental level (mental age) when interviewed with optimal techniques (free recall using open questions) although they may perform worse than children of the same chronological age (Dent, 1986; Michel, Gordon, Ornstein, & Simpson, 2000), especially in response to suggestive and/or specific questions (e.g., Agnew & Powell, 2004; Henry & Gudjonsson, 2003). In an ongoing program of research, we are exploring the ability of 7- to 12-year-old children with mental retardation (CWMR) to provide meaningful accounts of personally experienced staged events when interviewed using the NICHD protocol. We are also exploring the effects of the severity of mental retardation (mild vs. moderate), delay between the experience and the interview (2 weeks vs. 6 months), and repeated interviewing, as well as the children’s ability to answer suggestive questions that vary in structure (open vs. closed) and content (central vs. peripheral detail). Preliminary analyses indicate that children with mild mental retardation performed as well or better than mental age-matched typically developing children, and at times as well as their chronological age-matched peers, on measures of total information, accuracy, and suggestibility. A slightly different pattern was evident for children with moderate mental retardation. It was encouraging to see that children with mental ages as young as 4 years could cope with protocol-guided interviews. They responded to open questions and provided meaningful and reliable information, although they were less accurate than their typically developing peers, particularly in response to questions with a leading or suggestive edge (Brown, Lewis, Lamb, & Stephens, 2007).

As mentioned earlier, interviewer training depressingly often yields improvement in trainees’ knowledge but no meaningful changes in the ways in which they actually interview alleged victims. Recognizing this, training in use of the NICHD protocol has always been accompanied by efforts to provide continued support, guidance, and feedback on interviews conducted after starting to use the protocol. The incremental value of verbal and written feedback during the course of training was experimentally demonstrated
previously in individual (Adams, Fields, & Verhave, 1999; Clark, 1971; Frayer & Klausmeier, 1971; Sweet, 1966) and group (Gully, 1998) contexts, but the NICHD training model is the only one to systematically include feedback beyond the training period (i.e., in posttraining investigative interviews as well).

The importance of continuing quality control and feedback was assessed by comparing the effectiveness of four different training models designed to help interviewers implement recommended interviewing practices (Lamb, Sternberg, Orbach, Hershkowitz, et al., 2002). In all conditions, interviewers were first provided with a theoretical framework to help them understand how the recommended practices were consistent with basic research on children’s memory, linguistic, communicative, and social development. The first training condition only involved such conceptual training. In the second condition, interviewers were also urged to employ structured modules when building rapport, and they practiced using these modules. Interviewers in the third and fourth training conditions were also given copies of the NICHD protocol and practiced using it under close supervision. The third and fourth conditions differed with respect to the amount and type of supervision provided. Interviewers in the third condition attended intensive training courses, followed by monthly day-long group meetings in which their actual field interviews were scrutinized (using video recordings and transcripts) and desirable and undesirable practices were discussed. In addition, interviewers in the third condition received detailed written and verbal feedback on each of their subsequent field interviews. Interviewers in the fourth condition participated in the monthly meetings alongside those in the third condition but received no individual supervision and feedback on their own interviews.

The effects of these forms of training were assessed by examining the extent to which the interviewers employed open-ended as opposed to focused questions, the amount of information elicited using open-ended rather than focused prompts, and the extent to which the interviewers delayed introducing substantive information. In all cases, the performance of interviewers who had been trained using one of the four regimes was compared with that of the same interviewers conducting interviews with children of comparable age and circumstances in the 6 months prior to the training.

As expected, we found that meaningful long-term improvements in the quality of information obtained from young alleged victims of sexual abuse were observed only when well-established principles were operationalized in a clear and concrete fashion and when training was distributed over time, rather than provided in the form of a single initial session, however intensive. Significant differences between the baseline and training
conditions were clearly evident only when interviewers were guided by the NICHD protocol and continued to attend regular intensive training workshops. Such interviewers made more efforts to elicit information using open-ended prompts, elicited more information from the children's free recall, and were able to delay their first option-posing questions which, by definition, involved the introduction of information by the interviewer rather than by the child. By contrast, interviewers who received intensive short-term training but no continuing training generally performed little better than they had before training. There were no differences between interviews in conditions 3 and 4, however.

