The Effects of Intensive Training and Ongoing Supervision on the Quality of Investigative Interviews With Alleged Sex Abuse Victims

Michael E. Lamb, Kathleen J. Sternberg, and Yael Orbach
National Institute of Child Health and Human Development

Irit Hershkowitz
University of Haifa, Israel

Dvora Horowitz
Israeli Ministry of Labor and Social Affairs

Phillip W. Esplin
Private Practice, Phoenix, AZ

Four distinct strategies were employed to train 21 experienced forensic interviewers to interview alleged sex abuse victims (M = 9.20 years of age) in accordance with professionally recommended practices. The structure and informativeness of the 96 interviews they conducted following training were compared with the structure and informativeness of 96 matched interviews conducted by the same interviewers in the 6 months prior to the training. Didactic workshops and instruction in the utilization of highly structured presubstantive interview procedures had little effect on the number of open-ended prompts used to elicit information or on the amount of substantive information elicited in this way. By contrast, intensive training in the use of a highly structured interview protocol, followed by continuing supervision in the form of monthly day-long seminars, supplemented in some cases by detailed individual feedback on recent interviews, yielded dramatic improvements on these measures of interview quality.

Because alleged victims are often the only available sources of information about their abusive experiences, considerable efforts have been made to understand how children's testimony can be made as useful and reliable as possible. The research has been fruitful, and has resulted in surprisingly broad international consensus regarding optimal interview practices. Unfortunately, agreement regarding the ways in which interviews should be conducted has not been paralleled by changes in the way interviews are actually conducted in the field, and several researchers have shown how difficult it is to effect lasting changes through training. Building on recent demonstrations that intensive training and continuing supervision can change interviewer behavior, the goal of this study was to examine the relative effectiveness of several popular ways of training forensic interviewers to follow "best practice" guidelines when conducting investigative interviews in the field.

Expert professional groups agree that children should be interviewed as soon as possible after the alleged offenses by interviewers who themselves introduce as little information as possible, while encouraging children to provide as much information as possible in the form of narratives elicited using open-ended prompts (American Professional Society on the Abuse of Children, 1990; Bull, 1992, 1995; Jones, 1992; Lamb, 1994; Lamb, Sternberg, & Esplin, 1998; Lamb, Sternberg, Hershkowitz, Orbach, & Esplin, 1999; Memorandum of Good Practice, 1992; Poole & Lamb, 1998; Raskin & Esplin, 1991; Warren & McGough, 1996; Yuille, Hunter, Joffe, & Zaparniuk, 1993). Before substantive issues are discussed, interviewers are typically urged to explain their roles, the purpose of the interview, and the "ground rules" (for example, to limit themselves to descriptions of events "that really happened" to them and to correct the interviewer, request explanations or clarification, and acknowledge ignorance, as necessary). Because recognition prompts are more likely than open-ended questions to elicit erroneous information, investigators are consistently urged to use these

prompts only when needed to elicit forensically relevant information, and as late in the interview as possible.

The universal emphasis on the value of narrative responses elicited using open-ended prompts is rooted in the oft-replicated results of laboratory analog studies demonstrating that information elicited using such prompts is much more likely to be accurate than information elicited using more focused prompts (Dale, Loftus, & Rathbun, 1978; Dent, 1982, 1986; Dent & Stephenson, 1979; Goodman & Aman, 1990; Goodman, Bottoms, Schwartz-Kenney, & Rudy, 1991; Hutcheson, Baxter, Telfer, & Warden, 1995; Oates & Shrimpton, 1991; Ornstein, Gordon, & Larus, 1992) probably because open-ended questions force the respondent to recall information from memory, whereas more focused prompts often require the respondent to recognize one or more options suggested by the interviewer. Accuracy is much more difficult to establish in the field than in laboratory analog contexts, of course, because forensic interviewers seldom know what really happened, but the results of field studies in which accuracy was assessed confirm that, as in the laboratory, responses to open-ended questions posed by forensic interviewers are more likely to be accurate than responses to more focused prompts (Lamb & Fauchier, 2001; Orbach & Lamb, 1999, 2001). Moreover, these field studies have demonstrated that, as in the laboratory, erroneous information is more likely to be elicited by recognition memory prompts (Lamb & Fauchier, 2001; Orbach & Lamb, 2001).

Unfortunately, these research-based recommendations are widely endorsed but are seldom followed. Studies of forensic interviews in various parts of the United States, United Kingdom, Sweden, and Israel consistently show that forensic interviewers use open-ended prompts quite rarely, even though such prompts consistently elicit more information than more focused prompts do (e.g., Cederborg, Orbach, Sternberg, & Lamb, 2000; Craig, Scheibe, Raskin, Kircher, & Dodd, 1999; Davies, Westcott, & Horan, 2000; Lamb, Hershkowitz, Sternberg, Esplin, et al., 1996; Lamb, Hershkowitz, Sternberg, Boat, et al., 1996; Lamb et al., 2000; Sternberg, Lamb, Davies, & Westcott, 2001; Sternberg, Hershkowitz, Esplin, et al. 1996; Walker & Hunt, 1998; Walker & Warren, 1998).

Such findings are consistent with the results of studies showing that both intensive and less intensive training programs for investigative interviewers may impart knowledge about desirable practices but have little, if any, effect on the actual behavior of forensic investigators (Aldridge & Cameron, 1999; Freeman & Morris, 1999; Stevenson, Leung, & Cheung, 1992; Warren et al., 1999). More recently, however, researchers at the National Institute of Child Health and Human Development (NICHD) have shown that it is possible to effect major improvements in the quality of forensic interviewing when interviewers are trained to follow a very

detailed and specific interview protocol and are given continuing supervision and feedback on simulated and actual forensic interviews (Orbach et al., 2000; Sternberg, Lamb, Orbach, Esplin, & Mitchell, 2001).

