Repeated Self- and Peer-Review Leads to Continuous Improvement in Child Interviewing Performance

STACIA N. STOLZENBERG

School of Criminalogy and Criminal Justice, Arizona State University

THOMAS D. LYON

Gould School of Law, University of Southern California

The present study examined whether a training model that focuses on consistent exposure to protocol procedure, self-evaluation, and intensive peer-review sessions could improve interviewers' ability to adhere to best practices. Law students (N = 19) interviewed 5- to 10-year-old children on a weekly basis as part of a semester-long forensic child interviewing class. They transcribed their interviews, and participated in 1-hr self- and peer-reviews. The proportion of each question type was calculated (option-posing, Wh- questions [what, how, where, why, when, and who], and open-invitations) within each interview for each interviewer. Across 10 weeks of interviews, interviewers consistently improved their performance, decreasing the proportion of option-posing questions by 31% and increasing the proportion of open-invitations by 47%. All interviewers improved. The present study suggests that with consistent self-evaluation and peer-review, forensic interviewers can incrementally improve their performance.

Gold-standard interviews with children are often distinguished largely by the interviewer's questioning-style, with open questions being preferred as they are more likely to elicit free narratives from children while limiting the interviewer's suggestiveness (Craig, Scheibe, Raskin, Kircher, & Dodd, 1999; Lamb, Hershkowitz, Orbach, & Esplin, 2008; Sternberg et al., 1996).

However, research has found that trained interviewers are unlikely to consistently follow the suggested guidelines when interviewing children (Cederborg, Orbach, Sternberg, & Lamb, 2000; Sternberg, Lamb, Davies, & Westcott, 2001). The present study examined whether a training model that focuses on consistent exposure to protocol procedure, self-evaluation, and intensive peer-review sessions improved interviewers' ability to adhere to best practices. Furthermore, the present study did so in a classroom environment, providing a template for incorporating training into relevant curricula.

Guides for forensic interviewers instruct interviewers to use open-ended questions and avoid the use of option-posing prompts, including yes–no and forced-choice inquiries (American Professional Society on the Abuse of Children, 2012). Adherence to such protocols is critical, as following guidelines improves the quality of information obtained from children in investigative settings, increasing the likelihood that forensic and legal decisions are appropriate (Lamb et al., 2007). This is largely because children's responses vary drastically depending on how interviewers pose inquiries. Option-posing questions are discouraged, as they tend to elicit brief responses from children and increase the risk of error (Lamb et al., 2008; Lyon, 2014). In comparison, open-invitations are heralded for their productivity (Lamb et al., 2008), as they often elicit reliable, spontaneous, and elaborate descriptions from children (Lamb & Fauchier, 2001; Lamb, Orbach, Hershkowitz, Horowitz, & Abbott, 2007). What has been debated in the field is the role of Wh- questions (what, how, where, why, when, and who), which tap cued-recall memory by refocusing children on previously mentioned details of the allegations. Although it is well known that Wh- questions are less productive than open-invitations in encouraging children to produce spontaneous and elaborate reports, interviewers commonly rely upon them as a means of eliciting information that children do not spontaneously mention in their reports (Snow, Powell, & Murfett, 2009). For example, asking children "How did you feel when . . ." is productive in eliciting children's subjective reactions, which might otherwise go unreported (Lyon, Scurich, Choi, Handmaker, & Blank 2012). For this reason, Wh- questions offer interviewers a constructive compromise at eliciting necessary information, without consistently suggesting information as with closed-ended questions.

Researchers have struggled to improve interviewers' performance, defined as asking more open-invitations and fewer closed-ended questions. Aldridge and Cameron (1999) found that after a one-week training program intended to improve the quality of investigative interviews with children, whereby interviewers learned basic information about interviewing, practiced skills through video-taped role playing, and recorded interviews with school children following a staged incident, trained interviewers did not differ from untrained interviews in questioning strategies. Other similar work by Warren et al. (1999) found that after a 10-day intensive training program, whereby interviewers learned about basic child development,

research-based interviewing strategies, and practiced interviewing, the effects on post-training performance were mixed; interviewers significantly decrease their use of yes-no questions from 74 to 66%, but had no change in their use of open-ended questions. These results are echoed in other research examining interviewers' performance after training (Stevenson, Leung, & Cheung, 1992). However, these studies did not provide individual feedback or group peer-review.

