Testing the effectiveness of a blended vulnerable witness training for forensic interviewers

Melanie Lawrie, Sonja P. Brubacher, Becky Earhart, Martine B. Powell, Linda C. Steele & David Boud

Published online: 23 Mar 2021.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

Check for updates

Testing the effectiveness of a blended vulnerable witness training for forensic interviewers

Melanie Lawrie[a] , Sonja P. Brubacher[a] , Becky Earhart[a] ,
Martine B. Powell[a] , Linda C. Steele[b], and David Boud[c,d,e]

[a]Centre for Investigative Interviewing, Griffith Criminology Institute, Griffith University, Mount Gravatt, Australia; [b]National Children's Advocacy Center, Huntsville, Alabama, USA; [c]Centre for Research in Assessment and Digital Learning, Deakin University, Geelong, VIC, Australia; [d]Faculty of Arts and Social Sciences, University of Technology Sydney, NSW, Australia; [e]Centre for Research on Work and Learning, Middlesex University, London, UK

ABSTRACT

The current study tested the effectiveness of a compact (18 hour) and blended (involving online and face-to-face components) training course, adapted from a previously evaluated course found to be successful in fostering long-term change in interviewing skill. The compact course was developed by trimming the previous course to only include learning activities that empirically demonstrated improvement of interviewing skills. There were 41 US forensic interviewers, with prior training experience, who took part in the research. Their interviewing habits were assessed using standardized mock interviews immediately prior to, and at the conclusion of training. A subset were assessed 9-24 months later. Results demonstrated that, despite reductions in length and content, training was effective in the short term and the subset maintained trained behaviors up to 24 months after completion. Results suggest that adjustments to training can be effective if the training remains founded on principles of human learning.

To achieve ideal individual and organization performance, employees should be trained in the improvement of their knowledge, skills, and abilities (Combs et al., 2006) and apply these learned skills to the workplace (Goldstein & Ford, 2002). The effective use of learned skills in the field is perceived as the definitive measure of skill acquisition (Blume et al., 2010). In investigative interviewing, most academics and practitioners around the world agree on a core body of interviewing behaviors and phases, yet less is known about *how* to effectively train interviewers to use these skills on the job (Lamb, 2016). There is some evidence that transfer of training can be achieved, but such training has typically been extensive, involving

lengthy courses or multiple follow-up refresher sessions (Benson & Powell, 2015; Collins et al., 2010; Lamb, 2016; Rischke et al., 2011).

In 2015, Benson and Powell published an evaluation of a training program that educates professionals in the interviewing of children and other vulnerable witnesses such as individuals with complex communication needs. This course demonstrated sustained training benefits in participants one year after completion, but it is resource intensive with multiple exercises that require time and money to support trainers having individual sessions with trainees (Benson & Powell, 2015). The creation of compact courses that retain key learning exercises might improve the efficiency of learning. Such courses could make high quality interview training more widely available to organizations whose temporal and financial resources are limited. This would allow for additional resources to be set aside for training of select groups of individuals on more advanced topics (e.g., managing extreme reluctance, effective strategies for peer review, etc.). The goal of the current project was to establish an interviewer training program that minimizes required resources, while maximizing training outcomes that have the potential to be transferred to the workplace (Goldstein & Ford, 2002). In advance of presenting the current research we outline beneficial interviewing behaviors, and the research on training, specifically in the field of vulnerable witness interviewing.

Question types and interviewing behaviors

There is broad consensus about key skills and behaviors that are beneficial in forensic interviewing of vulnerable witnesses (Lamb et al., 2018; Powell & Brubacher, 2020). Open questions (e.g., "What happened then?") encourage narrative and allow interviewees to tell their stories with minimal influence from the interviewer (Powell et al., 2005). In contrast, other question types such as yes-no or option posing (e.g., "Did X happen?") and cued-recall questions (e.g., "When did they do X?") dictate what information the interviewer seeks (Powell et al., 2005). The integration of open questions with minimal encouragers (e.g., "Mmhmm") fosters a conversation that encourages interviewees to share what they remember (Hershkowitz, 2001; Powell & Snow, 2007). Effective training should also discourage the use of questions that could increase the opportunity for bias in responses (e.g., leading questions, Powell & Snow, 2007). Interviewers can bolster the usefulness of open questions with positive interviewer behaviors, such as keeping questions simple and developmentally appropriate, and including child-friendly preparatory activities. For example, when interviewers talk with children about a neutral topic prior to raising the topic of concern they provide a practice opportunity for children to become familiar with the

expectations of a narrative-focused questioning format (for a review of practice narratives, see Roberts et al., 2011). Notwithstanding the consensus in the research literature around these skills and behaviors, interviewers require high quality training to adopt them effectively (Lamb, 2016; Powell & Brubacher, 2020).

Training for interviewers of vulnerable witnesses

Despite an existing knowledge base for the features that support learning of abstract skills, a multitude of past training programs only had short-term success in positively changing interviewers' behaviors (Benson & Powell, 2015; Lamb, 2016). To increase students' potential to retain information, a training program developed by Powell and her colleagues addressed this separation of research and practice by creating an ideal learning environment for interviewers to develop and apply recommended behaviors (Powell, 2008). One course in particular was the subject of a large-scale evaluation (Benson & Powell, 2015). It contained 15 modules and was completed over a period of roughly one year. There were theoretical (e.g., readings, videos) and practical (e.g., mock interviews, quizzes) exercises, and the course blended online individualized learning with face-to-face opportunities to engage with trainers (see Appendix for an overview). The course contained the necessary resources and explanations to understand the concepts thoroughly, coupled with ample practice opportunities. In many forensic interviewing courses, trained skills return to baseline within months after training (e.g., Smith et al., 2009); however, in Powell's course, learners retained skills over a 6- to 12-month period. The training program was effective because it delivered content in a manner consistent with principles of human learning (Blume et al., 2010).