The incremental value of verbal and written feedback during the course of training has been experimentally demonstrated in individual (Adams, Fields, & Verhave, 1999; Clark, 1971; Frayer & Klausmeier, 1971; Sweet, 1966) and group (Gully, 1998) contexts, but only the NICHD training model includes feedback beyond the training period (i.e., in post-training investigative interviews as well). Because the success of their efforts contrasted with the failures of those whose efforts were limited to intensive but time-limited training seminars, Orbach et al. (2000) and Sternberg, Lamb, Orbach, et al. (2001) suggested that ongoing supervision and feedback might be crucial. In a recent study, furthermore, Lamb and his colleagues (2002) demonstrated the adverse effects of the termination of supervision and feedback on investigators' performance. Forensic interviews conducted by trained investigative 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

1. Declines in the use of open-ended prompts.
2. Corresponding increases in reliance on more focused prompts.
3. 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.

The importance of continuing quality control and feedback was assessed in this study by comparing the effectiveness of four different training models designed to help interviewers implement recommended interviewing practices. In all training 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 memorial, linguistic, communicative, and social development. The first training condition involved such conceptual training, without any concrete guidelines or supervised practice. Interviewers in the second training condition were not only introduced to scientific principles, but were also urged to employ structured modules in the presubstantive rapport building phase of their investigative interviews, and they practiced using these modules. Interviewers in the third and fourth training conditions were introduced to the scientific principles and were also given copies of the fully

structured NICHD interview protocol and practiced using it under close supervision.

The third and fourth conditions differed with respect to the amount and type of supervision provided and the types of investigative interviews for which the interviewers received specialized training. Interviewers in the third condition were trained to interview alleged victims in intensive courses, followed by monthly day-long group meetings in which their actual field interviews were analyzed (using video recordings and transcripts of their recent interviews) and desirable and undesirable practices were discussed. These monthly meetings continued until the end of the study, during which time these interviewers also received detailed written and verbal feedback on each of their forensic interviews. By contrast, interviewers in the fourth condition were trained to employ a structured protocol when interviewing juvenile suspects. They participated in the monthly meetings alongside those in the third condition but received no individual supervision and feedback on any of their 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 or questions, 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 described here 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.

The results of earlier studies (Aldridge & Cameron, 1999; Sternberg et al., 1997; Warren et al., 1999) led us to predict that the first two training conditions would have little apparent effect on the interviewers' behavior. As suggested by Orbach et al. (2000) and by Sternberg, Lamb, Orbach, et al. (2001), we hypothesized that the third and fourth training conditions, which provided interviewers with continuing supervision and instruction, would have the greatest effects on both the interviewers' practices and the quality of information provided by children. Interviewers in the third condition were expected to perform best because, in addition to the conceptual training, they benefitted from both continuing group training and supervision as well as detailed individualized feedback on many of their field interviews, whereas those in the fourth group only observed this feedback being provided to their peers.

Method

In this study, we examined 192 forensic interviews of alleged sexual abuse victims by 21 experienced youth investigators (15 women and 6 men) working in all regions of the state of Israel, where youth investigators are the only persons authorized to conduct forensic interviews of children and adolescents suspected of being victims, witnesses, or perpetrators of criminal acts (see Sternberg, Lamb, & Hershkowitz, 1996, for a description of the Israeli system). All were the first interviews of these children, conducted by youth investigators immediately following a formal report of the abuse. All of the interviews and interviewees were native Hebrew speakers who conversed in Hebrew. All of the training and supervision, except for some portions of the intensive training in Conditions I and II and some of the written materials provided to interviewers in all conditions, were provided in Hebrew as well.

The 54 boys and 138 girls interviewed averaged 9.20 (SD = 2.22) years of age (range = 4.0 to 14.25 years). The average ages of children in the 8 cells or groups described later ranged from 8.5 years to 10.1 years, but there were no significant differences among the ages of children in the different groups. Ninety-five children were under and 97 were over 9.25 years of age, so children in these two age groups were contrasted in analyses concerned with differences between older and younger children.

The Training Conditions

Condition 1: Validation condition. Training for this condition involved a week-long session taught by a multinational team of forensic and developmental psychologists and social workers. In this session, instructors explained the developmental factors and circumstances associated with variation in children's abilities to describe their experiences, with emphasis on the ways in which interviewers crucially affect the quality and richness of children's accounts. Videotaped examples of desirable interview practices were shown for illustrative purposes. Because the interviewers had agreed to assist in a field validation study of children's credibility analysis, furthermore, the literature on this topic was also explained. Further details about this condition were provided by Lamb et al. (1997) in their report of the field validation study.

Condition 2: Rapport building condition. Training specific to this condition (provided in addition to the Condition 1 training) involved a 2-day session focused on the importance of structuring the investigative interview carefully and of motivating child witnesses to be informative. Participants were given two structured rapport-building modules to use as the presubstantive phase of their investigative interviews. One version introduced children to open-ended interview prompts while the other introduced them to focused prompts. Both modules took an average of 7 min to implement,

and interviewers were instructed to alternate between the two modules in their investigative interviews. Only interviewers who faithfully employed the two modules were included in our study (see Sternberg et al., 1997, for further details).

Condition 3: Victims' protocol. Before the fully structured NICHD interview protocol (Orbach et al., 2000) was implemented, all interviewers in this condition participated in an intensive 2-day training seminar during which the conceptual and empirical rationale for all phases of the interview were explained by a team of forensic and developmental psychologists and social workers. Both appropriate and inappropriate interview techniques were illustrated and discussed. Trainees also conducted simulated interviews using the protocol which were then discussed with other trainees and the instructors.