In a related study, furthermore, Lamb, Sternberg, Orbach, Esplin, et al. (2002) showed the adverse effects of the termination of supervision and feedback on investigators' performance. Forensic interviews conducted by trained interviewers who received close and continuing supervision and intensive individual feedback were compared with interviews conducted by the same interviewers in the 6 months immediately following the completion of training and the termination of the supervision-and-feedback. As predicted, the quality of the later interviews was inferior to that of the earlier interviews, as indexed by (a) declines in the use of open-ended prompts, (b) corresponding increases in reliance on more focused prompts, and (c) the earlier introduction of focused prompts. The expected changes in the interviewers' questioning style were accompanied by decreases in the amount of information elicited using free-recall prompts.

These results have important, although somewhat sobering, implications for those attempting to apply information gleaned from basic research in the real world. Clearly, it is possible to employ our professions' accumulated knowledge of memory and communicative development to improve the quality of information elicited from alleged victims of child abuse, but these benefits are obtained only when extensive efforts are made not only to train interviewers to adopt recommended practices, but to ensure the maintenance of these practices as well. Regardless of their skillfulness, interviewers continue to maintain or improve their skills only when they regularly review their own and others' interviews closely, discussing their strategies, successes, and mistakes with other interviewers. In Israel, for example, all interviewers are required to continue attending regular peer-review sessions of this sort, and this seems to have ensured that the investigative interviews conducted there are of the highest quality.

In summary:

- The NICHD protocol was developed to operationalize guidelines for best practice.
• Research in several countries indicates that use of the protocol has positive effects on both children and interviewers, promoting developmentally appropriate interviewing techniques.

• Most of the research has been conducted in the field but recent laboratory-analogue studies provide complementary evidence that use of the protocol helps interviewers elicit detailed and accurate information from children.

• The protocol has been tested with children ranging in age from 4 to 13 years and is being modified for use in interviews with children assumed to have been abused who are reluctant to disclose and those with mental retardation.

• Gains made by interviewers trained to use the protocol are more likely to be maintained when the interviewers continue to receive feedback.

EVALUATION OF TESTIMONY: DEVELOPMENT OF CRITERION-BASED CONTENT ANALYSIS

"Consumers" of forensic interviews, including police departments, social work agencies, prosecutors, judges, and juries, often face situations in which the allegations made by a young witness are contradicted by other possible informants (e.g., alleged perpetrators) and cannot be substantiated by corroborative information. To facilitate decision making in such circumstances, experts are often forced or asked to evaluate the children's credibility.

Some promising techniques for evaluating the credibility of children's accounts were inspired by Statement Reality Analysis (SRA), which had its origins in the experiences of a German forensic psychologist, Udo Undeutsch (1982, 1989). Undeutsch proposed that descriptions of events that really happened differ in content and quality from descriptions of events that were not actually experienced (Undeutsch 1982, 1989). In particular, he believed that experienced events are reported in richer detail and with clearer links to other real-world events than events that have been fabricated or imagined. Undeutsch was especially concerned with the qualitative characteristics of narrative accounts, believing that credibility was reflected by the extent to which descriptions of the target incidents were placed in temporal and spatial context, logically coherent though marked by digressions, and contained the unique details that appeared to define specific incidents rather than generic or general situations. Undeutsch did not claim that credible accounts necessarily contained more details than implausible accounts. He also recognized lack of confidence as an index of credibility rather than its absence.

Stetler and Koehnkken (1989) and Raskin and Esplin (1991c) formalized the "Undeutsch hypothesis" by developing a list of 19 criteria that could be used
to quantify features of children’s statements or accounts and thereby systematically evaluate the credibility of children’s accounts. Two field studies were then conducted to determine whether accounts of incidents that actually happened were assigned higher scores on a Criterion-Based Content Analysis (CBCA) checklist than accounts of incidents that did not happen (Boychuk, 1991; Raskin & Esplin, 1991a, 1991b). In both studies, substantially more criteria were present in plausible accounts than in doubtful statements. Raskin and Esplin (1991b) in fact reported no overlap between the distribution of scores assigned to statements about both confirmed and doubtful incidents, and similar results were reported by Boychuk (1991).