Review and feedback may be critical to gain improvements in inter-viewer performance. In work conducted by Lamb et al. (2002b), the researchers demonstrated that training including practice and feedback leads to improvement in interviewer performance. In their research, participants attended a 2-day training seminar explaining the conceptual and empirical basis for interviewing protocol. Participants subsequently conducted field interviews, wherein they received written feedback on such transcripts and attended individual and group training sessions monthly for a 6-month per-iod, led by social workers involved in the initial training. Participants who received such training used dramatically more invitations and fewer option-posing prompts than interviewers who did not receive feedback on posttraining, compared to baseline interviews. These findings have been replicated across multiple studies, suggesting that ongoing supervision and feedback may be key for interviewers' adherence to recommended practices (Cederborg, Alm, da Silva Nises, & Lamb, 2013; Lamb, Sternberg, Orbach, Esplin, & Mitchell, 2002; Lamb et al., 2002). One might think that because of the productivity of improved methods, interviewers would naturally continue to increase their use of open-ended questions over time. However, research has shown drift, whereby interviewers revert to their old practices once feedback is no longer given. That is, they not only do not show improvement, but also actually lose the benefits of training (Lamb et al., 2002a).

What has yet to be clarified is whether continual improvement is poss-ible with repeated, consistent feedback. The goal of the present study was to continuously measure the effects of an intensive training format character-ized by weekly self and peer-reviews of interviews with children. More specifically, we assessed the questioning style of interviewers on a weekly basis as they received regular feedback throughout the training course. The training occurred in a university seminar, assessing the potential productivity of a novel format for training future investigative interviewers.

METHODS

Participants ($N = 19$; 79% female) enrolled in a child-interviewing seminar at a major university across four consecutive semesters (six enrolled in Semester 1, six enrolled in Semester 2, four enrolled in Semester 3, and three enrolled

in Semester 4). They were expected to have special interest in child advocacy and some experience with children (even if only informal). The seminar was a semester-long course (16 weeks), developed to teach a theoretical and practical foundation for best practices for interviewing children. The course required interviewers to conduct 10 weekly interviews (8–10 min each) with an unknown student between the ages of 5 and 10 at local public elementary schools. In general, students interviewed a child every week, excluding the first few weeks of the course (heavily based on introduction to the field and core concepts) and the last few weeks of the course (when students gave an individual final presentation on a topic of choice, utilizing the student conducted interviews from that semester). Interviewed children were enrolled through after-school programs in a major U.S. city. Eligibility required that they were (a) English-speaking, (b) between the ages of 5 and 10, and (c) that their parents consented. Across the four semesters, 71 children were interviewed ($M = 7.98$, $SD = 1.17$; 51% female). Although we did not systematically collect information on ethnicity and socioeconomic status, children were enrolled in schools serving predominantly working-class Latinos and African-Americans. The topic of each weekly interview corresponded with the lecture material from that week and covered an array of subjects such as: the last trip the child took with their family, the last time they went to the doctor, or the last time they got in trouble at school. The same interviewer never interviewed any given child participant more than once.

Once the interview was concluded, the interviewer transcribed their interview verbatim, provided comments and a self-review of their own performance, and submitted it on a class forum to their peers. Each week, every interviewer was assigned to peer-review a fellow classmate's interview, examining question type as well as identifying areas for interview improvement. Feedback was written on the electronic transcript by the peer-reviewer and re-posted to the class forum.

During the 2-hr weekly seminar session, all interviewers participated in a 1-hr lecture regarding best practices and research on child interviewing, followed by a 1-hr interactive group peer-review of the past week's assigned interviews. All interviewers received regular peer-review during these sessions, during which the transcript and video recording of their past interview was examined by their peers as moderated by the course instructor.

For the purposes of the present investigation, two research assistants coded all interviews, classifying questions as the following: option-posing, Wh- (e.g., what, where, when, who, and how) and open invitations (e.g., "Tell me more about X," "What happened next?"). Option-posing questions were defined as all closed-ended questions including forced-choice, tag, negative term (includes a contraction with "not" e.g., "Isn't it true that you told X about Y?"), and simple yes-no questions. The proportion of each question type, within each interview, for each interviewer, was calculated.

Further, reliability was assessed on 20% of the sample and Cohen's kappa =.84.