The goal of the structured protocol was to operationalize the recommendations of professional advisory groups regarding optimal interviewing techniques and to maximize the interviewers' adherence to these procedures. The investigative strategies reflected in the structured protocol thus gave priority to open-ended probes and retrieval cues, and delayed focused questions to later stages of the interview. The protocol was structured to encourage children to provide as much information as possible from free-recall and to report event-specific rather than generic information. After learning to use the protocol, interviewers were observed conducting simulation and field interviews using the protocol and were given feedback on their techniques. Written feedback was provided on transcripts of field interviews until the study ended. In addition, individual and group training sessions focused on adherence to the protocol and its adaptation to individual circumstances were conducted every 4 weeks by the social worker involved in the initial training. Problematic cases were reviewed with the group and techniques for addressing difficult issues were discussed. Further details about this training and protocol were provided by Orbach et al. (2000).

Condition 4: Suspects' protocol. Interviewers in this condition were initially trained to interview alleged juvenile perpetrators of sexual abuse, using a fully structured protocol. Like the NICHD victim interview protocol, the suspect protocol emphasized the value of recall memory retrieval, but coercive strategies were included in the suspect protocol because many of the alleged suspects were less motivated than the alleged victims to be truthful. The interviewers in Condition 4 attended the monthly sessions alongside interviewers in the victims' protocol condition in which group level feedback on victims' and suspects' inter-

view strategies were provided, and exemplary interviews were analyzed and discussed. The interviewers received no individual supervision and feedback on their own field interviews with either suspects or child victims, however. Fewer victim interviews were conducted by interviewers in this group because they were also required to interview alleged suspects. Only their interviews of alleged victims were considered for inclusion in this study.

Research Design

The 192 interviews were distributed across 8 cells with between 17 and 30 interviews per cell. Thus, 25 interviews (validation condition) were conducted by 6 experienced interviewers. The structure and yield of these interviews were compared with those of 25 matched interviews (validation baseline) conducted within the previous 6 months by the same interviewers. An additional 24 interviews (rapport building condition) were conducted by 5 experienced investigators who employed either of two tightly scripted introductory and rapport-building modules (Sternberg et al., 1997). The structure and yield of these interviews were compared with those of a matched group of 24 interviews (rapport-building baseline) conducted by the same 5 interviewers in the 6 months preceding the introduction of the scripted presubstantive modules. Another 30 interviews were conducted by 5 experienced investigators who had been trained to employ the fully structured investigative protocol (victims' protocol condition). The structure and yield of these interviews was compared with those of 30 matched interviews conducted by the same interviewers in the 6 months preceding introduction of the structured protocol (victims' protocol baseline). Lastly, 17 interviews were conducted by 5 experienced interviewers who received copies of the investigative protocol employed in the victims' protocol condition as well as a protocol developed for interviews of young suspects (suspects' protocol condition). The structure and yield of these interviews were compared with those of 17 matched interviews (suspects' protocol baseline) conducted by the same interviewers in the 6 months preceding onset of the suspects' protocol condition interview training.

For purposes of comparison, interviews in the training conditions were matched as closely as possible with interviews in specific comparison, or 'baseline' groups. All interviewers were included in both a training condition and its parallel baseline condition. All interviewers participating in the different training conditions were included in the study. To avoid confounding interviewer effects, however, no interviewer was included in more than one training condition. This precaution limited the number of interviewers who could be studied. Interviews in the baseline conditions

were selected from among all sexual abuse investigations conducted by the same interviewers prior to the respective training periods solely on the basis of their comparability with interviews in the training conditions. Unmatchable interviews in the training conditions were excluded from the study. Matching variables used to select interviews conducted by the same interviewers in the 6 months preceding the onset of training in each of the training conditions included the following: the age of the child (within 6 months); the number of alleged incidents experienced (103 alleged one incident and 89 alleged more than one incident); the relationship between the alleged perpetrator and the victim (the perpetrator was a familiar person in 112 of the cases and a stranger in 80 cases); the type of abuse (penetration [$n = 46$], touching over [$n = 43$] or under [$n = 78$] the clothes, or exposure [$n = 25$]); and gender of the alleged victim ($n = 138$ female). All of these factors are either known or believed to affect interview processes. None of the alleged perpetrators were members of the victims' families.

All the interviews were transcribed, checked for accuracy, and checked to ensure that all personal identifiers were deleted before transcripts were sent to the researchers.

Procedure

One of four trained Hebrew-speaking raters independently reviewed each of the transcripts, categorizing each utterance made by the interviewer during the substantive portion of the interview, operationally defined as the portion of the interview during which the incidents under investigation were discussed. Five categories introduced by Lamb and his colleagues (Lamb, Hershkowitz, Sternberg, Esplin, et al., 1996; Lamb, Hershkowitz, Sternberg, Boat et al., 1996) were used to characterize all interviewer utterances in the substantive portions of the interviews: facilitators, invitations, directive, option-posing, and suggestive.

1. Facilitators. Non-suggestive encouragement to continue with a response to the previous utterance. These include utterances like "O.K.," restatements of the child's previous utterance, and non-suggestive words of encouragement designed to prompt continuation of the child's narrative. Because Hershkowitz (in press) has recently shown that facilitators merely amplify the preceding interviewer prompt, they are not analyzed as independent utterances. Details provided following facilitators were attributed to the preceding prompt.

2. Invitations. Utterances, including questions, statements, or imperatives, prompting free-recall responses from the child. Such utterances do not delimit the child's focus except in a general way (for example, "Tell me everything that happened"), or use details disclosed by the child as cues (for example, "You mentioned that he touched you. Tell me everything about the touching.").

3. Directive utterances. These refocus the child's attention on details or aspects of the alleged incident that the child has already mentioned, providing a category for requesting additional information using "Wh-" questions (cued recall). For example, "When did it happen?" (When the child disclosed that something happened), or "What color was his t-shirt?" (When the child said he was wearing a t-shirt).