Although such findings suggest that the CBCA procedure can discriminate between truthful and nontruthful accounts, the procedure has been sharply criticized. Wells and Loftus (1991), for example, criticized the representativeness of the 40 cases included in Raskin and Esplin’s study; they emphasized the importance of using independent case facts (i.e., not judicial dismissal or lack of prosecution) to establish the plausibility of the allegations, questioned the evaluation of inter-rater reliability and asked about the association between children’s ages and their CBCA scores.

Before beginning our own validation studies, therefore, we undertook a study of the reliability of the CBCA coding system. Horowitz et al. (1997) found that trained raters showed high levels of agreement with one another regarding whether most of the CBCA criteria were present; the three raters were also highly reliable over a 3-month interval, and their high levels of agreement did not vary depending upon the age of the child or the degree to which the account appeared credible. It thus seemed that the CBCA system provided a reliable means of quantitatively evaluating children’s credibility, although variations in the levels of inter-rater and temporal reliability led Horowitz et al. to suggest that 5 of the 19 criteria be dropped and others redefined more carefully to enhance reliability.

In conducting validation studies, such as those conducted by Raskin and by Esplin and Boychuk, a major difficulty inevitably arises: How can researchers determine whether or not the alleged incident actually occurred? Because judicial disposition is itself influenced by the child’s statement, it does not constitute an independent validation of the child’s allegations. For research purposes, it is thus necessary to consider only independent validating information (Wells & Loftus, 1991). Raskin and Esplin (1991b) based their discrimination between “doubtful” and “confirmed” cases on the results of polygraphic and medical examinations, confessions, and eyewitness accounts, although it was obviously much more difficult to obtain independent evidence about the doubtful cases. Although recantations were obtained from several children in the doubtful groups, and there was no evidence that supported the children’s allegations of abuse, the confirmation that the
allegations were doubtful was weaker than comparable evidence about the confirmed cases. The absence of supporting or corroborative evidence should never be confused with contradictory evidence, however: corroborative evidence is lacking in the majority of cases. This is one of the reasons why field research on credibility is both so difficult to conduct and so important to administration of justice.

To strengthen and systematize the assessment of plausibility, Horowitz et al. (1995) suggested that multifaceted procedures should be used to synthesize the results of medical examinations, suspect statement, polygraphic examinations, witness statements, and other circumstantial or physical evidence when determining, with varying degrees of certainty, the probability that the alleged events actually occurred. An elaborate multifaceted procedure of this sort was then developed and used by Lamb et al. (1997) in Israel, where the established system made it easier to obtain a representative sample of cases. Because interviews were selected only on the basis of external characteristics, rather than following examination of the interviews themselves, the interviews varied widely in quality and length, with no standardized interview procedures having been followed.

From a pool of 1,187 interviews, Lamb et al. selected 98 for which independent validating information was available and the alleged victims were between 4 and 13 years of age. The interviews were rated by at least two native Hebrew speakers who had been trained to reliably employ the revised CBCA codes and were totally unfamiliar with the independent validating material. Meanwhile, other researchers used all investigative information other than the child’s statement to evaluate, on a 6-point scale, the likelihood that the events described by the child indeed occurred: Very Likely, Quite Likely, Questionable, Quite Unlikely, Very Unlikely, and No Judgment Possible.

Very few (13) of the cases were rated as either Quite Unlikely or Very Unlikely, but there was a significant association between the plausibility of the allegations and the total CBCA scores, with the highest scores assigned to statements about events deemed Very or Quite Likely to have occurred on the basis of independent case facts. Five of the criteria were significantly more likely to be present in Plausible than in Implausible accounts.

Although the CBCA scores indeed differentiated significantly between more and less credible accounts in this study, the differences between them were much less dramatic than reported by Boychuk (1991) or Raskin and Esplin (1991a, 1991b). The results of another validation study in the United States, using CBCA rating procedures similar to Lamb et al.’s, yielded surprisingly similar findings (Craig et al., 1999).