A useful practice and assessment tool for training forensic interviewers is the mock interview, in which participants practice their interviewing skills with a trained role-player acting in the role of the interviewee (e.g., Lawrie et al., 2020; Powell et al., 2008). This is an interactive practice format that positively benefits participants' skill development (Shea & Barney, 2007) and fosters collaboration (Wills et al., 2011). Strong performance in simulated scenarios is shown to be a good indicator of successful application of communication skills in real-world situations (Weersink et al., 2019). Benefits obtained from engaging in mock interviews can be supported further with coding tasks in which participants use a manual to guide them in the identification of question types in their interviews (Yii et al., 2014). Mock interviews have been used in past evaluations for assessment of interviewer performance-growth (e.g., Benson & Powell, 2015), as well as in the present study.

Current study

Despite being based in the same principles, the efficacy of one training program does not necessarily translate to the usefulness of its derivatives (Bryk et al., 2015). Due to the potential decrease in effectiveness that could occur with course alterations, adaptations of the training evaluated by Benson and Powell (e.g., decreasing length, using fewer theoretical exercises) might change how the course affects the learner. The present study was needed to determine whether a compact version (roughly one-third to half of the original length) would confer benefits to learners. We hypothesized that the adapted course would be effective in changing interviewing skills because it was designed using the primary, evidenced-based exercises from the course evaluated by Benson and Powell (2015). We further predicted that at least some of these learned skill changes, such as a higher percentage of open questions, would be maintained post-training.

Method

Participants

There were 50 forensic interviewers from across the US who commenced the training program. They took part in three cohorts. A total of 41 completed the course; nine were unable due to other time commitments (completion rate: 82%). They completed an approximately 18-hour course over the length of 12 weeks. They had all previously received varying amounts of classroom-based training during their careers and were all trained in a nationally- or state-recognized interview protocol. Specific information about the amount of on-the-job interviewing and prior training was only available for half the sample, due to the information not being collected for the first two cohorts. For participants in the third cohort, 19% indicated that forensic interviewing comprised 26-50% of their job, 24% indicated 51-75%, and 57% said that forensic interviewing was 76% or more of their daily job. These same participants reported that they received prior training in using one or more of the following protocols: the National Children's Advocacy Center's (NCAC) child forensic interview structure (62%; National Children's Advocacy Center, 2019), the National Child Protection Training Center's (NCPTC) ChildFirst protocol (57%; National Child Protection Training Center, 2001), Tom Lyon's 10-Step investigative interview (24%; Lyon, 2005), The CornerHouse forensic interview protocol: RATAC (19%; Anderson et al., 2009), the National Institute of Child Health and Human Development investigative interview protocol (5%; Lamb et al., 2018), or a state protocol like Michigan's forensic interviewing protocol (10%; State of Michigan Governor's Task Force on Child Abuse and Neglect and Department of Human Services, 2017).

Participants were invited to participate in the research via an informational letter stating that their learning data would be de-identified and utilized in research for course evaluation. The letter explained that the program was looking for forensic interviewers wanting to improve their interviewing skills through a research-based course. Participants were not aware of specific course content prior to course commencement. If they did not want to partake in the research, they were able to withdraw their data without penalty (and still receive the training). No one withdrew data. There was no compensation for participating in the research. Griffith university's Human Research Ethics Board approved the study.

Follow-up interviews were completed between 9-24 months post-training ($M = 385$ days, $SD = 129$), similar to the 12-month follow-up time frame in Benson and Powell (2015). All prior participants received three emails over the course of three months inviting them to participate in the follow-up interview. A total of 13 participants agreed to participate, but due to availability, only 11 completed the follow-up interview. These 11 differed from the overall sample only in that they had a higher rate of repeating their post-mock interview due to initial poor performance (27%) compared to the full sample rate (17%). All other performance measures were similar to the full sample.

Materials and procedure

The training course evaluated in the present study was based on training that showed long-term sustainability (Benson & Powell, 2015). Elements of that training were underpinned by empirical research; learning components that were directly associated with positive change in interviewing skill were employed in the current course (e.g., Powell et al., 2008; Yii et al., 2014). These components included identification of question types, narrative structured mock interviews, practice, feedback, and consistent assessment (Powell & Wright, 2008). Modules and exercises that were identified as possibly unnecessary, and not directly related to this core foundation, were not utilized. This resulted in a shorter, less resource intensive course. The truncated course was then adapted to the needs of forensic interviewers from an established professional organization. Thus, it included the development of some new content as well.

The content was similar to the original course, with the removal of some 'special topics' modules (e.g., modules on notetaking, relationship evidence [grooming], cultural considerations, and interviewing people with complex communication needs). While these modules are certainly important, the goal of the compact course was to focus intensively on core interviewing skills. Thus, topics not directly connected to this foundation were removed.

Some readings and other non-interactive activities were cut entirely, and other readings were converted to brief narrated PowerPoint presentations. Newly developed content pertaining to conducting a practice narrative was included, in addition to a reading and quiz related to barriers to disclosure (see Appendix for a comparison of the current study's training program to Benson & Powell's [2015] program).

The modules in the compact course included a series of videos, quizzes, readings, exercises, and three mock interviews. Mock interviews were an opportunity to practice interviewing a child with a trained role-player via video chat. Importantly, they were also used as an opportunity for interviewers to practice exhausting open questions. Instructions for mock interviews explained to learners that the interview was not intended to be a full forensic interview, but rather should be considered akin to the free narrative phase. One of the goals of mock interviews in the training was to teach learners how to use open questions skillfully, such that they could elicit a narrative with as few specific questions as possible (recognizing that these may be asked at the end of the interview after a break; Lamb et al., 2018). Scenarios were based on actual cases to aid realism.

