



A Long-Term Evaluation of Adult Sexual Assault Investigative Interview Training



Joyce van der Mark, Sonja P. Brubacher and Martine Powell

¹ Centre for Investigative Interviewing, Griffith Criminology Institute, Griffith University

Corresponding author:

Sonja P. Brubacher
s.brubacher@griffith.edu.au

Acknowledgement

The authors wish to thank the organisation that made this research possible and the participants from the organisation who gave their valuable time to participate in this study.

The content of this manuscript is solely the responsibility of the authors and does not represent the official views of the participating organisation.

ABSTRACT

This study evaluated the impact of a specialist investigative interview training program on police officers' adherence to best-practice techniques in the context of adult sexual assault. Twenty police interviewers from a metropolitan sexual assault unit in Australia participated in a repeated-measures study. Using a longitudinal design, interview performance was assessed at baseline (T1), immediately post-training (T2), and at a 10-week follow-up (T3) using mock interviews with trained actors portraying adult sexual assault complainants with varied emotional responses. Analyses indicated significant post-training enhancements in open-ended questioning and interviewer behaviours, such as ground rule application, and complete narrative elicitation, which remained stable at follow-up. Concurrently, a reduction in less effective questions (wh-, yes-no, forced-choice) was observed. Effect sizes ranged from small to large, suggesting both statistically and practically meaningful improvements. The findings support the effectiveness of targeted, skill-based interview training in improving core components of evidence-based interviewing with adult complainants. The sustained impact observed across multiple behavioural indicators reinforces the value of structured training programs that incorporate scenario-based practice and expert feedback. This study offers novel insights into the retention of best-

practice interviewing skills in adult sexual assault contexts and addresses a critical gap in the empirical evaluation of adult-focused police training.

Key Words: *Investigative Interviewing, Training Evaluation, Police, Adult Sexual Assault, Adult Interviewing.*

Introduction

Investigative interviews are central to the evidentiary foundation of adult sexual assault cases, as these interviews allow investigators to obtain information about the crime or event in question from persons directly involved (Powell et al., 2005). Beyond establishing what happened, high quality interviewing is designed to elicit narrative recall with sufficient contextual detail to guide investigations (McMillan & Thomas, 2009) and prosecutions (Chenier et al., 2021). High quality interviews are grounded in cognitive and developmental psychology and contain best-practice narrative-facilitating strategies such as asking open-ended questions, delivering ground rules that support witness autonomy and accuracy, and using minimal encouragers. These are discussed in the next section.

Narrative Facilitating Strategies

Narrative-facilitating strategies are important to elicit a comprehensive account from complainants that extends beyond merely detailing the crime itself. Especially in interpersonal crimes such as adult sexual assault, it is important to understand not only what happened, but also the context of what happened. Information about the lead-up to the event, as well as a detailed narrative account of the event itself and the events immediately following the event in question, may all be important for the investigation (Brubacher & Powell, 2025).

In adult sexual assault contexts, interviewers may also need to determine whether the presenting complaint reflects a single or repeated experience, as this has implications for eliciting distinct episode-

level narratives (Brubacher & Earhart, 2019; Hartwig et al., 2012; Snow et al., 2020). Ideally, frequency should be queried using simple questions (e.g., “Did this happen one time or more than one time?”). This is because while adults’ estimates of number and frequency are often reasonable, they are imprecise and prone to cognitive biases such as the availability heuristic (Tversky & Kahneman, 1973); with regard to repeated violence, adults might estimate the frequency of occurrences by assessing the availability of specific occurrences, which could lead to errors in the report (under- or over-estimation). As an alternative to asking if an event happened once or more than once, interviewers can use open-ended questions to elicit elaborate responses, which should contain clues to frequency of experience (Brubacher & Earhart, 2019).

Open-ended questions are critical to facilitating narratives. These questions do not dictate the content of the interviewee’s answer, allowing the interviewee to report what they remember in their own words and pace, thereby eliciting an elaborate story (narrative) about what happened (Powell & Brubacher, 2020; Powell & Snow, 2007). In comparison, specific cued-recall (e.g., ‘when’, ‘where’) and closed (e.g. ‘Did you say anything?’) questions focus the interviewee on a certain aspect of the event or direct the interviewee towards a certain answer category, thereby eliciting short responses. Cognitive and developmental research has consistently demonstrated that non-leading open-ended questions elicit the most accurate, coherent and elaborate (narrative) responses from interviewees (Lamb et al., 2018; Powell, 2013; Powell et al., 2009; Vrij et

al., 2014). Furthermore, a variety of open-ended question stems (e.g. "Tell me more. What else happened. And then what happened.") can be more effective in eliciting new information than repeating the same stem (e.g., "Tell me... Tell me... Tell me..." Powell & Guadagno, 2008; Zekiroski et al., 2024a).

