The Use of a Guided Peer Review Assessment for Investigative Interviewers of Child Witnesses

Sonja P. Brubacher[a], Martine B. Powell[b], Linda C. Steele[c], & David Boud[d]

a. Centre for Investigative Interviewing, Griffith Criminology Institute, Griffith University, Brisbane, Australia. ORCID ID: 0000-0002-2422-9913
b. Centre for Investigative Interviewing, Griffith Criminology Institute, Griffith University, Australia. ORCID ID: 0000-0001-5092-1308
c. National Children's Advocacy Center, Huntsville, USA
d. Centre for Research in Assessment and Digital Learning, Deakin University, Victoria, Australia; Faculty of Arts and Social Sciences, University of Technology Sydney, NSW, Australia; Work and Learning Research Centre, Middlesex University, London, UK. ORCID ID: 0000-0002-6883-272

Brubacher, S. P., Powell, M. B., Steele, L. C., & Boud, D. (2021). The use of a guided peer review assessment for forensic interviewers. *Journal of Forensic Practice.* Accepted October 27, 2021. doi: 10.1108/JFP-07-2021-0040

Author Note: Correspondence and reprint requests should be sent to Sonja P. Brubacher, s.brubacher@griffith.edu.au Centre for Investigative Interviewing, Griffith Criminology Institute, Griffith University, 176 Messines Ridge Road, Mount Gravatt, 4122, QLD.

Declaration of Interest Statement: 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.
Funding statement: This research was funded in part by an Australian Research Council (ARC) Discovery Grant to DP180100715 to MBP, DB, & SPB
Data Availability Statement: The quantitative data that support the findings of this study can be made available upon reasonable request.

Abstract

Purpose: Investigative interviewers assess their colleagues' interviews ('peer review') as a necessary part of their practice, and for their self-development. Yet, there is little guidance around what the process involves and how they might do it. Research suggests that effective peer review is supported by using guidance material. The goal of the present work was to describe the use of such a guide by a group of professionals who regularly conduct investigative interviews with children, to share what was learned with other professionals seeking to create a formalized peer review process. Methodology: Sixty US child witness interviewers completed a guided peer review assessment of an anonymous interview, as an assignment at the conclusion of an 18-hour training program that focused on developing their interviewing skills. They consented to the use of their learning data in research, and the research was approved by the university's research ethics board. Peer reviews were coded for the extent to which they used the guide to support their evaluations, and the overall quality of the review to assess the utility of the guide in supporting them to conduct effective assessments. Findings: In general, the guide and instructions for providing feedback were moderately effective in supporting the peer assessments, but results suggested specific training in how to deliver peer review would be useful. Practical implications: Through this process, we identified components that would be helpful to further increase the efficacy of peer review, summarized in 5 tips for organizations and trainers to develop effective peer review.

Keywords: investigative interview; training; peer review; adult learning; assessment; collaboration; guide

The Use of a Guided Peer Review Assessment for Investigative Interviewers of Child Witnesses

How might investigative interview trainers and organizations construct effective assessments to evaluate the quality of interviews conducted by their teams, and best support professional learning? Despite the relatively widespread nature of this evaluative process among investigative interviewers (Rivard & Schreiber Compo, 2017; Wolfman et al., 2017), the literature on the topic is scarce. Peer learning is one of the earliest methods by which we acquire knowledge (Falchikov, 2007). It can take numerous formats, but for the purposes of the current paper we adopt the definition "the use of teaching and learning strategies in which students learn with and from each other without the immediate intervention of a teacher" (Boud et al., 1999, p. 413). Even within this definition, a wide variety of activities can take place. One such example is investigative interviewers' assessments of the quality of each other's interviews for the purposes of integrating, maintaining, or improving interviewing skills, frequently termed "peer review".

Peer review of coworkers' conduct has a relatively well documented history in some professions, such as nursing. In the nursing field, the original framework and principles that guide contemporary peer review were developed over 30 years ago (see Haag-Heitman & George, 2011). We briefly summarize what is known about peer review broadly, with a focus on elements that might aid investigative interviewers with the development of their own peer review processes. The goal of the present study was to describe the use of a guided assessment among a sample of US investigative interviewers enrolled in an interviewer training course, with the aim to generate further research and guidance for peer review among interviewers of children and other vulnerable witnesses. Our focus in this paper is on child and vulnerable witness interviews

because–while all investigative interviews should be of good quality–interviews with this population necessitate extra skill and considerations (e.g., Bull, 2010).

Peer Review as a Mode of Workplace Learning

Peer review, as used to describe the process of professional interviewers evaluating each other's interviews, can be thought of as socio-constructivist learning (Vygotsky, 1962); learners interact with each other to build their knowledge and enhance their skills (Falchikov, 2007). It also increases responsibility and may help learners to assess their own practice so they can develop self-monitoring skills (Boud & Falchikov, 2006; Collins et al., 1991). Thus, peer review can accrue as much benefit to the reviewer as to the reviewee. Indeed, amongst a group of law students, ongoing peer review sessions combined with training over ten weeks improved students' own skills in conducting interviews with children (Stolzenberg & Lyon, 2015). An integrative review of 24 studies published between 1988-2018 on peer review in nursing education suggested that some benefits of peer review include greater ownership over one's own learning and performance improvement process, development of critical thinking, increased motivation, collegiality, and a sense of belonging. The core challenges associated with peer review involved peers not feeling competent and confident enough to give accurate and effective feedback information and concerns about giving or receiving negative feedback from peers. The reviewed research suggested that education around peer review, including clear and specific criteria for assessment, is needed (Tornwall, 2018).