One mock interview took place prior to commencement, one midway, and one at course completion. The first two mock interviews were followed by verbal feedback, while the third was followed with written feedback. The second mock interview was associated with a reflection exercise where learners evaluated mock interviews as a learning tool (Lawrie et al., 2020). All mock interviews took place via Adobe Connect. As per Benson and Powell (2015), the first and last mock interviews were recorded and transcribed, and used to establish pre- and post-training performance. A follow-up mock interview was conducted 9 to 24 months (263 to 721 days) post-completion, and verbal feedback was offered as incentive. Data from the first, last, and follow-up mock interviews are the focus of the present research.

Coding and reliability

A professional transcriber transcribed the mock interviews. A single experienced research assistant coded all transcripts, and a second research assistant familiar with coding procedures double-coded 25% to establish reliability. Ten additional transcripts were double-coded during initial training of research assistants, which resulted in 50% of the transcripts having been double-coded at some point.

Assistants coded four question types (open, cued recall, closed, and leading) and ten interviewer behaviors. It should be noted that questions coded as leading also received a code for question type (e.g., open question-

Table 1. Coding definitions and reliability statistics for interviewer behaviors.

Behavior	Coding definition	Reliability
Less desirable behaviors		
Changes terms	Interviewer changed a term. For example: C: "He touched my minny" I: "Tell me about him touching your private."	85%
Complex	Double negatives, embedded clauses, multiple questions.	84%
How/why	Questions that began with "how", "why", or "how come". *Note:* This did not include the use of this phrasing in the transition prompt (e.g., "Tell me why you are here to talk to me today") or feeling questions (e.g., "How does that make you feel").	86%
Developmentally inappropriate	Questions about time, duration, measurement, specific location, body parts, or other abstract concepts. How/why questions were not double coded here.	93%
"Can you"	Open questions beginning with the phrasing "Can you … "	99%
Desirable behaviors		
Practice topic	During rapport building, interviewer asked the child to talk about what happened during a specific, real event the child experienced (not a series of questions about likes/dislikes, a story or film, etc.).	1.00
Transition prompt	Interviewer transitioned to the allegation phase using a variation of the question, "Tell me what you have come to talk about."	.82
Maintains open questions (OQ)	Interviewer asked mainly open questions for the duration of the interview (not including the practice narrative).	.88
Uses minimal encouragers (ME) effectively	Interviewer used "Uh-huh," "Ok," "Mmm-hmm," or similar, alone (not in combination with a question) to encourage further narrative at least three times during the mock interview.	.88
Open question (OQ) variety	Interviewer used at least two types of breadth and two types of depth question formats. No more than two identically phrased questions were asked in a row.	.87

leading). Coders adhered to a manual depicting an explanation of each aspect. Reliability for coding question types was assessed with Cohen's kappa and ranged from $.90 - 1.00$. Agreement for interviewing behaviors was assessed with percent agreement for quantitative variables (e.g., counting the number of developmentally inappropriate questions) and Cohen's kappa for variables that captured the presence or absence of behaviors (e.g., using an appropriate transition prompt to the topic of concern, or not). Descriptions of interviewer behaviors, operational definitions for their coding, and the reliability statistics can be found in Table 1.

Results

Assumptions testing and analytic plan

Quantitative data were screened for missing values, outliers, and assumptions. One participant was an outlier for having a very high proportion of closed questions at the post-training assessment (> 3 SD), but excluding this participant did not affect results so they were retained. Three of the question type variables (proportions open, cued recall, and closed) satisfied

statistical assumptions. As such, their usage was assessed with a multivariate repeated measures analysis of variance (MANOVA) across two timepoints (pre- and post-training) for the full sample and across three timepoints for the 11 participants who were available for long-term follow-up interviews. The distribution for proportion leading questions was non-normal (positively skewed). Further, any question type can be coded as leading, so they are not independent data from the other question types. Due to these reasons, proportion leading questions over time were assessed separately with non-parametric tests.

The less desirable quantitative interviewer behaviors (changes terms, complex questions, how/why, developmentally inappropriate, and "can you") were all positively skewed, non-normal distributions. This pattern was anticipated because interviewers were expected to use few to none of these behaviors, particularly after training. Due to non-normality, we collapsed scores into dichotomous variables (0-1 incidences of the behavior versus 2+ incidences). This allowed us to analyze the less desirable interviewer behaviors along with the desirable dichotomous behaviors using McNemar tests. The McNemar test is the repeated measures equivalent of the Pearson chi-square test and should be used to assess nonrandom change in dichotomous paired variables across two assessment points (Adedokun & Burgess, 2012). We conducted the McNemar tests using the SPSS Macro written by Garcia-Granero (http://www.how2stats.net/2011/09/two-proportions-test-related-spss.html), which avoids the addition of the problematic Yates correction with small samples (Camilli & Hopkins, 1979) and supplies 95% confidence intervals for the difference in observations (Newcombe, 1998).

Change in question type usage over time

A repeated-measures MANOVA on pre- and post-training proportions of open, cued recall, and closed questions demonstrated a significant overall effect of time, $F (2, 39) = 56.05$, $p < .001$, $\eta_p^2 = .74$. As predicted, the proportion of open questions increased from pre- to post-training (F [1, 40] $= 114.29$, $p < .001$, $\eta_p^2 = .74$), and the proportions of cued recall and closed questions decreased ($Fs \geq 58.47$, $ps < .001$, $\eta_p^2 s \geq .59$), as can be seen in Figure 1, top panel. At both time points, leading questions made up 3% of the questions asked ($SDs = 3\%$), and Wilcoxon signed rank tests indicated no significant difference, $p = .98$.