4. Option-posing utterances. These focus the child's attention on details or aspects of the alleged incident that the child has not previously mentioned, asking the child to affirm, negate, or select options provided by the investigator using recognition memory processes, but do not imply that a particular response is expected. For example, "Did he say anything to you?", or "Did he touch you over or under your clothes?" (When the child mentioned that he touched him/her).

5. Suggestive utterances. These are stated in such a way that the interviewer strongly communicates what response is expected (e.g., "He forced you to do that, didn't he?") or they assume details that have not been revealed by the child (for example: Child: "We laid on the sofa." Interviewer: "He laid on you or you laid on him?")

When a single turn in the dialogue included two or more statements or questions that could be coded differently, the highest category defined by the numerical label in the list above was applied. Coders then employed a technique introduced by Yuille and Cutshall (1986) and elaborated by Lamb, Hershkowitz, Sternberg, Esplin, et al. (1996) to tabulate the number of new details conveyed by the child. By definition, details involved the identification of individuals, objects, and events, and descriptions of their features (e.g., appearance, action, location). Details were only counted when they added to understanding of the target incidents, so restatements of facts were not counted. Details provided following facilitators were attributed to the preceding substantive utterance (invitation, directive, option-posing, or suggestive). The number of facilitators was not tabulated.

Interrater reliability. All rating of the utterance types were conducted by one of four coders who trained on an independent set of transcripts until they agreed with one another concerning the classification of at least 85% of the utterance types and details. During the course of rating, 23% of the transcripts were independently coded by two or more of the raters to ensure that they remained equivalently reliable. In these assessments, raters agreed regarding the classification of 87% of the interviewer utterances and 88% of the details provided by the children.

Results

Analyses of Interviewer Utterance Types

Several 4 (Condition: Validation, Rapport building, Victims' protocol, Suspects' protocol) × 2 (Stage: Training, Baseline) × 2 (Age: Under 9.25 years, Over 9.25 years) Multivariate Analyses of Variance (MANOVAs) with the absolute and proportional numbers of invitations, directives, option-posing, and suggestive interviewer prompts as dependent variables were conducted to assess the primary sources of variation in the interviewers' behavior. These analyses revealed significant effects for condition, $F(12, 458) = 7.810$, $p < .001$ for absolute numbers, and $F(9, 424) = 14.905$, $p < .001$ for proportions; stage, $F(3, 174) = 14.934$, $p < .001$; $F(3, 174) = 39.025$, $p < .001$; and age, $F(3, 174) = 3.429$, $p < .01$; $F(4, 173) = 5.492$, $p < .001$, as well as significant two-way Condition × Stage interactions, $F(12, 458) = 5.176$, $p < .001$; $F(9, 424) = 12.302$, $p < .001$. The 3-way interaction among Condition, Stage, and Age was significant for proportion, $F(9, 424) = 2.166$, $p < .023$, but not for absolute numbers of utterances.

Subsequent univariate analyses revealed significant effects for condition on the numbers and proportions of invitations, $F(3, 176) = 11.593$, $p < .001$ for number, and $F(3, 176) = 38.232$, $p < .001$ for proportion; directives, $F(3, 176) = 5.866$, $p < .001$ for number, and $F(3, 176) = 4.693$, $p = .004$ for proportion; and option-posing, $F(3, 176) = 10.488$, $p < .001$ for number, and $F(3, 176) = 24.391$, $p < .001$ for proportion; but no significant effects on the number and proportion of suggestive prompts. Scheffé tests showed that invitations were both absolutely and proportionately more frequent ($p < .001$) and option-posing prompts both absolutely and proportionately less frequent in the victims' ($p < .001$) and suspects' ($p < .05$) protocol conditions than in the validation and the rapport building conditions (see Table 1). Interviewers in the validation condition used absolutely and proportionally more directive prompts and proportionally fewer option-posing prompts than they did in the rapport building condition ($ps < .05$), as well as absolutely ($p < .001$) but not proportionally more directive prompts than in the victims' protocol conditions. There were no significant differences between the suspects' and victims' protocol conditions.

Significant effects for stage revealed that interviewers in the training conditions used absolutely and proportionally more invitations, $F(1, 176) = 36.346$, $p < .001$ for number, and $F(1, 176) = 115.916$, $p < .001$ for proportion; fewer directives, $F(1, 176) = 7.797$, $p = .006$ for number, and $F(1, 176) = 17.470$, $p < .001$ for proportion; and fewer option-posing prompts, $F(1, 176) = 4.893$, $p = .028$ for number, and $F(1, 176) = 8.502$, $p = .004$ for proportion; than interviewers in the baseline conditions (see Table 1). Interviewers in the training stage also tended to use nonsignificantly fewer suggestive utterances than interviewers in the baseline conditions, $F(1, 176) = 3.679$, $p = .057$ for number.

There were significant univariate effects for age only in the number and proportion of suggestive utterances, $F(1, 176) = 10.430$, $p < .001$ for number, and $F(1, 176) = 15.608$, $p < .001$ for proportion; with younger children asked more suggestive questions than older children, and there was a nonsignificant tendency for older children to be asked proportionally but not absolutely more directive questions, $F(1, 176) = 3.64$, $p = .058$.

Univariate tests of the Condition × Stage interaction revealed significant effects on the number and proportion of invitations, $F(3, 176) = 11.055$, $p < .001$ for number, and $F(3, 176) = 40.345$, $p < .0001$ for proportion; and option-posing prompts, $F(3, 176) = 5.698$, $p < .001$ for number, and $F(3, 176) = 10.429$, $p < .0001$ for proportion; and on the proportion of directive prompts, $F(3, 176) = 3.029$, $p < .031$.