A guide for investigative interviewers to use in evaluating their peers may be a necessary but not sufficient tool. Interviewers would have to be sufficiently skilled interviewers themselves to accurately evaluate performance (Lamb et al., 2018; Powell, 2013; Powell & Brubacher, 2020). A study with 50 nursing students in Thailand showed that peer ratings using an

assessment scale were uncorrelated with ratings by experts using the same scale the *first* time it was used, but by the second assessment, ratings from students and experts were significantly related (Lai, 2016). The first round of assessments took place during the mid-term course assessment and the second during the final assessment week after the students had completed their training and had a better understanding of course material. The students' own communication skills had improved over the duration of their training, so they were better able to use the assessment scale to think critically about the skills of their peers (e.g., Kaminske, 2019). Interestingly, they were not exposed to the experts' ratings until after both assessments were complete, so the students did not receive direct information from experts regarding their peer reviews. In other words, it was their learning of the relevant knowledge, not external assessment of their evaluations, that was the main impetus for improvements in their abilities to evaluate.

The existing literature suggested that at least three elements would be needed for effective peer review of interviews: a guidance document with specific criteria for evaluation, in-depth knowledge of the criteria to be scored, and an understanding of how to translate the guided assessment into actionable feedback for the learner. A standardized guide would provide learners with explicit information as to what was expected of them, and consistent standards across evaluators (Price et al., 2007; see also Wolfman, 2016, Study 3). The use of a guide would also support assessment. Assessment has two key elements: identifying and making judgments about the expected standards of performance and comparing the quality of performance in relation to those standards (Boud, 1995). The following features are important to making assessment an effective learning activity: instilling a sense of responsibility in the learner, genuine reflection, being able to identify and apply standards, and giving and receiving feedback (Boud, 1995). Feedback information must be appropriate and manageable (Kluger & DeNisi, 1996). Further, it

must be congruent with training; when peers provided feedback that differed from training recommendations, interviewer performance declined (Cyr et al., 2021). These findings further support the argument that peer evaluators would be well-served by a strong understanding of the skills and behaviors they are assessing. In the next section we review some of the key features of a best-practice investigative interview upon which investigative interviewers might evaluate their peers. These features were included in the peer review guide used in the current study.

General Overview of Best Practices in Interviewing

It is widely accepted that interviewers should maximize their use of open-ended questions and minimize use of specific and leading ones; the proportions of these question types in an interview has been the dominant yardstick by which to evaluate interview quality over the past two decades (e.g., Benson & Powell, 2015; Johnson et al., 2015; Luther et al., 2015; Otgaar et al., 2019; Warren et al., 1999). Open-ended questions (such as "Tell me what happened") encourage elaborate responses and do not constrain response parameters, whilst specific questions encourage brief responses to information the interviewer seeks. Leading questions introduce information not yet raised by an interviewee and they can be open-ended or specific (see Powell & Snow, 2007).

Priority should be given to non-leading open-ended questions at least until a witness' narrative account is exhausted (Lamb et al., 2018). Many interview guidelines for children suggest a narrative questioning phase at the beginning of the interview and specific questions (if needed) at the end of the interview paired with open-ended questions (e.g., "You talked about a person named Lolo being in the tent. Who is Lolo?" [after child's response] "Tell me more about what happened in the tent with Lolo"). A non-exhaustive list of guidelines structured in this manner include the National Institute of Child Health and Human Development (NICHD)

protocol (Lamb et al., 2018); Tom Lyon's ten step version (Lyon, 2005); the National Children's Advocacy Center's (NCAC) Investigative Interview Structure (NCAC, 2019); Guidance for Achieving Best Evidence (ABE) in Criminal Proceedings (Ministry of Justice, 2011); and the Standard Interview Method (SIM; Powell & Brubacher, 2020).

In addition to using best-practice questioning approaches, interviewers are also encouraged to follow prescribed interview phases (e.g., ground rules, episodic memory training; see La Rooy et al., 2015; Powell & Brubacher, 2020, for reviews). Further, there are effective behaviors that interviewers can use to enhance the quality of their interviews. For example, not interrupting a child, using their words and terminology, and using minimal encouragers all convey the message that the child is being heard (Powell & Snow, 2007). Eliciting an account of an episodic event during the practice narrative has been shown to yield greater episodic detail later in the interview than eliciting generic information (Brubacher et al., 2011). Crafting prompts that query actions rather than descriptions is associated with more productive child responses (Ahern et al., 2018). In a related vein, using a variety of open-ended question stems (instead of repeating a stem like "tell me") is necessary for shaping children's accounts because different question stems have varying functions and therefore elicit different kinds of details (Feltis et al., 2010; Lamb et al., 2003). Some interview evaluations use checklists to identify whether prescribed interview phases and effective behaviours were present (e.g., Benson & Powell, 2015; Luther et al., 2015).

Current Study

Professionals enrolled in an online advanced investigative interviewing course completed a peer review assignment at the conclusion of their training. The assignment is the focus of this research. Learners accessed an instruction booklet online to guide them through the peer review.

We coded their assignments for the extent to which they used the guide to aid their reports and the utility of their reports (i.e., thoroughness of report, whether they were able to make concrete and actionable recommendations for improvement). One co-author trainer and two independent experts also evaluated the quality of all peer reviews. We used both measures to evaluate whether the guidance was useful, in which aspects it was most helpful, and where improvements to the guidance could be made. Our goal was to provide investigative interview trainers and organizations with a framework and "lessons learned" upon which to develop new–or adjust existing–peer review procedures.

Method

Participants

Interviewers applied to a major US organization to take a blended interview training course. The course involved online exercises and one-on-one mock interview sessions with trainers who delivered feedback information via videochat. Of approximately 300 applicants, 80 were selected to take the training in four cohorts commencing between September 2017 and November 2019. Selection was made based on application date and geographical representativeness (with priority given to earlier applicants and those from regions with fewer applicants). The focus of the training was on teaching investigative interview skills, not specifically on how to conduct peer review. Participants in the first three cohorts were part of an evaluation of the training program (i.e., all participants that had completed the training up to the time the evaluation was conducted; blinded, 2021).