Next, we repeated the MANOVA for the 11 participants who were able to engage in a long-term follow-up interview. Results should be interpreted with caution due to the small sample. Time was again significant, $F (4, 38) = 4.05$, $p = .008$, $\eta_p^2 = .30$. As with the whole sample, proportion of open questions
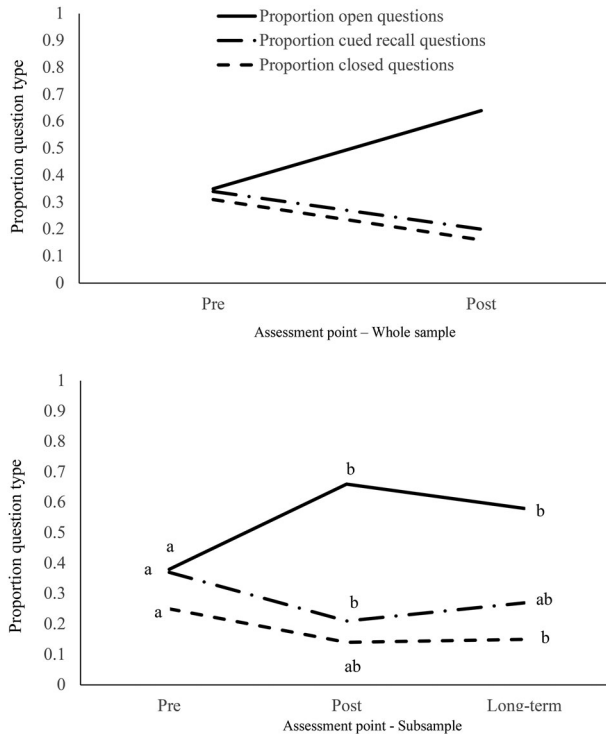
Figure 1. Proportion of open, cued recall, and closed questions across assessment points.
Note: All dependent variables differed significantly from pre- to post-assessment for the whole sample (top panel). Different letters indicate significant differences across assessments for each dependent variable in the subsample (bottom panel).

increased, $F (2, 20) = 8.76$, $p = .02$, $\eta_p^2 = .47$ and cued recall and closed questions decreased, $Fs \geq 5.28$, $ps \leq .014$, $\eta_p^2 s \geq .35$. Post hoc tests (Bonferroni $p < .05$) confirmed the proportion of open questions at time 1 was significantly lower than time 2 and time 3, the latter two not differing (see Figure 1 bottom panel). In other words, there was no drop-off in proportion of open questions at the follow-up interview compared to the post interview. For cued recall questions, there was a significant reduction from time 1 to time 2, but at time 3 the proportion of cued recall questions did not differ from either of the previous assessments. For closed questions, the reduction was not significant from time 1 to time 2 (in this subgroup of 11 participants), but a significantly lower proportion of the time 3 mock interview questions were closed questions compared to time 1; that is, they continued to improve in reducing closed questions after training ended.

Interviewer behaviors

Five of the measured interviewer behaviors were less desirable; behaviors that one would expect training to reduce. The other five were desirable
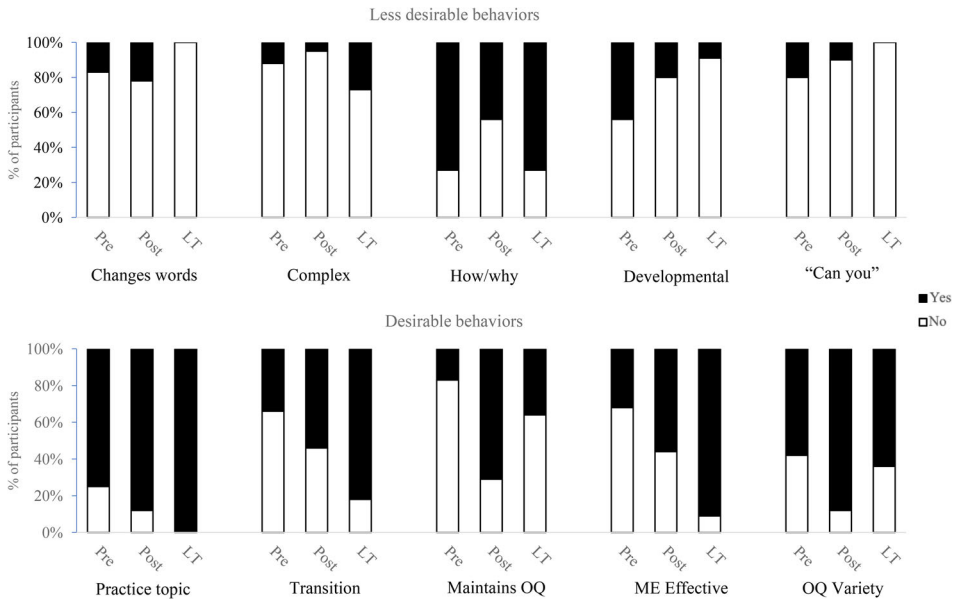
Figure 2. Observed frequencies of behaviors across assessment points.

behaviors that would be expected to increase with training. The prevalence of these ten behaviors across the three assessment points is visually depicted in Figure 2, with pre- and post-training data reflecting the whole sample and long-term (LT) data reflecting the frequency of observations for the subsample of 11. To test the hypotheses that a) the training would produce different (improved) interviewer behaviors from pre- to post-training, and b) this change would still be present at the long-term follow-up interview, we conducted sets of McNemar tests comparing performance at pre- and post-test for the whole sample and at pre- and follow-up test for the subsample (see Table 2 for observation rates and statistics).