Post hoc analyses conducted to clarify these interactions revealed, as expected, that the multivariate effect for condition was not significant when data from the baseline stage only were analyzed, but was highly significant when data from the training stage were analyzed separately. Following training, interviewers in the victims' protocol and suspects' protocol conditions used more invitations, proportionally fewer directive ($ps < .001$), and fewer option-posing utterances ($ps < .05$) than did interviewers in the validation and the rapport-building conditions.

The multivariate Condition × Stage × Age interaction in the proportion of investigative utterances of each type was clarified by significant univariate interaction effects for invitations, $F(3, 176) = 4.382$, $p < .006$. Whereas there was no Age × Condition interaction prior to training, following the training younger children were more likely than older children to be asked more invitations in the victims' and suspects' protocol conditions than in the validation and rapport building conditions ($ps < .001$).

Analyses of the Children's Responses

The total numbers of forensically relevant details obtained in each of the study conditions are displayed in Table 2. Several 4 (Condition: Validation, Rapport building, Victims' protocol, Suspects' protocol) × 2 (Stage: Training, Baseline) × 2 (Age: Younger, Older) MANOVAs with the absolute number and proportion of all details elicited in response to invitations, directives, option-posing prompts, and suggestive prompts as dependent variables were conducted to assess the primary sources of variation in the children's informa-

119

Table 1. *Condition and Stage-Based Differences in the Types of Utterances Used to Elicit Information From the Children*

Utterance Types	Pre-training								Training							
	Validation		Rapport		Victims'		Suspects'		Validation		Rapport		Victims'		Suspects'	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Invitation	5.20	3.72	5.29	3.38	4.50	8.02	7.00	5.12	5.76	6.00	5.75	4.95	18.87	11.04	19.59	14.27
Directive	56.68	31.33	35.83	16.95	41.30	28.49	55.29	17.50	50.20	32.09	37.25	27.93	27.53	17.21	29.41	19.20
Option-posing	29.80	15.09	27.79	13.97	24.27	18.46	26.76	13.63	30.64	21.33	32.96	22.71	8.13	4.78	12.65	12.21
Suggestive	9.48	8.73	6.38	5.29	7.00	6.38	9.47	8.30	6.08	4.71	7.58	6.45	3.93	3.65	7.41	12.32

Table 2. *Total Numbers of Details Provided by Children in the Four Experimental Conditions*

Stage	Validation		Rapport building		Victims' Protocol		Suspects' Protocol	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pretraining	249.96	126.96	214.75	109.23	289.47	191.62	275.76	114.79
Training	296.52	187.32	297.08	235.64	267.57	134.77	252.82	180.21

tiveness. The analyses revealed significant effects for condition, $F(12, 458) = 5.825$, $p < .001$; $F(12, 458) = 8.686$, $p < .001$ for absolute numbers and proportions, respectively; stage, $F(4, 173) = 11.142$, $p < .001$ for number, and $F(4, 173) = 13.902$, $p < .001$ for proportion; and age, $F(4, 173) = 4.773$, $p < .001$ for number, and $F(4, 173) = 2.445$, $p = .048$ for proportion; as well as significant Condition × Stage interactions, $F(12, 458) = 4.408$, $p < .001$ for number, and $F(12, 458) = 5.560$, $p < .001$ for proportion; and a Condition × Age interaction that was significant for the numbers, $F(12, 458) = 2.281$, $p < .008$, but not proportion of details.

Subsequent univariate analyses revealed condition effects, apparent in Table 3, with respect to the number and proportion of details elicited using invitations, $F(3, 176) = 5.784$, $p < .001$, and $F(3, 176) = 16.304$, $p < .001$, respectively; and option-posing prompts, $F(3, 176) = 10.069$, $p < .001$, and $F(3, 176) = 24.771$, $p < .001$, respectively; as well as a significant effect on the proportion of details elicited using directive prompts, $F(3, 176) = 6.411$, $p < .001$. Scheffé tests showed that both absolutely ($p < .01$) and proportionally ($p < .001$) more details were elicited using invitations in the victims' and suspects' protocol conditions than in the validation condition, whereas proportionally more details were elicited using invitations in the victims' protocol condition than in the validation and rapport-building conditions ($p < .001$). Absolutely and proportionally fewer details were elicited using option-posing prompts in the victims' protocol condition than in the validation ($p < .05$) and rapport-building ($p < .001$) conditions. Proportionally more details were elicited using directives in the validation than in the victims' protocol and rapport-building conditions.

Significant effects for stage on the number and proportion of details elicited using invitations, $F(1, 176) = 34.230$, $p < .001$ for number, and $F(1, 176) = 54.581$, $p < .001$ for proportion; and directive prompts, $F(1, 176) = 4.906$, $p < .028$ for number, and $F(12, 176) = 25.533$, $p < .001$ for proportion; as well as a significant effect on the proportion of details elicited using option-posing prompts, $F(1, 176) = 6.124$, $p < .014$, revealed that invitations were more useful and both directive and option-posing prompts less useful for eliciting details after training than in the baseline conditions (see Table 3).

Significant effects for age on the number of details elicited using invitations, $F(1, 176) = 9.571$, $p < .002$; directive, $F(1, 176) = 12.435$, $p < .001$; and option-posing prompts, $F(1, 176) = 4.220$, $p < .041$; and on the

proportion of details elicited using suggestive prompts, $F(1, 176) = 5.391$, $p < .021$; indicated that all types of prompts elicited more details from older than from younger children.