Although the course was pass/fail, the peer review assignment that is the focus of the current study was not graded. It was not accessed by the research team until after all four cohorts had completed the training. Twenty participants (4-6 per cohort) withdrew during training,

usually due to work commitments, and did not complete the assignment. These interviewers did not differ from participants who were retained in terms of their baseline interview performance, assessed by mock interview (see blinded et al. [2021], for a detailed overview of the coding and analyses used with the subsample described above). As a result, the final sample included 60 investigative interviewers (59 female). They were in the following age groups: 20-29 (18%), 30-39 (42%), 40-49 (34%), 50+ (6%). Most (83%) indicated that conducting investigative interviews with children made up at least 50% of their employment role.

The participants (henceforth "learners") were informed via a consent form that their anonymous learning data (e.g., quiz scores, assignments) would be used for research evaluation and improvement of training materials. They were permitted to decline the use of their data without penalty, but none did. There was no financial compensation for participation. The research was approved by [blinded] university.

Materials

Interview to evaluate

An eight-page interview transcript with a 7-year-old female alleging inappropriate touching by an adolescent acquaintance was developed to provide learners an interview to assess. It was based on an actual interview conducted by a prior learner of a similar training program, but we edited it for anonymity and to create various opportunities for evaluation (i.e., to maximize the number of features learners could discuss in their peer review). For example, we altered some of the question types to create better or worse questions, purposely omitted an interview instruction, made a child response ambiguous regarding body touch location, etc.

Assessment guidance and score sheets

Learners accessed a coding manual that explained the various question types and how to identify them, a peer review guide, and three score sheets to help them assess the interview. The peer review guide outlined the steps that would be taken in preparing to write the assessment: 1) coding and evaluating the types of questions asked, 2) identifying strengths and limitations in the interview practice, and 3) developing strategies to aid the learner in future practice. A tally sheet for recording the total number of each question type supported the first step. Learners prepared for the second step by using a rubric to assess the peer interviewer's practice. The rubric contained 27 checklist items of best practice spanning the full interview (e.g., delivered an instruction not to guess, engaged in episodic memory training, used appropriate prompts to elicit the topic of concern, provided a respectful closure, etc.). Checklist items were created by consulting best practice guidelines and expert/peer review checklists used by several training organizations (including, but not limited to, the organization that participated in the current study). Alongside the checkbox column was a column for notes to elaborate if necessary. To prepare for the third step, learners read some examples of concrete feedback associated with commonly observed interviewer challenges (e.g., "interviewer uses open-ended questions, but the questions are of limited variety: this interviewer could practice different open-ended question stems so that they come to mind easily").

After completing these three steps, learners read instructions to prepare a written evaluation of approximately 500 words, taking into account the information they prepared when reviewing the interview. These instructions were embedded into the online course, and learners typed or copied and pasted their response into a submission box in their web browsers.

Procedure

The peer review activity was the last substantive exercise in a six-module course on best practice interviewing with young children. The training course covered the following topics: Orientation and terminology (including identifying question types), choosing the most effective questions, conducting episodic memory training (narrative practice), interviewing about allegations of repeated abuse, introducing the topic of concern and eliciting a disclosure, and assessment (self and peer). When learners reached Module 6, they were instructed to complete the peer review exercise. After learners submitted their peer review evaluation, the learning system automatically generated a model answer written by the course trainers. Learners could compare their own responses with the model response. They were not able to edit their answers after viewing the model answer.

Coding of Peer Review Assignments

We evaluated the peer reviews in two primary ways: objective criteria based on materials with which the participants were provided to support their reviews, and subjective judgments of interview trainers who regularly give peer feedback. For the objective criteria, we created a list of features upon which learners might comment from the interview transcript (see Table 1). The features were derived from the overarching instructions (Group 1), the checklist learners were given (Group 2; e.g., "Interviewer delivered ground rules"), and the remaining guidance documents (Group 3; i.e., the *assignment guide* and the *strategies and examples for providing feedback*). The latter was included because the guidance documents contained some broader features not included on the checklist (e.g., considerations as to whether questions were appropriate for child's age/development).

Once coding began, nine additional items were added to the list because they appeared in learners' written assignments. The first seven were arguably advanced concepts that learners

might be expected to identify because the content was included in the broader training course, but not explicitly present on the peer review guidance or tally sheets (Group 4, e.g., using prescribed wording for certain prompts). Finally, two added items were not part of the broader training course or the assignment instructions, and/or they suggested non-recommended practice (Group 5). These were: 1) learner provides a concrete suggestion for improvement that is *not* evidence-based and 2) learner uses the peer review as an opportunity to talk about own challenges. To qualify as not evidence-based, the recommendation had to be something that was not covered in the training course, not a recommendation of the organization providing the training, and either explicitly discouraged (e.g., providing contingent support, asking specific questions at the outset of the interview) or controversial (e.g., anatomical dolls). When coding whether learners used the peer review as a forum for their own challenges, we assigned it a positive or negative valence. Some learners shared how they overcame their own challenges in a manner that could be helpful to a peer. Other learners used the forum to expound on their own difficulties without constructive insight. Previously coded assignments (n = 10) were revisited considering the coding additions.

The purpose of dividing the items into groups was to allow us to examine the complexity of the peer reviews. The first four groups represent a continuum from using the tally sheet and guide explicitly–examining the surface features of the interviews (or micro-level analysis of the interview: items in Groups 1 and 2)–to lifting from the guided instructions to a broader and more advanced analysis of the interview (towards a macro-level analysis: items in Groups 3 and 4).