The majority of interviewing behaviors showed significant change in the expected direction from the start to the end of the training program (Table 2, top panel). Of the less desirable behaviors, fewer participants used how/why and other developmentally inappropriate questions post-training. There was no significant change in observations for changing children's terms, asking complex questions, and prefacing questions with "can you". These behaviors, however, were already infrequent at the beginning of training (only observed in 12-20% of trainees at pre-test). For desirable behaviors, the only behavior that did *not* increase post-training was the choice of an appropriate topic for practice narrative, which was already observed in 75% of trainees' pre-training mock interviews.

Considering just the subsample available for long-term follow-up interviews, a reduction in developmentally inappropriate behaviors was observed, but there was no change in the use of how/why questions

Table 2. Change in interviewer behaviors across assessment points.

Whole Sample[a]		Pre No[b]	Pre Yes	Post No	Post Yes	Long Term No	Long Term Yes	X^2	p	Difference	95% CI
Less Desirable[c]	Changes terms	83	17	78	22			0.33	.56	−5	−22, 12
	Complex	88	12	95	5			3.00	.08	7	−3, 20
	How/why	27	73	56	44			**7.20**	**.007**	**29**	**8, 47**
	Developmental	56	44	80	20			**4.17**	**.041**	**24**	**1, 44**
	"Can you"	80	20	90	10			4.00	.045	10	−1, 22
Desirable[d]	Practice Topic	25	75	12	88			1.92	.17	12	−29, 5
	Transition	66	34	46	54			**4.57**	**.03**	**−20**	**−35, −2**
	Maintains OQ	83	17	29	71			**22.00**	**<.001**	**−54**	**−67, −35**
	ME effective	68	32	44	56			**5.56**	**.02**	**−24**	**−42, −4**
	OQ Variety	42	58	12	88			**8.00**	**.005**	**−29**	**−46, −10**
Subsample[e]											
Less Desirable	Changes terms	82	18			100	0	–	–	18	–
	Complex	82	18			73	27	0.33	.56	−9	−41, 25
	How/why	27	73			27	73	0	.99	0	−29, 29
	Developmental	46	54			91	9	**5.00**	**.03**	**45**	**6, 71**
	"Can you"	91	9			100	0	–	–	9	–
Desirable	Practice Topic	27	73			0	100	–	–	−27	–
	Transition	55	45			18	82	2.67	.10	−36	−64, 4
	Maintains OQ	82	18			64	36	1.00	.32	−18	−49, 18
	ME effective	73	27			9	91	**7.00**	**.008**	**−64**	**−83, -23**
	OQ Variety	46	54			36	64	0.33	.56	−9	−38, 23

[a]All observations, difference scores, and confidence intervals are percentages based on whole sample ($N = 41$).
[b]For less desirable behaviors a "No" observation can represent up to one incidence of the behavior (0-1).
[c]The incidence of less desirable behaviors should decrease across assessments (decrease in Yes observations).
[d]The incidence of desirable behaviors should increase across assessments (increase in Yes observations).
[e]All observations, difference scores, and confidence intervals are percentages based on the subsample ($n = 11$).
Bolded values are significant at $p < .05$.

(Table 2, bottom panel). Inspection of the frequencies indicates that participants returned to baseline use of how/why questions after 9–24 months post-training. There were no observations of changing children's terms or prefacing questions with "can you" in the long-term follow-up interview. Therefore these variables were not subject to statistical analysis, but can be taken as a positive indication of sustained skill. For desirable behaviors, the change in effective use of minimal encouragers was significant and was much larger than the difference observed between the pre- and post-training assessment; trainees continued to improve in their use of minimal encouragers post-training. All practice narrative topics were appropriate. Nearly twice as many trainees used an appropriate transition prompt at the long-term follow-up interview compared to pre-training but the difference was not significant. Maintaining open questions for the duration of the mock interview and using a variety of open question stems did not show a significant change across the assessment points, and in fact showed some evidence of a return to baseline.

Discussion

Similar to the full vulnerable witness course evaluated by Benson and Powell (2015), the course in the present study was successful in improving

interviewer skills when the behaviors were measured immediately after course completion. The shortening of the course length and focus on practical exercises yielded positive behavioral change when reviewed in close time proximity to the training. The expedited training format appears to be useful in training interviewing professionals, at least when they have previously been exposed to some interview training.

When examining data from the subsample of trainees who participated in the follow up after 9 to 24 months, we found that many of the best-practice behaviors present during the post-training assessment were maintained over the long term. Open questions, arguably the most fundamental skill, maintained improved performance across the post-training retention interval. On most other measures, there was evidence of sustained or improved performance from pre-training to the long-term follow-up assessment. This broad interpretation of the results—that the training is a successful tool for improving interviewers' behaviors—is consistent with the initial review of results at post-training. These comprehensive results are similar to those of the full vulnerable witness course, which showed long-term maintenance of open questions and interviewing behaviors (Benson & Powell, 2015).

Question type usage improved from pre- to immediate post-training across all question types except leading. The latter was infrequent and showed no change across assessments. Long-term follow up results showed that the level of open question performance was maintained, and the reduction of closed questions was further improved upon, even after training ceased. Use of cued recall prompts (i.e., specific "Wh-" questions) at the long-term follow up assessment showed some return to baseline performance but was not significantly different from any other timepoint. Cued recall prompts may be necessary in interviews with some populations, such as very young children (Hershkowitz et al., 2012) and interviewees with severe verbal impairments (Bearman et al., 2019). Highly effective open questioning can reduce the need for cued recall questions (Lamb et al., 2018), but cued recall questions are likely the hardest for interviewers to avoid using because they directly and explicitly request needed information (e.g., "Where are your pajamas now?"). Further, such questions may be needed to some extent to elicit information that interviewees did not think to include because they may be unaware of investigation requirements. Nevertheless, learners were explicitly instructed to maintain open questioning as much as possible during the mock interviews, so the increase in use of cued recall questions at the follow-up mock interview hints that these are probably fairly pervasive in learners' corresponding field interviews.