In light of Raskin and Esplin’s (1991c) insistence on the need for high-quality interviews to be available whenever credibility was assessed,
Hershkowitz, Lamb, Sternberg, and Esplin (1997) then sought to determine whether there was a relationship between the interview characteristics empirically associated with the elicitation of greater amounts of information and the presence of the CBCA criteria that ostensibly index credibility. As noted earlier, open-ended invitations elicit longer (more words) and richer (more details) responses than more focused utterances. The superiority and importance of invitational prompts might be further underscored if it could be shown that these prompts were also more likely to elicit responses that contain CBCA criteria. Accordingly, Hershkowitz et al. asked not simply whether the criteria were more likely to be present in truthful accounts, but examined the association between the presence of these criteria and features of the interview process, particularly the types of interviewer utterances known to elicit greater amounts of information from children. As expected, invitations evoked responses containing significantly more details and CBCA criteria than did all the other types of utterances, including facilitators, and the number of criteria present was significantly correlated with the number of words spoken by and the number of details provided by the child.

Because the presence of CBCA criteria in children's statements is believed to reflect greater reliance on the recall memory of events that actually occurred, the results thus underscored the importance of open-ended questions for facilitating access, via free-recall memory, to narrative accounts of events. In the absence of such narratives, it is difficult to evaluate the credibility of children's statements and thus high quality interviews that tap recall memory are essential for evaluations of credibility. The results of this study, then, support previous recommendations that interviewers use open-ended strategies whenever possible and also amplify the arguments by proponents of the CBCA system that criteria are more likely to be present when the material is obtained using open-ended interview procedures (e.g., Raskin & Esplin, 1991a; Raskin & Yuille, 1989). As noted, only the NICHD Protocol has been shown to increase the use of open-ended questioning procedures in forensic contexts.

In summary:

- Few attempts have been made to validate credibility assessment procedures in the field.
- Although the CBCA reliably discriminates between true and false statements, its precision is not good enough for forensic application.
- Credibility assessment is easier and better when interviews are well-conducted and children are encouraged to provide narrative accounts.
Guidelines

Considerations and Cautions

The development of interview protocols has certainly helped forensic interviewers to conduct developmentally appropriate interviews with children. Whereas the theoretical knowledge acquired in most training does not get translated into improved practices, only use of the structured NICHD protocol, which operationalizes guidelines and recommended practices, has repeatedly been shown to improve the quality of interviewing. It is crucially important to protect the integrity of interviews with child witnesses because alternative or corroborating evidence is seldom available to investigators and triers-of-fact. As highlighted throughout this book, forensic interviews with children represent a two-way street: we must be aware what both interviewers and children bring to the interview context. Interviewing techniques that may compromise the actual or perceived reliability of children's reports make it less likely that children will be adequately protected and increase the likelihood that false allegations will be pursued.

We have given particular emphasis in this chapter to research evaluating the NICHD interview protocol but, of course, the protocol is not a panacea. It emphasizes techniques that help children to report information about experienced events that is more likely to be accurate but it does not really address motivational factors that make some children reluctant to disclose abuse (Pipe, Lamb, Orbach, & Cederborg, 2007). Even experienced interviewers guided by the protocol have difficulty conducting forensic interviews well unless they continue to review their interviews with other experienced interviewers even after they have been trained.

The following considerations and cautions are therefore proposed:

- Before beginning an interview, careful planning and preparation is necessary with respect to: (a) developmentally appropriate language, (b) the timing and duration of the interview session, (c) the types of prompts that might be used to identify the target event and to elicit relevant details (including those for particularization), and (d) techniques and topics on which to focus when establishing rapport.
- At the outset of the interview, it is important to establish ground rules and ensure children both understand and have practiced them (e.g., correcting the interviewer, responding "I don't know" or "I don't understand").
- Children should have an opportunity to practice recalling a recent past event in detail so that they also become familiar with the types of prompts that will be used in the substantive portion of the interview.
- It is difficult for interviewers to maintain high standards. Regular supervision, evaluation of interviews (including self-evaluation), supportive and constructive feedback about past interviews, and discussion of techniques to enhance the quality of future interviews are essential. Agencies also need to recognize the need and provide opportunities for interviewers to attend conferences and workshops that review empirically based advances, and to have access to current literature and emerging research so as to continue their professional development.
- Any single technique will not be suitable for all children or contexts. Interviewers need to recognize the challenges posed by reluctant and mentally impaired witnesses, as well as the potential for interviewers to contaminate the quality of the child's testimony or elicit false reports.
REFERENCES