To provide an anchor against which to judge the quality of the peer review, three interviewer trainers (each with 3-5 years of experience giving feedback to interview trainees on a regular basis) read the reviews and provided global assessments. The raters included one paper

co-author and two trainers unassociated with this study and blind to participant details. Raters assigned each review a score from 0-2. A score of 0 was assigned when the review only summarized what the interviewer did without any specific feedback information (e.g., "the interviewer included ground rules and conducted a good practice narrative") and did not make concrete, actionable recommendations for improving performance. A score of 1 was assigned when the review contained only positive feedback or made some recommendations for improving performance but the recommendations were vague (i.e., did not provide clear, actionable direction). These frequently tended to be circular criticisms (e.g., "this interviewer asked many open-ended but also many specific questions, so she could work on increasing her use of open-ended questions"). A score of 2 was assigned when the review included specific actionable recommendations (e.g., "Line 206: The interviewer said, 'what's the very next thing that happened.' Instead, she could have used a depth prompt, which tend to be more beneficial with young children, like 'Tell me more about the part where his hand went like that'").

Reliability

Of the 60 peer review assignments, 30% (n = 18) were double-coded for the list of features by a research assistant otherwise unassociated with the study. The most appropriate reliability measure was percent agreement because all coding involved identifying the presence of each checklist item within the written assignment. Overall agreement ranged from 93-100%. There were 0-2 disagreements (out of 30 items) per assignment. Reliability was performed in two waves: 15% were coded partway through coding and the remaining 15% were coded at the end. Average reliability at both time points was 97%. Disagreements were resolved by re-reading the assignment and agreeing on the most appropriate code. In most cases, a disagreement resulted from one coder having missed an item.

Reliability between the expert raters was assessed with Kappa because each review received a discrete code. The co-author coded all reviews, and the two expert trainers each coded half (i.e., all 60 reviews were double-coded). Kappa was 0.76, and there were nine disagreements (in each case, differing by one value), which were resolved through discussion.

Results

We first examined the peer review assessments with respect to the instructions given and the guidance documents provided. These results are provided in advance of summarizing the scores of the 'expert' reviewers.

Learners were instructed to write an evaluation of approximately 500 words, and they did so (M = 568, SD = 230, range 185-1815). Only 13% wrote fewer than 480 words, and 15% wrote more than 600 words. These data provide evidence that learners took the assignment seriously, even though it was not graded. Next, we looked at the overall number of checklist items they included. On average, each report included 10.88 items (SD = 3.30, range 4-20). As expected, there was wide variety in the thoroughness of the reports, despite use of a guide. Unsurprisingly, there was a positive association between the number of words in the review and the number of items discussed; longer reports were more complete, r (58) = .37, p = .004. While this conclusion seems obvious, it does mean that longer reviews were not simply full of vague praise (or criticism) to meet the assignment word count goal.

Subsequently, we examined the checklist items in greater detail (see Table 2 for frequencies). Frequencies of behaviors observed in the first two groups, which included items directly from the tally sheets and coding instructions, were noticeably higher than for the third and fourth groups. This is to be expected because they were the items upon which learners were explicitly encouraged to comment. To test whether they were indeed more frequently observed,

we converted the number of behaviors reported in each group to proportions of the total number of behaviors in each (i.e., Group 1 only had four behaviors while Group 2 had 11). A Friedman test indicated a significant difference across frequencies, X^2 (3, N = 60) = 68.18, p < .001. Follow-up Wilcoxon Signed Ranks tests (corrected alpha = .05/6 = .008) indicated that all frequency comparisons differed significantly Zs ≥ 4.14, ps < .001 except Groups 3 and 4 (p = .44). Learners' evaluations contained 64% of the behaviors in Group 1 (*SD* = 25), 38% of the behaviors in Group 2 (*SD* = 17), 24% of the behaviors in Group 3 (*SD* = 20%), and 22% of the behaviors in Group 4 (*SD* = 16). Therefore, the content of learners' feedback was closely connected to the direct instructions they were given for feedback. Put another way, organizations developing peer review guidelines need to be mindful to include the specific features they want their interviewers to address.

Very few learners used the evaluation as a platform to discuss their own challenges. Of the six who did, five also explained what they do (or did) to overcome the challenges. Unexpectedly, 24 learners made a non-recommended suggestion. A few of these involved introducing anatomical dolls, but most of the non-recommended suggestions were coded when learners said that the interviewer should have been asking more specific questions to direct the interview at the outset (e.g., "If the interviewer had asked more direct questions initially, she may have obtained more detailed information"; "Once she gave you the big disclosure you continued to ask open-ended questions. Instead, ask more direct questions to help her understand what details you need, such as 'what did he do with his hand when he did that?'").

Comparisons to Expert Opinions

The experts gave scores of 0 to seven reviews and all noted that this category was difficult to assign (i.e., a rare occurrence). This finding means that most of the reviews were at

least of moderate quality, likely a reflection of the thorough guide for conducting the review. Of the remaining, 32 reviews received a score of 1 and 21 a score of 2. Review scores were not correlated with the number of words [r (58) = .12, p = .36] or items [r (58) = .02, p = .90] mentioned in the review. Thus, it was the quality of the recommendations that led experts to give reviews higher scores, not the length or the quantity of items included.

Discussion

The present descriptive study examined the utility of a guided peer review exercise for use by investigative interviewers. Through a broader interviewing course, the interviewers had received relevant training on the interviewing skills they were asked to evaluate in a peer. They were provided a list of criteria to evaluate, and instructions and examples for giving peer feedback to complete the assignment. Our goal in this paper was to provide an overview of our process and experience, with the aim to share what we learned with organizations developing their own peer review models and to encourage further research on this critical topic. We summarize what we learned (in combination with the existing literature) in '5 tips for designing effective peer review for investigative and investigative interviewers', followed by '3 outstanding questions' for trainers and researchers.

Implications for Practice

Five Tips for Designing Effective Peer Review for Interviewers

1. Interview training should include an overview of the value of doing peer review.
2. Peer reviewers need to be well-trained in the concepts they are evaluating in others.
3. Training to give effective feedback information is needed in addition to content knowledge.

4. If there are specific features upon which peer reviewers should comment, these should be explicit.
5. Oversight is needed to ensure that recommendations are consistent with best practice.