Open questioning should be supported by effective behaviors, such as using a variety of question stems, which maximizes the effectiveness of the questioning (Brubacher et al., 2020). At the immediate post-training assessment, nearly all of the ten behaviors demonstrated improvement. At the

long-term follow up interview, the most obvious decline to baseline performance was for how/why questions, mirroring the question type analysis that demonstrated an increase in the use of cued recall questions during free narrative after a delay. Some decrease in trained skill would be expected after a lengthy break from training, as there are many factors, including the frequency of behaviors used on the job, that affect retention. Transfer of individual skill is dependent on the support that learned skills receive from the team and organizational environment in which they are being applied (Ford et al., 2018). A recent study testing the effects of various modalities of refresher training showed that police interviewers who engaged in refresher sessions through peer discussion (compared to receiving feedback from an interviewing expert or completing online exercises) were the most likely to return to baseline performance in use of cued recall questions (Cyr et al., 2020). Taken together, the results suggest that these questions are particularly difficult for interviewers to avoid (and indeed, will be needed to some extent in most interviews).

Repetition of practice exercises can improve skill retention (Wang et al., 2013). Although the training program evaluated in the present study retained effective practical exercises from the original course, the number of these exercises was necessarily reduced. In particular, the number of mock interviews decreased from eight to three. Mock interviews may be a critical component of fostering long-term skill change because they are practical exercises with opportunities for feedback and discussion (Clark et al., 2012; Nicol & Macfarlane-Dick, 2006). We speculate that the abbreviated course was effective because it employed critical learning features but that the balance between maximum effectiveness and course length may still require some refinement (and likely depends on the target learner group; the present sample had some prior training experience). Next steps in this line of research would be to test performance gains incrementally as further learning activities are added to find the point where learning plateaus.

Three of the desirable behaviors, which improved from pre- to post-training for the whole sample, did not show improvements from baseline to long-term follow up for the subsample: using a recommended transition prompt, maintaining open questions for the duration of the mock interview, and using a variety of open question stems. These findings provide further evidence that some skills are more difficult to teach and maintain than others (Blume et al., 2010). We speculate that maintenance of these behaviors (particularly the latter two) relies on consistent application. Thus, to achieve long-term maintenance, more purposeful practice might be needed. This remains an empirical question for further research. Training developers should focus on maximizing interactive exercises and encouraging consistent practice and repetition. Future studies could assess whether increased repetition can be achieved through workplace

support of the training program (e.g., by having colleagues practice with each other at intervals during or after training), because other studies have shown that use of training skills in workplace teams has the potential to increase positive outcomes (Salas et al., 2007).

Another possible solution to minimizing decay of skills post-training is to follow an abbreviated training course with a refresher intervention (including practice and feedback), which has been shown to improve retention (Wang et al., 2013). Further, it could provide an opportunity for memory and skill retrieval, which has also been shown to increase retention abilities (Bjork & Bjork, 2011). Continuing research in this area should explore which specific course alterations impacted observed long-term outcomes. In order to apply the results from this study to length reduction of other courses, it needs to be determined whether course length, decreased number of mock interviews, or minimized theoretical exercises played the primary role in altering some behavior changes.

Limitations

The present study's results should be evaluated in light of several limitations. In particular, the sample size achieved for the follow-up evaluation was quite small. Only 11 out of 41 participants were able to participate in the long-term assessment. This subsample may have been particularly motivated to engage in further skill development, and/or had more time to partake in the assessment. It is possible that results would have been different if a greater percentage of the original sample took part. It should be noted, however, that the smaller sample appeared to be fairly representative of the full sample. Further, the whole sample of participants was a specific group of interviewing professionals who had been exposed to prior face-to-face training at some point in the past few years. Application of results may vary for individuals who have different levels of experience. However, the trainees evaluated by Benson and Powell (2015) were largely novice interviewers before training, which suggests that the exercises and activities employed in both training programs encourage long-term retention of many positive interviewing behaviors regardless of interviewer background.

Furthermore, it should be noted that this study relied on mock, rather than field interviews, for the follow-up assessment. This could be perceived as a limitation, but in Benson and Powell's study (2015) follow-up mock interview performance was significantly and positively associated with follow-up field interview performance. It is reasonable to suggest that the mock and field performance of the current sample would be similarly correlated. Future studies can also include a field follow-up to reestablish this performance connection.

Conclusions

Past research has shown that the success of an original program may not translate to modified versions that are created subsequently (Bryk et al., 2015). As such, even a training model demonstrated to be effective must continually be evaluated whenever it is modified, and the progression and use of the modifications carefully tracked (Powell & Brubacher, 2020). In the present study, we obtained evidence that a shorter, less intensive training course based on the same learning principles as the original maintained most of the desired outcomes over the long term, at least with a group of previously trained interviewers. The current results showcase the importance of testing all training programs, regardless of their relationship to previously evaluated programs or incorporation of evidence-based activities.

Acknowledgments

Disclosure of interest

The data for this study arose from a training exercise included in not-for-profit training offered by MBP and SPB to the organization where LCS is employed. The research was conducted as part of the dissertation of ML. The quantitative data that support the findings of this study are available upon reasonable request.

Ethical standards and informed consent

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation [institutional and national] and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all participants for being included in the study.