To determine the effect of regular peer-review on interviewer performance we conducted three generalized estimating equation analyses (one per question-type), allowing for a within-subjects analysis of participants' interviewing growth, as a function of time. Because the dependent variable was the proportion of question-types in weekly interviews, we used a linear model in these analyses to assess interviewer growth across time. Standard errors were estimated using a robust variance estimator. Analyses were conducted with a first-order autoregressive structure, assuming that the relationship between variances change in systematic ways across time for interviewers. Parameter estimates were achieved using hybrid maximum likelihood estimation.

RESULTS

Age of the child interviewed was unrelated to the proportion of option-posing, Wh-, and open invitations asked, and as such, is not considered further. Figure 1 descriptively shows the progress of student's performance across the 10-weeks, displaying the mean proportion of each question type across time. Table 1 presents the results of the three consecutive general estimating equations. Across time, interviewers consistently improved the quality of their interviews by significantly decreasing the proportion of
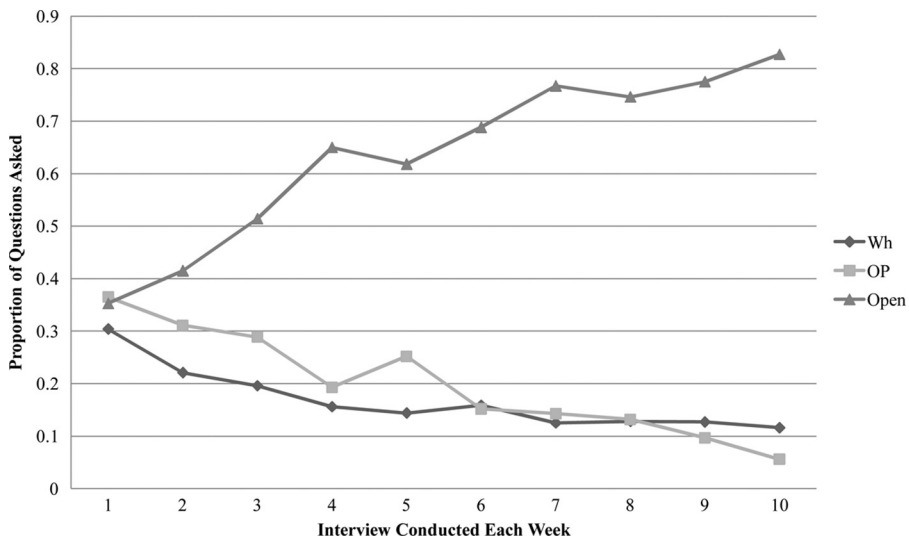

FIGURE 1 Interview performance, as defined by proportion of option-posing (OP), Wh-, and open-invitation questions, across 10 weeks with continuous peer review feedback given weekly.

TABLE 1 Three General Estimating Equations Determining the Effect of Regular Peer-Review across time on interviewers' performance, as defined by their proportion of option-posing, Wh- and open-invitation questions

	B	SE	Wald χ^2	*p*-value	95% Confidence interval	
					Lower	Upper
Option-Posing Proportion						
Intercept	0.37	0.04	93.16	<0.001	0.29	0.44
Time (in weeks)	−0.03	0.01	22.70	<0.001	−0.05	−0.02
Time × Interviewer	0.00	0.01	0.23	0.631	0.00	0.00
Wh- Proportion						
Intercept	0.27	0.03	61.25	<0.001	0.20	0.33
Time (in weeks)	−0.01	0.01	3.71	0.054	−0.03	0.00
Time × Interviewer	0.00	0.00	1.41	0.236	0.00	0.00
Open Invitation Proportion						
Intercept	0.35	0.04	85.99	<0.001	0.28	0.42
Time (in weeks)	0.05	0.01	23.17	<0.001	0.03	0.07
Time × Interviewer	0.00	0.00	0.23	0.628	0.00	0.00

option-posing questions and significantly increasing the proportion of open-invitations. There was no significant change in the utilization of Wh- questions across time. All participants improved from their baseline performance to their final interview; on average, participants decreased their proportion of option-posing questions by 31% (with the most significant decreasing being 53%) and increased their proportion of open-invitations by 47% (with the most significant increase being 88%). However, there was variability in the skills that each participant acquired, especially for the few students who started off with rather impressive baseline scores. For example, one participant began in his first interview asking no option-posing questions, but by the final interview doubled his proportion of open-invitations from 50% to 100%. Another student who began with a rather high open-invitation rate (65%) and relatively low option-posing rate (19%) was able to still improve his performance by the final interview (open-invitation rate increased to 73% and option-posing rate decreased to 13%). It should be noted, though that the majority of participants initially asked very few open-invitations and many option-posing questions. Without exception, these participants also improved. For example, the participant with the highest proportion of option-posing questions at the baseline interview (60%) decreased her usage to 21% by the final interview, while also increasing her open-invitation inquires from 32% to 75% of questions asked.