Univariate Condition × Stage interactions were evident in the number and proportion of details elicited using invitations, $F(3, 176) = 4.434$, $p < .005$ for number, and $F(3, 176) = 15.528$, $p < .001$ for proportion; directive prompts, $F(3, 176) = 4.017$, $p < .009$ for number, and $F(3, 176) = 2.954$, $p < .034$ for proportion; option-posing utterances, $F(3, 176) = 10.059$, $p < .001$ for number, and $F(3, 176) = 9.650$, $p < .001$ for proportion; and suggestive prompts, $F(3, 176) = 4.416$, $p < .005$ for number, and $F(3, 176) = 3.008$, $p < .032$ for proportion. As Table 3 makes clear, more details were elicited using invitations, and fewer were elicited using directive, option-posing, and suggestive prompts in the victims' and suspects' protocol conditions after training than before. Whereas the rapport-building and validation training conditions had no effect on the amount of information elicited using invitations and directive prompts, more details were elicited using option-posing and suggestive prompts after training than in the matched baseline conditions (ps < .05).

Univariate Condition × Age interactions were evident in the numbers of details elicited using invitations, $F(3, 176) = 3.603$, $p < .015$; and option-posing prompts, $F(3, 176) = 4.204$, $p < .007$. Invitations elicited more details from older than from younger children in the rapport building and suspects' protocol conditions. Option-posing prompts elicited more details from older than from younger children in the validation and rapport-building condition, whereas the reverse was true in the victims' protocol condition ($p < .05$).

Number of Details Before First Option-Posing Prompt

Another important index of interview quality is the number of details provided by the interviewee before the interviewer introduces any information. Condition (4) × Stage (2) × Age (2) ANOVAs revealed significant effects for stage, $F(1, 176) = 22.630$, $p < .001$ for number, and $F(1, 176) = 14.864$, $p < .001$ for proportion; and age, $F(1, 176) = 14.872$, $p < .001$ for number, and $F(1, 176) = 10.269$, $p < .001$ for proportion; in the number and proportion of details provided before the first option-posing or suggestive prompt along with a

Table 3. *Condition and Stage-Based Differences in the Number of Details Elicited From the Children Using the Four Types of Prompts*

	Pretraining								Training							
	Validation		Rapport		Victims'		Suspects'		Validation		Rapport		Victims'		Suspects'	
Utterance Types	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Invitation	39.28	38.02	49.21	46.80	51.37	81.11	40.53	30.76	46.24	44.52	84.25	104.10	142.23	73.06	162.00	159.89
Directive	138.16	103.08	95.71	62.00	144.67	120.15	153.41	93.36	144.00	97.39	92.50	67.20	89.90	70.35	66.41	46.03
Option-posing	54.36	28.09	55.37	35.46	55.87	54.67	57.82	29.20	76.24	67.74	95.50	90.29	19.83	20.11	13.76	14.29
Suggestive	18.16	14.99	14.46	13.82	37.57	41.48	24.00	26.88	30.04	49.11	24.83	28.26	15.60	21.28	10.65	13.62

nonsignificant trend for condition with respect to the number (not proportion) of details, $F(3, 176) = 2.156$, $p < .095$; and a significant 3-way interaction, $F(3, 176) = 5.361$, $p < .001$; $F(3, 176) = 4.718$, $p < .003$, for number and proportion, respectively. Interviewers in all the training conditions elicited more information before their first intrusive prompt than did those in the baseline conditions, while interviewers allowed older children to provide more information before intruding than did those interviewing younger children. The significant interaction reflected the absence of differences with respect to condition prior to training, whereas, after training, interviewers in the victims' protocol condition were especially successful when interviewing younger children, while those in the suspect and to a lesser extent, rapport building condition, were especially successful at delaying their intrusions when interviewing older children.

Discussion

These results have important, although somewhat sobering, implications for those attempting to transfer information gleaned from basic research to those attempting to apply that information in the real world. Clearly, it is possible to employ our 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.

Because there is substantial agreement in the field regarding the practices that are most likely to elicit accurate information from alleged victims about their experiences, many professionals have developed intensive training programs in which these principles are explained, often in the context of role-play exercises designed to consolidate learning. Unfortunately, systematic evaluations of such training program consistently reveal effects on the trainees' knowledge but no demonstrable impact on the quality of their behavior when conducting interviews (Aldridge, 1992; Aldridge & Cameron, 1999; Freeman & Morris, 1999; Stevenson et al., 1992 ; Warren et al., 1999). Such reports are especially disconcerting because investigative interviews in the field are seldom conducted in accordance with best practice guidelines (Bruck, 1999; Cederborg et al., 2000; Craig et al., 1999; Davies et al., 2000; Lamb, Hershkowitz, Sternberg, Boat, et al., 1996; Lamb, Hershkowitz, Sternberg, Esplin, et al., 1996; Lamb et al., 2000; Sternberg et al., 2001; Sternberg, Lamb, Hershkowitz, Esplin, et al., 1996; Walker & Hunt, 1998).

Previous reports by both Orbach et al. (2000) and Sternberg, Lamb, Orbach, et al. (2001) showed that the quality of interviewing improved when forensic interviewers in Israel and the United States, respectively, were trained to implement a protocol that operationalized the consensus recommendations of diverse professionals and scholars, practiced using that protocol, and received written and verbal feedback on their interviews thereafter. In both studies, investigators following detailed interview guidelines relied much more heavily on open-ended rather than focused questions, thereby increasing the amount of information obtained in the form of recall narratives. Such information is preferable because it is more likely to be accurate (Dale et al., 1978; Dent, 1982, 1986; Dent & Stephenson, 1979; Goodman et al., 1991; Goodman & Aman, 1990; Hutcheson et al., 1995; Lamb & Fauchier, 2001; Oates & Shrimpton, 1991; Orbach & Lamb, 2001; Ornstein et al., 1992), which is why all professional and expert groups recommend that as much information as possible be obtained using open-ended prompts (see Lamb et al., 1999, and Poole & Lamb, 1998, for reviews). Both Orbach et al. (2000) and Sternberg, Lamb, Orbach, et al. (2001) argued that continued supervision played a central role in ensuring that gains obtained through implementation of the structured protocol were maintained over time.