Funding

ORCID

Melanie Lawrie http://orcid.org/0000-0002-5335-2902
Sonja P. Brubacher http://orcid.org/0000-0002-2422-9913
Becky Earhart http://orcid.org/0000-0003-1222-7256
Martine B. Powell http://orcid.org/0000-0001-5092-1308
David Boud http://orcid.org/0000-0002-6883-2722

References

Adedokun, O. A., & Burgess, W. D. (2012). Analysis of paired dichotomous data: A gentle introduction to the McNemar test in SPSS. *Journal of MultiDisciplinary Evaluation*, *8*(17), 125–131. http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/336

Anderson, J., Ellefson, J., Lashley, J., Miller, A. L., Olinger, S., Russell, A., Stauffer, J., & Weigman, J. (2009). The CornerHouse forensic interview protocol: RATAC. *Thomas M. Cooley Journal of Practical & Clinical Law*, *12*(3), 193–331. https://www.cornerhousemn. org/images/CornerHouse_RATAC_Protocol.pdf

Bearman, M., Brubacher, S. P., Timms, L., & Powell, M. (2019). Trial of three investigative interview techniques with minimally verbal adults reporting about occurrences of a staged repeated event. *Psychology, Public Policy, and Law*, *25*(4), 239–252. https://doi.org/ 10.1037/law0000206

Benson, M. S., & Powell, M. B. (2015). Evaluation of a comprehensive interactive training system for investigative interviewers of children. *Psychology, Public Policy, and Law*, *21*(3), 309–322. https://doi.org/10.1037/law0000052

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher & J. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (2nd ed., pp. 59–68). Worth Publishers.

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, *36*(4), 1065–1105. https://doi.org/10.1177/ 0149206309352880

Brubacher, S. P., Benson, M. S., Powell, M. B., Goodman-Delahunty, J., & Westera, N. (2020). An overview of best practice investigative interviewing of child witnesses of sexual assault. In I. Bryce, & W. Petherick (Eds.), *Childhood sexual abuse: Forensic issues in evidence, impact, and management*. Elsevier.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.

Camilli, G., & Hopkins, K. D. (1979). Testing for association in 2 * 2 contingency tables with very small sample sizes. *Psychological Bulletin*, *86*(5), 1011–1014. https://doi.org/10. 1037/0033-2909.86.5.1011

Clark, R., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator*, *36*(1), 6–11. https://www.aft. org/sites/default/files/periodicals/Clark.pdf

Collins, M. E., Kim, S. H., & Amodeo, M. (2010). Empirical studies of child welfare training effectiveness: Methods and outcomes. *Child and Adolescent Social Work Journal*, *27*(1), 41–62. https://doi.org/10.1007/s10560-009-0190-0

Combs, J., Liu, Y., Hall, A., & Ketchen, D. (2006). How much do high-performance work practices matter? A meta-analysis of their effects on organizational performance. *Personnel Psychology*, *59*(3), 501–528. https://doi.org/0.1111/j.1744-6570.2006.00045.x https://doi.org/10.1111/j.1744-6570.2006.00045.x

Cyr, M., Dion, J., Gendron, A., Powell, M., & Brubacher, S. P. (2021). A test of three refresher modalities on child forensic interviewers' post-training performance. *Psychology, Public Policy, & Law*. https://doi.org/10.1037/law0000300

Ford, J. K., Baldwin, T. T., & Prasad, J. (2018). Transfer of training: The known and the unknown. *Annual Review of Organizational Psychology and Organizational Behavior*, *5*(1), 201–225. https://doi.org/10.1146/annurev-orgpsych-032117-104443

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation.* Wadsworth.

Hershkowitz, I. (2001). Children's responses to open-ended utterances in investigative interviews. *Legal and Criminological Psychology*, 6(1), 49–63. https://doi.org/10.1348/135532501168190

Hershkowitz, I., Lamb, M., Orbach, Y., Katz, C., & Horowitz, D. (2012). The development of communicative and narrative skills among preschoolers: Lessons from forensic interviews about child abuse. *Child Development*, 83(2), 611–622. https://doi.org/10.1111/j.1467-8624.2011.01704.x

Lamb, M. E. (2016). Difficulties translating research on forensic interview practices to practitioners: Finding water, leading horses, but can we get them to drink? *The American Psychologist*, 71(8), 710–718. https://doi.org/10.1037/amp0000039

Lamb, M. E., Brown, D. A., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2018). *Tell me what happened: Questioning children about abuse.* Wiley.

Lawrie, M., Brubacher, S. P., Powell, M. B., & Boud, D. (2020). Forensic interviewers' perceptions of the utility of mock interviews with trained actors as a training tool for child interviewing. *Child Abuse & Neglect*, 106, 104553. https://doi.org/10.1016/j.chiabu.2020.104553

Lyon, T. D. (2005). *Ten step investigative interview.* http://works.bepress.com/thomaslyon/5

National Child Protection Training Center. (2001). *ChildFirst forensic interview protocol.* https://www.zeroabuseproject.org/education-training/childfirst-forensic-interview-protocol/

National Children's Advocacy Center. (2019). *National Children's Advocacy Center's child forensic interview structure.* https://www.nationalcac.org/wp-content/uploads/2016/07/Child-Forensic-Interviewing-Best-Practices.pdf

Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Statistics in Medicine*, 17(8), 873–890. https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<873::AID-SIM779>3.0.CO;2-I

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. https://doi.org/10.1080/03075070600572090

Powell, M. B. (2008). Designing effective training programs for investigative interviewers of children. *Current Issues in Criminal Justice*, 20(2), 189– 208. https://doi.org/10.1080/10345329.2008.12035804

Powell, M. B., & Brubacher, S. P. (2020). The origin, experimental basis, and application of the Standard Interview Method (SIM): An information-gathering framework. *Australian Psychologist*, 55(6), 645–659. https://doi.org/10.1111/ap.12468

Powell, M. B., Fisher, R. P., & Hughes-Scholes, C. H. (2008). The effect of using trained versus untrained adult respondents in simulated practice interviews about child abuse. *Child Abuse & Neglect*, 32(11), 1007–1016. https://doi.org/10.1016/j.chiabu.2008.05.005

Powell, M. B., Fisher, R. P., & Wright, R. (2005). Investigative interviewing. In N. Brewer, & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 11–42). Guilford Press.