Interview Training Should Include an Overview of the Value of Peer Review

We observed that learners took the peer review exercise very seriously despite that they did not know or meet the person who conducted the interview, and they were not graded on the assignment. Motivation is likely important to completing an effective peer review evaluation (Liu, Lin, Chiu, & Yuan, 2001). Expectancy theory suggests that, to deliver a useful peer review, interviewers must anticipate that their comments would have some positive outcome (Friedman, Cox, & Maher, 2008). These outcomes may not be limited to expectations for the evaluated peer; some interviewers may recognize the beneficial effects that delivering peer review has on their own abilities (Boud & Falchikov, 2006). We suggest that peer review should be introduced early in an interviewer's professional life due to its benefits of fostering self-reflection on one's own learning and practice (Boud & Falchikov, 2006). Organizations working to develop their own peer review processes should explicitly explain to their interviewers how peer review is a learning exercise itself, and how it can foster professional development of both reviewer and reviewee.

Reviewers Need Knowledge of the Reviewed Concepts

Learners were trained in a six-module course on best practices in investigative interviewing, and they had to meet a criterion level of performance to pass the course. Even though the learners were already practicing interviewers, an evaluation of the training demonstrated that it improved their skills pre- to post-training, including 9-24 months later for a small sub-sample of trainees (blinded, 2021). Research shows that peer reviewers must develop

the underlying content knowledge before they can effectively evaluate it in others (e.g., Lai, 2016). This finding may seem like common sense, but rarely is interviewers' concept knowledge evaluated unless they are involved in a training study. If training is not evaluated (by assessing trainees pre- and post-training, and at regular intervals thereafter), it cannot be assumed that knowledge is transferred. Further, when peer reviewers have greater expertise in the content they are evaluating, they think more critically about it (Kaminske, 2019).

Training in Giving Effective Feedback Information is Needed

Despite the use of a guide with specific criteria, learners only discussed around one-third of the items in their evaluations. They were not, however, explicitly directed to comment on each one. Doing so could be viewed as an overwhelming amount of feedback and therefore not helpful (e.g., Kluger & DeNisi, 1996). The "right" amount of information should strike a balance between providing a summary of what was done well with specific comments on where improvement would be beneficial and associated actionable suggestions for improvement; these were the criteria the experts looked for when evaluating the quality of the peer reviews in the current study. The experts' opinions were not related to the length of the reviews, or the number of items discussed, so they did not view an effective review simply as one with an abundance of information.

Giving concrete, actionable, and manageable feedback information is a skill (Kluger & DeNisi, 1996), just like interviewing. The training course that our learners completed did not involve direct training in giving feedback information. Because the peer review was a learning activity, it was not ethical to manipulate the instructions such that group comparisons between interviewers given and not given guidance could be made. Nevertheless, existing literature on

peer review shares strong consensus that specific guidance and training is critical to successful peer review (Tornwall, 2018).

Specific Features for Review Should be Explicit

Learners in the present study were much more likely to comment on the items from the peer review checklist than other features from the peer review guidance documents or the broader course. Organizations and trainers designing peer review materials may want to consider a hierarchy of elements that staff should identity when evaluating interviews. These elements should be operationalized and made explicit. Many organizations and trainers use interview checklists, but these are frequently a list of chronological interview components. Reviewers identify whether the phase was present/absent and make comments where applicable (as was done in the current study). Our experience suggested that the checklist was a helpful tool to evaluate the interview quality, but learners who focused on the checklist items tended to provide simpler reviews with less actionable feedback. We propose that peer review might benefit from a higher order set of instructions that help reviewers translate micro-level analysis into macro-level learning goals (e.g., instead of identifying that a ground rule was awkwardly worded and a later question was multi-faceted, broader and more actionable feedback might be "practice wording questions as simply as you can").

Periodic Oversight Must Occur to Keep the Process Effective

Unexpectedly, 40% of the sample in the current study made a non-recommended suggestion somewhere in their review, which mostly involved earlier use of specific questioning (in particular, cued recall "wh-" questions). Other research has similarly identified a heavy reliance by investigative interviewers on wh- questions (Brubacher et al., 2020; Wolfman et al., 2017). The finding suggests that even well-trained interviewers (who had to meet criteria levels

to pass the current training course) may still provide feedback that is not evidence-based. Supervisors and, where possible, external trainers may want to check in on the quality of peer review sessions periodically (see also Cyr et al., 2021, for discussion).

The ultimate goal of developing peer review guides and associated training should be to support this activity with only minimal ongoing intervention from researchers or trainers. In her discussion of the role of peers in learning and assessment, Falchikov (2007) wrote about how students first received evaluation criteria from their teachers but learned over time to develop their own measures. The idea is that more advanced learners are better acquainted with the criteria and require less support to develop them. Collins and colleagues (1991) refer to this stage of their apprenticeship model as "fading". Fading, however, relies on the fundamental assumption that the training received by the peers is up-to-date and effective, and that skills do not deteriorate over time. In the field of child interviewing, training is a moving target because the research industry is vibrant and continually improving (Powell & Brubacher, 2020). Further, training is not always effective in promoting long-term change (Lamb, 2016). Thus, to support in-house peer review that is relatively independent, trainers need to develop refresher modules that can be updated as new findings emerge from ongoing research.

Outstanding Questions and Further Investigation

The present work raised three outstanding questions of interest, and we urge the research community to explore these issues in collaboration with interviewing professionals. Firstly, do improvements to the peer review process in fact translate to improved interviews (and for whom: the reviewer, the reviewee, or both)? The literature supports this assumption (e.g., Boud & Falchikov, 2006; Collins et al., 1991) but it has not been directly evaluated in the investigative interviewing field. Stolzenberg and Lyon (2015) found that training law students in best practice

interviewing and peer review improved their interviewing skills over ten weeks, but their design did not permit answering the question of whether it was the content knowledge or the peer review (or both) that improved performance.