The results of this study are largely consistent with this: Significant differences between the baseline and training conditions we studied were largely accounted for by differences in the performance of interviewers in the victims' and suspects' protocol training conditions. In these conditions, the interviewers were guided by highly structured investigative interview protocols and continued to attend regular intensive training workshops. The effects were most clearly marked by improvements in the extent to which interviewers tried to elicit information using open-ended prompts, in the amount of information actually elicited from the children's free recall, and in the extent to which the interviewers were able to delay their first option-posing questions which, by definition, involved the introduction of information by the interviewers rather than by the children. By contrast, interviewers who received intensive short-term training but no continued training generally performed little better than they had before training.

It was noteworthy, however, that interviewers in the suspects' protocol condition performed at least as well as their peers in the victims' protocol condition on the measures of interview quality. Interviewers in both of these groups employed a highly structured interview protocol and attended intensive day long workshops every month, but those in the victims' protocol condition also received individual feedback on most of their interviews. The fact that this continuing individualized feedback did not bring about greater improvements in their performance (relative to peers in the suspects' protocol condition) suggests that this especially tedious and costly form of extended supervision may not be necessary to bring about and maintain improve-

ments in interview quality. Of course, interviewers in both conditions did benefit from detailed examination of interviews conducted recently by the seminar participants, and this close study of successful and unsuccessful strategies employed by their peers in interviews like those they were themselves conducting was obviously beneficial. Relatively few interviews were included in the suspect's protocol condition, furthermore, so additional research on the value of individual feedback is needed.

Overall, the results of this study strongly suggest that meaningful long-term improvement in the quality of information obtained from young alleged victims of sexual abuse are observed only when (a) well-established principles are operationalized in a clear and concrete fashion and (b) training is distributed over time, rather than provided in the form of a single initial session, however intensive. It is, of course, costly to continue providing intensive support and training to interviewers, but researchers have yet to identify any less costly techniques that are equivalently effective and we have shown that the termination of continuing supervision is associated with rapid declines in the quality of forensic interviewing (Lamb et al., 2002).

References

Adams, B. J., Field, L., & Verhave, T. (1999). Effects of unreinforced conditional selection training, multiple negative comparison training, and feedback on equivalence class formation. *Psychological Record, 49,* 685–702.

Aldridge, J. (1992). The further training of professionals dealing with child witnesses. In H. Dent & R. Flin (Eds.), *Children as witnesses* (pp. 231–244). Chichester, England: Wiley.

Aldridge, J., & Cameron, S. (1999). Interviewing child witnesses: Questioning techniques and the role of training. *Applied Developmental Science, 3,* 136–147.

American Professional Society on the Abuse of Children. (1990). *Guidelines for psychosocial evaluation of suspected sexual abuse in young children*. Chicago: Author.

Bruck, M. (1999). A summary of an affidavit prepared for Commonwealth of Massachusetts v. Cheryl Amirault LeFave. *Applied Developmental Science, 3,* 110–127.

Bull, R. (1992). Obtaining evidence expertly: The reliability of interviews with child witnesses. *Expert Evidence, 1,* 5–12.

Bull, R. (1995). Innovative techniques for the questioning of child witnesses, especially those who are young and those with a learning disability. In M. S. Zaragoza, J. R. Graham, G. C. N. Hall, R. Hirschman, & Y. S. Ben-Porath (Eds.), *Memory and testimony in the child witness* (pp. 179–194). Thousand Oaks, CA: Sage.

Cederborg, A., Orbach, Y., Sternberg, K. J., & Lamb, M. E. (2000). Investigative interviews of child witnesses in Sweden. *Child Abuse and Neglect, 24,* 1355–1361.

Clark, D. C. (1971). Teaching concepts in the classroom: A set of prescriptions derived from experimental research. *Journal of Educational Psychology Monograph, 62,* 253–278.

Craig, R. A., Scheibe, R., Raskin, D. C., Kircher, J., & Dodd, D. (1999). Interviewer questions and content analysis of children's statements of sexual abuse. *Applied Developmental Science, 3*, 77–85.

Dale, P. S., Loftus, E. F., & Rathbun, L. (1978). The influence of the form of the question of the eyewitness testimony of preschool children. *Journal of Psycholinguistic Research, 74*, 269–277.

Davies, G. M., Westcott, H. L., & Horan, N. (2000). The impact of questioning style on the content of investigative interviews with suspected child sexual abuse victims. *Psychology, Crime, and Law, 6*, 81–97.

Dent, H. R. (1982). The effects of interviewing strategies on the results of interviews with child witnesses. In A. Trankell (Ed.), *Reconstructing the past: The role of psychologists in criminal trials* (pp. 279–297). Stockholm: Norstedt.

Dent, H. R. (1986). An experimental study of the effectiveness of different techniques of questioning mentally handicapped child witnesses. *British Journal of Clinical Psychology, 25,* 13–17.

Dent, H. R., & Stephenson, G. M. (1979). An experimental study of the effectiveness of different techniques of questioning child witnesses. *British Journal of Social and Clinical Psychology, 18,* 41–51.

Frayer, D. A., & Klausmeier, H. J. (1971). *Variables on concept learning: Task variables*. Madison: Wisconsin Research and Development Center for Cognitive Learning.

Freeman, K. A., & Morris, T. L. (1999). Investigative interviewing with children: Evaluation of the effectiveness of a training program for child protective service workers. *Child Abuse and Neglect, 23,* 701–713.

Goodman, G. S., & Aman, C. (1990). Children's use of anatomically detailed dolls to recount an event. *Child Development, 61*, 1859–1871.

Goodman, G. S., Bottoms, B. L., Schwartz-Kenney, B. M., & Rudy, L. (1991). Children's testimony about a stressful event: Improving children's reports. *Journal of Narrative and Life History, 1*, 69–99.

Gully, S. M. (1998). The influences of self-regulation processes on learning and performance in a team training context. *Dissertation Abstracts International*, *58*(9-B), 5175.