DISCUSSION

The purpose of the present study was to assess the effectiveness of continuous peer-review feedback on interviewer's performance. More specifically,

we examined adherence to best practices for question-type recommenda-tions in interviews with children. Across a 10-week training and review per-iod with weekly sessions, interviewers consistently and incrementally improved their performance; interviewers decreased the proportion of option-posing questions by 31% and increased the proportion of open-invitations by 47%. The findings from the present study suggest that with consistent self- and peer-review feedback, interviewers gradually replaced less productive option-posing questions with open-invitations (e.g., "Tell me everything that happened"), thereby improving the quality of their interviews. This finding is consistent with prior work suggesting that feedback is critical for interviewer improvement (Cederborg et al., 2013; Lamb et al., 2002a, 2002b), while making a unique contribution by demon-strating that incremental improvement is achieved across time.

Of note was that the proportion of Wh- questions did not change across time. This is likely a positive, and practical, outcome of the training program. Interviewers often need methods of eliciting disclosures and details about past experiences that maximize completeness without compromising accu-racy yet are unable to elicit complete reports through open-invitations. Wh- questions afford interviewers a useful tool. Although Wh- questions may not be as productive as open-invitations (Lamb et al., 2008, 2011), they do provide a method of tapping free-recall memory by focusing children on aspects or details that they have previously mentioned, to expand the child's report when specificity is lacking through responses to open-invitations. In addition, when used as suggested, Wh- questions do not introduce undis-closed information, thereby avoiding the risk of contamination that is often present when closed-ended questions are used.

Some limitations and future directions should be noted. First, there was no comparison group. It is possible that students would have shown improvement even without feedback. However, this seems unlikely, because previous research has found that feedback is essential for improvement (Lamb et al., 2002b), and that interviewers regress without ongoing feedback (Lamb et al., 2002a). Second, the sample was limited in that only 19 novice student interviewers were examined. Future work should include a larger sample, with more variability in experience, including active interviewers. It is possible that forensic interviewers with more experience will have more difficulty in improving their performance. Third, the present study did not assess drift in interviewing technique after the conclusion of the training per-iod; while interviewers improved over the course of the academic term, stu-dies assessing progress at longer intervals are required to assess true mastery. Fourth, we did not give students practice interviewing preschool-aged chil-dren, which may present unique challenges in eliciting complete reports due to young children's developing language abilities and memory capaci-ties. Future work should explore how training can facilitate high interview quality with children of all ages.

Finally, future training should instruct interviewers on how to use Wh-questions to maximize children's productivity, when open-invitations have been exhausted and more specificity is still needed. Recent work has begun to distinguish among different types of Wh- questions, finding that when used by prosecutors interviewing children about abuse in court, Wh- questions focusing on actions (e.g., "How did you get hurt?") were nearly three times as productive in eliciting details from children than static Wh- questions that inquire about specific contextual information (e.g., "What did he wear?"; Ahern, Stolzenberg, & Lyon, 2015).

In conclusion, the findings from the present study are consistent with prior work suggesting that feedback is helpful in interviewer's utilizing recommended practices (Lamb et al., 2002a, 2002b). This study demonstrated that with weekly feedback and peer review, continuous improvement is possible. Moreover, it demonstrates the feasibility of integrating interview training and feedback into the classroom, potentially serving as a model for future instruction.

FUNDING

REFERENCES

Ahern, E. C., Stolzenberg, S. N., & Lyon, T. D. (2015). Do prosecutors use interview instructions or build rapport with child witnesses? *Behavioral Sciences and the Law*, *33*. doi:10.1002/bsl.2183

Aldridge, J., & Cameron, S. (1999). Interviewing child witnesses: Questioning techniques and the role of training. *Applied Developmental Science*, *3*, 136–147. doi: 10.1207/s1532480xads0302_7

American Professional Society on the Abuse of Children. (2012). *Practice guidelines: Investigative interviewing in cases of alleged child abuse*. Retrieved from http://www.apsac.org/ (last accessed 6/11/15).