Secondly, is anonymized peer review (as in the current study) more or less effective than non-anonymized (and if anonymous, should it be double- or single-blind)? We used an interview transcript rather than a live interview for multiple reasons including that it allowed us to maintain confidentiality, manipulate some of the interview questions to create opportunities for evaluation, and because transcripts permit coding of question types. It is possible that the anonymity of the exercise decreased learners' perceived responsibility to provide useful feedback (see Tornwall, 2018 for similar discussion). Our data suggest that this was unlikely, however, given that learners wrote lengthy reviews that included specific information. In contrast, anonymity might make colleagues feel more comfortable to give and receive feedback.

Finally, and related to the previous two questions, is conducting peer review (especially on an anonymous interview) more effective for one's own learning than self-review? Over the course of our training work, we have gained anecdotal insights that interviewers may give more reflective and honest feedback in a double-blind review situation than they would give themselves (i.e., their own interviews) or live colleagues with whom they work each day.

Several studies have examined the use of web technology to facilitate peer review (e.g., Lai, 2016; Schubert et al., 2019). Indeed, the peer review exercise and associated training in the current study was mounted in a web-based course. Schubert and colleagues (2019) used iPeer (https://ipeer.elearning.ubc.ca) for graduate nursing students to provide written assessments of each other's work. iPeer is an open-source application where instructors can customize the guidance and students can submit peer evaluations. Schubert and colleagues' (2019) participants

prepared for the review by completing a content module that taught them about key characteristics for effective feedback: Specific, Measurable, Attainable, Relevant, and Timely (SMART). They received specific performance standards, examples of neutral observations, constructive language, feedback that blended positive comments with areas for improvement, and actionable recommendations for improvement. The nursing students found it easier to be honest in the online environment compared to giving feedback in person, even though all learners participated non-anonymously. They reported it being more focused, faster, and appreciated the lack of direct visible access to the peer when crafting the feedback.

The utility of online peer review lies not in the modality per se, but rather in what online enables–a way to give and receive feedback information without immediacy cues (e.g., facial expressions). Online learning also allows for the development of refresher resources that could be updated and accessed by learners on their own schedules such that they have the underlying knowledge to provide useful evaluation (Benson & Powell, 2015; Lamb, 2016). In the investigative interviewing field, challenges with this approach would include privacy and legal concerns about sharing actual investigative interviews. This problem could be mitigated by developing a library of interviews for peer review where explicit permission has been granted and cases are resolved.

Conclusions

In the present study, we described the use of a guided peer review with accompanying training on best practice principles amongst a group of child witness interviewers. The current research furthers the notion that peer review activities are best supported by specific and concrete criteria to guide the evaluation, comprehensive understanding of the skills being evaluated, and training in how to translate assessment criteria into actionable and attainable recommendations.

Implications for Practice

- Trainers and organizations should invest in training and structured guidance around the importance and value of conducting peer review, the specific concepts that should be evaluated during peer review, and how to provide effective feedback information.
- Oversight is needed to make sure that recommendations made during peer review are consistent with best practice guidelines.

Ethical Approval

Ethical approval for this project was given by ([*blinded name and location of competent body*]) [XX-2018/248].

References

Ahern, E. C., Andrews, S. J., Stolzenberg, S. N. and Lyon, T. D. (2018), 'The productivity of wh-prompts in child forensic interviews', *Journal of Interpersonal Violence*, Vol 33 No. 13, pp. 2007-2015. doi: 10.1177/0886260515621084

Benson, M. S., and Powell, M. B. (2015), 'Evaluation of a comprehensive interactive training system for investigative interviewers of children', *Psychology, Public Policy, and Law*, Vol 21 No. 3, pp. 309-322. doi: 10.1037/law0000052

Boud, D. (1995), *Enhancing learning through self- assessment.* Routledge, New York, USA.

Boud, D., Cohen, R. and Sampson, J. (1999), 'Peer learning and assessment', *Assessment and Evaluation in Higher Education,* Vol 24, pp. 413-426. doi: 10.1080/0260293990240405

Boud, D. and Falchikov, N. (2006), 'Aligning assessment with long-term learning', *Assessment and Evaluation in Higher Education,* Vol 31 No. 4, pp. 399-413. doi: 10.1080/02602930600679050

Brubacher, S. P., Benson, M. S., Powell, M. B., Goodman-Delahunty, J. and Westera, N. (2020), 'An overview of best practice investigative interviewing of child witnesses of sexual assault', In I. Bryce and W. Petherick, W (Eds.), *Childhood Sexual Abuse: Forensic Issues in Evidence, Impact, and Management.* Elsevier, San Diego, USA, pp. 445-466.

Brubacher, S. P., Roberts, K. P. and Powell, M. B. (2011), 'Effects of practicing episodic versus scripted recall on children's subsequent narratives of a repeated event', *Psychology, Public Policy, and Law,* Vol 17 No. 2, pp. 286-314. doi: 10.1037/a0022793

Bull, R., (2010), 'The investigative interviewing of children and other vulnerable witnesses: Psychological research and working/professional practice', *Legal and Criminological Psychology*, Vol 15 No. 1, pp. 5-23. doi: 10.1348/014466509X440160

Collins, A., Brown, J. S. and Holum, A. (1991), 'Cognitive apprenticeship: Making thinking visible', *American Educator*, Vol 15 No. 3, pp. 6-11. Retrieved from pdf.semanticscholar.org

Cyr, M., Dion, J., Gendron, A., Powell, M. and Brubacher, S. P. (2021), A test of three refresher modalities on child forensic interviewers' post-training performance, *Psychology, Public Policy, & Law*, Vol 27 No. 2, pp. 221-230. https://doi.org/10.1037/law0000300

Falchikov, N. (2007), 'The place of peers in learning and assessment', In D. Boud and N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term,* Routledge, London UK, pp. 128-143.