Hershkowitz, I. (in press). The role of facilitative prompts in interviews of alleged sexual abuse victims. *Legal and Criminological Psychology*.

Hutcheson, G. D., Baxter, J. S., Telfer, K., & Warden, D. (1995). Child witness statement quality: Question type and errors of omission. *Law and Human Behavior, 19*, 631–648.

Jones, D. P. H. (1992). *Interviewing the sexually abused child: Investigation of suspected abuse* (4th ed.). London: Gaskell.

Lamb, M. E. (1994). The investigation of child sexual abuse: An interdisciplinary consensus statement. *Child Abuse and Neglect, 18*, 1021–1028.

Lamb, M. E., & Fauchier, A. (2001). The effects of question type on self-contradiction by children in the course of forensic interviews. *Applied Cognitive Psychology 15,* 483–491.

Lamb, M. E., Hershkowitz, I., Sternberg, K. J., Boat, B., & Everson, M. D. (1996). Investigative interviews of alleged sexual abuse victims with and without anatomical dolls. *Child Abuse and Neglect, 20,* 1251–1259.

Lamb, M. E., Hershkowitz, I., Sternberg, K. J., Esplin, P. W., Hovav, M., Manor, T., & Yudilevitch, L. (1996). Effects of investigative utterance types on Israeli children's responses. *International Journal of Behavioral Development, 19*, 627–637.

Lamb, M. E., Sternberg, K. J., & Esplin, P. W. (1998). Conducting investigative interviews of alleged sexual abuse victims. *Child Abuse and Neglect, 22*, 813–823.

Lamb, M. E., Sternberg, K. J., & Esplin, P. W. (2000). Effects of age and delay on the amount of information provided by alleged sex abuse victims in investigative interviews. *Child Development*, *71*, 1586–1596.

Lamb, M. E., Sternberg, K. J., Esplin, P. W., Hershkowitz, I., Orbach, Y., & Hovav, M. (1997). Criterion-based content analysis: A field validation study. *Child Abuse and Neglect*, *21*, 255–264.

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*, 35–41.

Lamb, M. E., Sternberg, K. J., Orbach, Y., Hershkowitz, I., & Esplin, P. W. (1999). Forensic interviews of children. In. A. Memon & R. A. Bull (Eds.), *Handbook of the psychology of interviewing* (pp. 253–277). New York and Chichester, England: Wiley.

Memorandum of Good Practice. (1992). London: Her Majesty's Stationery Office.

Oates, K., & Shrimpton, S. (1991). Children's memories for stressful and non-stressful events. *Medical Science and Law, 31*, 4–10.

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 abuse victims. *Child Abuse and Neglect, 24*, 733–752.

Orbach, Y., & Lamb, M. E. (1999). Assessing the accuracy of a child's account of sexual abuse: A case study. *Child Abuse and Neglect, 23*, 91–98.

Orbach, Y., & Lamb, M. E. (2001). The relationship between within-interview contradictions and eliciting interviewer utterances. *Child Abuse and Neglect, 25*, 323–333.

Ornstein, P. A., Gordon, B. N., & Larus, D. M. (1992). Children's memory for a personally experienced event: Implications for testimony. *Applied Cognitive Psychology, 6*, 49–60.

Poole, D. A., & Lamb, M. E. (1998). *Investigative interviews of children: A guide for helping professionals*. Washington, DC: American Psychological Association.

Raskin, D. C., & Esplin, P. W. (1991). Statement validity assessment: Interview procedures and content analyses of children's statements of sexual abuse. *Behavioral Assessment*, *13*, 265–291.

Sternberg, K. J., Lamb, M. E., Davies, G. M., & Westcott, H. L. (2001). The Memorandum of Good Practice: Theory versus application. *Child Abuse and Neglect, 25*, 669–681.

Sternberg, K. J., Lamb, M. E., & Hershkowitz, I. (1996). Child sexual abuse investigations in Israel. *Criminal Justice and Behavior, 23*, 322–337.

Sternberg, K. J., Lamb, M. E., Hershkowitz, I., Esplin, P. W., Redlich, A., & Sunshine, N. (1996). The relation between investigative utterance types and the informativeness of child wit-

nesses. *Journal of Applied Developmental Psychology, 17*, 439–451.

Sternberg, K. J., Lamb, M. E., Hershkowitz, I., Yudilevitch, L., Orbach, Y., Esplin, P. W., & Hovav, M. (1997). Effects of introductory style on children's abilities to describe experiences of sexual abuse. *Child Abuse and Neglect*, *21*, 1133–1146.

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.

Stevenson, K. M., Leung, P., & Cheung, K. M. (1992). Competency-based evaluation of interviewing skills in child sexual abuse cases. *Social Work Research and Abstracts, 28*, 11–16.

Sweet, R. C. (1966). *Educational attainment and attitude toward school as a function of feedback in the form of teachers' written comments.* Madison: Wisconsin Research and Development Center for Cognitive Learning.

Walker, N., & Hunt, J. S. (1998). Interviewing child victim-witnesses: How you ask is what you get. In C. P. Thompson, D. J. Herrman, J. D. Read, D. Bruce, D. Payne, & M. P. Toglia (Eds.), *Eyewitness memory: Theoretical and applied perspectives* (pp. 55–87). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Warren, A. R., & McGough, L. S. (1996). Research on children's suggestibility: Implications for the investigative interview. *Criminal Justice and Behavior, 23*, 269–303.

Warren, A. R., Woodall, C. E., Thomas, M., Nunno, M., Keeney, J. M., Larson, S. M., & Stadfeld, J. A. (1999). Assessing the effectiveness of a training program for interviewing child witnesses. *Applied Developmental Science, 3*, 128–135.

Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology, 71*, 291–301.

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: Guilford.