Cederborg, A. C., Alm, C., da Silva Nises, D., & Lamb, M. (2013). Investigative interviewing of alleged child abuse victims: An evaluation of a new training programme for investigative interviewers. *Police Practice and Research*, *14*, 242–254. doi: 10.1080/15614263.2012.712292

Cederborg, A. C., Orbach, Y., Sternberg, K. J., & Lamb, M. E. (2000). Investigative interviews of child witnesses in Sweden. *Child Abuse & Neglect*, *24*, 1355–1361. doi: 10.1016/S0145–2134(00)00183–6

Craig, R. A., Scheibe, R., Raskin, D. C., Kircher, J. C., & Dodd, D. H. (1999). Interviewer questions and content analysis of children's statements of sexual abuse. *Applied Developmental Science*, *3*, 77–85. doi: 10/1207/s1532480xads0302_2

Lamb, M. E., & Fauchier, A. (2001). The effects of question type on self-contradictions by children in the course of forensic interviews. *Applied Cognitive Psychology*, *15*, 483–491. doi: 10.1002/acp.726

Lamb, M. E., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2008). *Tell me what happened: Structured investigative interviews of child victims and witnesses.* Chichester, UK: Wiley.

Lamb, M. E., La Rooy, D. J., Malloy, L. C., & Katz, C. (Eds.). (2011). *Children's testimony: A handbook of psychological research and forensic practice.* Chichester, UK: Wiley-Blackwell.

Lamb, M. E., Orbach, Y., Hershkowitz, I., Esplin, P. W., & Horowitz, D. (2007). Structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: A review of research using the NICHD Investigative Interview Protocol. *Child Abuse and Neglect*, *31*, 1201–1231. doi: 10.1016/j.chiabu.2007.03.021

Lamb, M. E., Orbach, Y., Hershkowitz, I., Horowitz, D., & Abbott, C. B. (2007). Does the type of prompt affect the accuracy of information provided by alleged victims of abuse in forensic interviews? *Applied Cognitive Psychology*, *21*, 1117–1130. doi: 10.1002/acp.1318

Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., & Mitchell, S. (2002a). Is ongoing feedback necessary to maintain the quality of investigative interviews with allegedly abused children? *Applied Developmental Science*, *6*, 35–41. doi: 10.1207/S1532480XADS0601_04

Lamb, M. E., Sternberg, K. J., Orbach, Y., Hershkowitz, I., Horowitz, D., & Esplin, P. W. (2002b). The effects of intensive training and ongoing supervision on the quality of investigative interviews with alleged sex abuse victims. *Applied Developmental Science*, *6*, 114–125. doi: 10.1207/S1532480XADS0603_2

Lyon, T. D. (2014). Interviewing children. *Annual Review of Law and Social Science*, *10*, 73–89. doi: 10.1146/annurev-lawsocsci-110413-030913

Lyon, T. D., Scurich, N., Choi, K., Handmaker, S., & Blank, K. (2012). "How did you feel?": Increasing child sexual abuse witnesses' production of evaluative information. *Law & Human Behavior*, *36*, 448–457. doi: 10.1037/h0093986

Snow, P. C., Powell, M. B., & Murfett, R. (2009). Getting the story from child witnesses: Exploring the application of a story grammar framework. *Psychology, Crime & Law*, *15*, 555–568. doi: 10.1080/10683160802409347

Sternberg, K. J., Lamb, M. E., Davies, G. M., & Westcott, H. L. (2001). The memorandum of good practice: Theory versus application. *Child Abuse & Neglect*, *25*, 669–681. doi: 10.1016/S0145-2134(01)00232-0

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 witnesses. *Journal of Applied Developmental Psychology*, *17*, 439–451. doi: 10.1016/S0193-3973(96)90036–2

Stevenson, K. M., Leung, P., & Cheung, R. M. (1992). Competency-based evaluation of interviewing skills in child sexual abuse cases. *Social Work Research and Abstracts*, *28*, 11–16. doi: 10.1093/swra/28.3.11

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. doi: 10.1207/s1532480xads0302_6