Powell, M. B., & Snow, P. C. (2007). Guide to questioning children during the free-narrative phase of an investigative interview. *Australian Psychologist*, 42(1), 57–65. https://doi.org/10.1080/00050060600976032

Powell, M. B., & Wright, R. (2008). Investigative interviewers' perceptions of the value of different training tasks on their adherence to open-ended questions with children. *Psychiatry, Psychology and Law*, 15(2), 272–283. https://doi.org/10.1080/13218710802014493

Roberts, K. P., Brubacher, S. P., Powell, M. B., & Price, H. L. (2011). Practice narratives. In M. E. Lamb, D. La Rooy, C. Katz, & L. Malloy (Eds.), *Children's testimony: A handbook of psychological research and forensic practice* (pp. 129–145). Wiley-Blackwell.

Rischke, A. E., Roberts, K. P., & Price, H. L. (2011). Using spaced learning principles to translate knowledge into behaviour: Evidence from investigative interviews of alleged abuse victims. *Journal of Police and Criminal Psychology*, *26*(1), 58–67. https://doi.org/10.1007/s11896-010-9073-8

Salas, E., Nichols, D. R., & Driskell, J. E. (2007). Testing three team training strategies in intact teams: A meta-analysis. *Small Group Research*, *38*(4), 471–488. https://doi.org/10.1177/1046496407304332

Shea, S. C., & Barney, C. (2007). Macrotraining: A "how-to" primer for using serial role-playing to train complex clinical interviewing tasks such as suicide assessment. *Psychiatric Clinics of North America*, *30*(2), e1–e29. [Mismatch] https://doi.org/10.1016/j.psc.2007.03.002

Smith, R., Powell, M. B., & Lum, J. (2009). The relationship between job status, interviewing experience, gender and police officers' adherence to open-ended questions. *Legal and Criminological Psychology*, *14*(1), 51–63. https://doi.org/10.1348/135532507X262360

State of Michigan Governor's Task Force on Child Abuse and Neglect and Department of Human Services. (2017). *Forensic interviewing protocol.* https://www.michigan.gov/documents/dhs/DHS-PUB-0779_211637_7.pdf

Wang, X., Day, E. A., Kowollik, V. K., Schuelke, M. J., & Hughes, M. G. (2013). Factors influencing knowledge and skill decay in training: A Meta-Analysis. In W. Arthur, E. Anthony, W. Bennett, & A. M. Portrey (Eds.), *Individual and team skill decay: The science and implications for practice* (pp. 92–140). Routledge.

Weersink, K., Hall, A. K., Rich, J., Szulewski, A., & Dagnone, J. D. (2019). Simulation versus real-world performance: A direct comparison of emergency medicine resident resuscitation entrustment scoring. *Advances in Simulation*, *4*(1), 1–10. https://doi.org/10.1186/s41077-019-0099-4

Wills, S., Leigh, E., & Ip, A. (2011). *The power of role-based e-learning: Designing and moderating online role play.* Routledge.

Yii, S. L. B., Powell, M. B., & Guadagno, B. (2014). The association between investigative interviewers' knowledge of question type and adherence to best-practice interviewing. *Legal and Criminological Psychology*, *19*(2), 270–281. https://doi.org/10.1111/lcrp.12000

Appendix

	Benson and Powell (2015)	Compact Course
Modules[1]	1. Establishing what constitutes best-practice interviewing 2. Defining the various questions 3. Understanding memory and language development 4. Choosing the most effective open-ended questions 5. Putting the right questions into practice 6. Introducing the topic of concern and eliciting a disclosure 7. Assessing your progress 8. Introducing the interview protocol 9. Interviewing about repeated abuse 10. Evidentiary requirements 11. Understanding relationship evidence 12. Considering cross-cultural issues 13. Interviewing witnesses with complex communication needs (CCN) 14. Recording the interview process 15. Putting it all together	1. Orientation & terminology - Modules 1-2 2. Choosing the most effective questions - Modules 2-5 3. Conducting a practice narrative - new content 4. Interviewing about repeated abuse - Module 9 5. Introducing the topic of concern - Module 6 & new content 6. Self-evaluation and assessment - Modules 7 & 15, new content
# Mock interviews	8	3
# Videos[2]	32	10
# Narrated PowerPoints	3	3
# Offline activities[3]	13	3
# Online activities[4]	16	5
# Non-assessed quizzes	23	8
# Hurdle quizzes[5]	3	2
# Assignment Submissions	7	3

[1]For the compact course, new module titles are listed with reference to the origin of module materials from the course evaluated by Benson and Powell (2015) below each title. Not all material from the original modules was included.

[2]Videos are accompanied by written commentaries, and typically show clips of interviews with children.

[3]Offline activities include any non-quiz guided exercise that can be done outside of the course (e.g., completing a reflection, practicing question stems). Responses to these activities may or may not be assignment submissions.

[4]Online activities include any non-quiz guided exercise that is completed within the course (e.g., responding to a short answer question, matching concepts). Responses to these activities may or may not be assignment submissions.

[5]Hurdle quizzes require a certain minimum score (usually 80%) before allowing the user to move forward. Multiple attempts are possible, and questions are randomly presented from a larger pool.