Feltis, B. B., Powell, M. B., Snow, P. C. and Hughes-Scholes, C. H. (2010), 'An examination of the association between interviewer question type and story-grammar detail in child witness interviews about abuse', *Child Abuse & Neglect*, Vol 34 No. 6, pp. 407-413. doi: 10.1016/j.chiabu.2009.09.019

Friedman, B. A., Cox, P. L. and Maher, L. E. (2008), 'An expectancy theory motivation approach to peer assessment', *Journal of Management Education,* Vol 32 No. 5, pp. 580-612. doi: 10.1177/1052562907310641

Haag-Heitman, B. and George, V. (2011), 'Nursing peer review: Principles and practice,' *American Nurse Today,* Vol 6 No. 9, pp. 48-52. Retrieved from www.myamericannurse.com

Johnson, M., Magnussen, S., Thoresen, C., Lønnum, K., Burrell, L. V. and Melinder, A. (2015), 'Best practice recommendations still fail to result in action: A national 10-year follow-up study of investigative interviews in CSA cases', *Applied Cognitive Psychology*, Vol 29 No. 5, pp. 661-668. doi: 10.1002/acp.3147

Kaminske, A. N. (2019), 'Can we teach critical thinking?' Retrieved from www.thelearningscientists.org

Kluger, A. N. and DeNisi, A. (1996), 'The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory', *Psychological Bulletin*, Vol 119 No. 2, pp. 254-284. doi: 10.1037/0033-2909.119.2.254

Lai, C. Y. (2016), 'Training nursing students' communication skills with online video peer assessment', *Computers & Education*, Vol 97, pp. 21-30. doi: 10.1016/j.compedu.2016.02.017

Lamb, M. E. (2016). 'Difficulties translating research on forensic interview practices to practitioners: Finding water, leading horses, but can we get them to drink?' *American Psychologist,* Vol 71 No. 8, pp. 710-718. doi: 10.1037/amp0000039

Lamb, M. E., Brown, D. A., Hershkowitz, I., Orbach, Y. and Esplin, P. W. (2018). *Tell me what happened: Questioning children about abuse* (2nd ed.). John Wiley & Sons, Hoboken, NJ.

Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., Stewart, H. and Mitchell, S. (2003), 'Age differences in young children's responses to open-ended invitations in the course of forensic interviews', *Journal of Consulting and Clinical Psychology*, Vol 71 No. 5, pp. 926-934. doi: 10.1037/0022-006X.71.5.926

La Rooy, D., Brubacher, S. P., Aromaki-Stratos, A., Cyr, M., Hershkowitz, I., Korkman, J., … and Lamb, M. E. (2015), 'The NICHD Protocol: A review of an internationally-used evidence-based tool for child forensic interviews', *Journal of Criminological Research, Policy, and Practice,* Vol 1, pp. 76-89. doi: 10.1108/JCRPP-01-2015-0001

Liu, E. Z. F., Lin, S. S., Chiu, C. H. and Yuan, S. M. (2001), 'Web-based peer review: The learner as both adapter and reviewer', *IEEE Transactions on Education,* Vol 44 No. 3, pp. 246-251. Retrieved from https://ieeexplore.ieee.org

Luther, K., Snook, B., Barron, T. and Lamb, M. E. (2015), 'Child interviewing practices in Canada: A box score from field observations', *Journal of Police and Criminal Psychology*, Vol 30 No 3., pp. 204-212. doi: 10.1007/s11896-014-9149-y

Lyon, T. (2005), 'Ten step investigative interview, version 2', Retrieved from: https://works.bepress.com/thomaslyon/5/

Ministry of Justice. (2011). *Achieving best evidence in criminal proceedings: Guidance on interviewing victims and witnesses, and guidance on using special measures*. Home Office, London, UK.

National Children's Advocacy Center. (2019). *National Children's Advocacy Center's Child Forensic Interview structure*. Retrieved from https://www.nationalcac.org/wp-content/uploads/2019/02/NCAC_CFIS_Feb-2019.pdf

Newlin, C., Steele, L. C., Chamberlin, A., Anderson, J., Kenniston, J., Russell, A. and Vaughan-Eden, V. (2015), 'Child forensic interviewing: Best practices', *Juvenile Justice Bulletin*, pp. 1-20. https://ojjdp.ojp.gov/sites/g/files/xyckuh176/files/pubs/248749.pdf

Otgaar, H., La Rooy, D., Horselenberg, R., Hershkowitz, I., de Ruiter, C., Blezer, L., ... and Kollau, R. (2019), 'Assessing the quality of child investigative interviewing in the Netherlands', *Applied Cognitive Psychology,* Vol 33 No. 5, pp. 889-897. doi: 10.1002/acp.3521

Powell, M. B. (2013), 'An overview of current initiatives to improve child witness interviews about sexual abuse', *Current Issues in Criminal Justice*, Vol 25 No 2., pp. 711-720. doi: 10.1080/10345329.2013.12035992

Powell, M. B. and Brubacher, S. P. (2020), 'The origin, experimental basis, and application of the Standard Interview Method (SIM): An information-gathering framework', *Australian Psychologist,* Vol 55 No. 6, pp. 645-659. https://doi.org/10.1111/ap.12468

Powell, M. B. and Snow, P. C. (2007), 'Guide to questioning children during the free-narrative phase of an investigative interview', *Australian Psychologist*, Vol 42 No. 1, pp. 57-65. doi: 10.1080/00050060600976032

Price, M., O'Donovan, B. and Rust, C. (2007), 'Putting a social-constructivist assessment process model into practice: building the feedback loop into the assessment process through peer review', *Innovations in Education and Teaching International*, Vol 44 No. 2, pp. 143-152. doi: 10.1080/14703290701241059

Rivard, J. R. and Schreiber Compo, N. (2017), 'Self-reported current practices in child forensic interviewing: Training, tools, and pre-interview preparation', *Behavioral Sciences & the Law*, Vol 35 No. 3, 253-268. doi: 10.1002/bsl.2290

Schubert, C., Momeyer, M. A. and Tornwall, J. (2019), 'Nursing peer review in an online course', *Journal of Nursing Education,* Vol 58 No. 6, pp. 374-374. doi: 10.3928/01484834-20190521-13

Stolzenberg, S. N. and Lyon, T. D. (2015), 'Repeated self-and peer-review leads to continuous improvement in child interviewing performance', *Journal of Forensic Social Work,* Vol 5 No. 1-3, pp. 20-28. doi: 10.1080/1936928X.2015.1068721

Tornwall, J. (2018), 'Peer assessment practices in nurse education: An integrative review', *Nurse Education Today,* Vol 71*,* pp. 266-275. doi: 10.1016/j.nedt.2018.09.017

Vygotsky, L. S. (1962), *Thought and language* (E. Hanfmann & G. Vakar, Trans.). MIT Press, Cambridge, USA.

Warren, A. R., Woodall, C. E., Thomas, M., Nunno, M., Keeney, J. M., Larson, S. M. and Stadfeld, J. A. (1999), 'Assessing the effectiveness of a training program for interviewing child witnesses', *Applied Developmental Science,* Vol 3 No. 2, pp. 128-135. doi: 10.1207/s1532480xads0302_6

Wolfman, M. (2016), *Understanding and addressing challenges faced by forensic interviewers in their work with children.* (Unpublished doctoral dissertation). Victoria University of Wellington, NZ.

Wolfman, M., Brown, D. and Jose, P. (2017), 'Examining forensic interviewers' perceptions of practice-focused supervision', *Australian & New Zealand Journal of Criminology*, Vol 50 No 4., pp. 566-581. doi: 10.1177/0004865816655588

Table 1

Coding Definitions

Group 1: Overarching elements of the peer review	
Use of open-ended questions	– Learner recognizes that approximately half of the interview prompts were open, and evaluates global interview questioning quality.
Variety of open-ended questions	– Learner reflects on whether open-ended questions have a variety of different stems or are largely the same.
Elicitation of disclosure	– Learner evaluates how disclosure was elicited.
Recommendations for improving performance	– Learner makes an actionable recommendation for improving an evidence-based skill.
Group 2: Core checklist items	
Introduction of self/role	– Learner mentions introduction component of interview.
Explanation of equipment/room	– Learner mentions phase of introductions where recording equipment/room layout is explained (absent in evaluated interview).
Interview instructions	– Learner mentions use of ground rules/instructions (three rules given, one practiced in evaluated interview).
Narrative practice phase	– Learner mentions narrative practice phase (and otherwise gets no codes for practice phase variables).
Quality of narrative practice	– Learner provides a global evaluation of the quality of the practice phase (and otherwise gets no codes for practice phase variables).
Use of initial invitations	– Learner mentions the initial invitations used for narrative practice and/or the substantive topic.
Use of minimal encouragers	– Learner mentions the use of minimal encouragers.
Use of child's words	– Learner discusses whether interviewer used child's words.
Leading questions	– Learner evaluates whether interviewer asked any leading questions (between 2-4 in the evaluated interview depending on strictness of criteria).
Interruptions	– Learner comments on any interruptions by the interviewer.
Closure of interview	– Learner comments on how the interviewer terminated the interview.

Group 3: Sub-items or implicit in peer review instructions	
Coding of question types	– Learner reports the proportion/frequency of one or more question types.
Child development	– Learner discusses any component of the interview with regards to child development/child's age.
Use of pairing (specific + open)	– Learner discusses extent to which interviewer used pairing of specific and open-ended questions (e.g., "Who was all watching the movie with you?" "Tell me more about what happened while you were watching.").
Clarification	– Learner mentions interviewer attempts (or failures) to clarify key information.
Episodic nature of practice narrative	– Learner explicitly comments on choice of episodic topic for practice narrative.
Practice narrative questions	– Learner recognizes that nearly all of the practice narrative questions are open-ended.
Group 4: Advanced behaviours	
Action vs description prompts	– Learner comments on extent to which prompts ask for actions versus descriptions (e.g., "Tell me what happened in the lounge" versus "tell me about the lounge").
Transition prompt	– Learner comments on the transition prompt used ("Tell me what you've come to talk to me about today").
Effectiveness of open-ended questions/minimal encouragers	– Learner recognizes how the open-ended questions or minimal encouragers affect child's ability to give narrative.
Episodic reorientation	– Learner recognizes that the child provides some script details in practice and the interviewer responds by re-directing her to an episode.
Frequency question	– Learner evaluates the use of the frequency question (e.g., timing, whether it was needed).
Specific questions after break	– Learner recognizes that specific questions were mostly retained until after the break.
Missed opportunities	– Learner identifies specific locations in the transcript where a better question could have been asked and provides examples of better questions.
Group 5: Other	
Personal challenges	– Learner uses peer review evaluation as a platform to share personal interviewing challenges.
Non-recommended suggestion	– Learner provides a suggestion that is not evidence-based.

Table 2

Frequencies of Observations in Peer Review Reports

Group 1: Overarching elements of the peer review guidance	
Use of open-ended questions	25
Variety of open-ended questions	39
Elicitation of disclosure	48
Recommendations for improving performance	42
Group 2: Core checklist items	
Introduction of self/role	24
Explanation of equipment/room	13
Interview instructions	43
Narrative practice phase	20
Quality of narrative practice	14
Use of initial invitations	13
Use of minimal encouragers	53
Use of child's words	13
Leading questions	25
Interruptions	16
Closure of interview	15
Group 3: Sub-items or implicit in peer review instructions	
Coding of question types	15
Child development	26
Use of pairing (open + specific)	13
Clarification	18
Episodic nature of practice narrative	8
Practice narrative questions	7
Group 4: Advanced behaviors	
Action vs description prompts	1
Transition prompt	4
Effectiveness of open-ended questions/minimal encouragers	24
Episodic reorientation	3
Frequency question	10
Specific questions after break	22
Missed opportunities	27
Group 5: Other	
Personal challenges	6
Non-recommended suggestion	24