Ground rules can be considered a narrative-facilitating strategy because of their purpose in empowering an interviewee to be the expert regarding his or her own experience. Ground rules are instructions for the interviewee to better understand the process and structure of the conversation during the investigative interview (Brubacher et al., 2015). For example, the interviewer might explicitly instruct the interviewee to correct mistakes made by the interviewer, alert the interviewer when they don't understand a question, or say 'I don't know' when they are unsure of an answer. The use of ground rules in an interview is aimed at reducing anxiety and discomfort of the interviewees by informing them of the conversational expectations of the interview and increasing the accuracy of the interviewee's responses (Brubacher et al., 2015; Fisher & Geiselman, 2010).

Finally, minimal encouragers are interviewer utterances that encourage interviewees to continue providing information without disrupting their narrative (Powell et al., 2016). For example, the interviewer says, "Uh-huh" or "Mm-hmm" when the interviewee stops talking. To function as a minimal encourager, the utterance must be used alone, rather than followed immediately by a question (e.g., "Uh-huh. Then what?").

Guidelines and Training

Widespread agreement on the value of these interview components has led to their incorporation in interview guidelines around the world aimed at use with children and adults (e.g. Cognitive Interview, Fisher & Geiselman, 2010; National Institute of Child Health and Human Development NICHD Protocol, Lamb et al., 2018; Achieving Best Evidence in Criminal Proceedings, Ministry of Justice, 2022; Standard Interview Method, Powell & Brubacher, 2020), with the Cognitive Interview (CI) being the dominant protocol in investigative interviewing of adult witnesses (Fisher & Geiselman, 2010). However, to transfer the guidelines to actual practice, training is essential to develop skills. This is reflected in the notion that despite acknowledgement of best practice techniques in contemporary interview protocols and training programs, adherence to best practice interviewing proves to be difficult for police interviewers (Akca et al., 2021; Brubacher et al., 2020). Training evaluation literature underscores the complexity, yet feasibility, of modifying interviewer behaviours through targeted training. Most research evaluating police interview training has found limited uptake, with post-training results indicating predominant use of specific and closed questions and, as a result, a lack of relevant detail in the narrative (Akca et al., 2021; Clarke and Milne, 2001; Snook et al., 2012). Before protocols and guidelines can be used to improve interviewing and aid evidence gathering, training must effectively promote transfer of skills. Although conventional instructional approaches often show limited lasting impact, experiential training is more effective at fostering enduring behavioural changes (Powell et al., 2010).

The training evaluation literature on child interviewing has revealed the importance of different training components, such as incremental learning (Brubacher et al., 2022), variety in tasks (Powell et al., 2016) and educational resources (Bos & Brand-Gruwel, 2017), spaced-practice through mock interviews (Powell et al., 2022), followed by expert feedback (Powell et al., 2022), as well as formal evaluation studies to better understand training delivery and maximize improvement in interview quality. Training evaluation studies on adult interviewing have yielded mixed results, and few have included open-ended questions as a measure. Acka et al. (2021) systematically reviewed the efficacy of investigative interviewing training courses and found that multiple training sessions yield the most consistent improvement in interviewing behaviour. This review also indicated a need for more systematic research on training effectiveness with more uniform and longer-term measures of effectiveness.

Current Study

The present study evaluates whether a targeted, modular training programme for specialist police interviewers is associated with (a) an increased proportion of recommended open ended question types and (b) improved interviewer behaviours that support narrative elicitation. While rapport is widely recognised as important for facilitating disclosure, rapport was not measured or analysed in this study and is therefore not considered further.

The training we developed for this study encompasses components that are found to be effective, such as gradual learning over multiple learning sessions (Brubacher et al., 2022), learning different question stems

(Powell & Wright, 2008), identifying different question types and understanding their effect on interviewee responses (Powell et al., 2016; Yui et al., 2014), and mock interviews (Powell et al., 2022). Mock interviews provide realistic, low-stakes environments that enable trainees to apply theoretical knowledge practically, receive immediate feedback, and consequently develop effective interviewing skills (Cederborg et al., 2013; Lawrie et al., 2020).

Evaluations of interview training that encompass the aforementioned components have revealed significant improvements in open-ended question usage immediately after training and sustained use of open-ended questions long after the training had ceased with no intervening follow up (Benson & Powell, 2015; Brubacher et al., 2022; Cederborg et al., 2021; Powell et al., 2016). Therefore, our hypotheses predict increases in desired behaviour – i.e., the proportion of recommended open-ended question types, the likelihood of eliciting a complete narrative account, the use of a frequency question, variety in open-ended question stems, the use of ground rules and minimal encouragers – from baseline (T1) to immediately after training (T2) and at a long-term follow-up assessment (T3).

This study makes several contributions to the literature. First, it focuses on adult complainants—a population that remains underrepresented in training evaluations. We reference child interview research where concepts directly inform adult interviewing, but our empirical focus is exclusively adult complainants. Second, it includes multiple behavioural indicators of interview quality beyond question type, thereby evaluating different components of skill. Third, it employs a three-point design that allows for

the assessment of training retention—a key, but often neglected, dimension in training evaluations (cf. Benson & Powell, 2015; Brubacher et al., 2022; Cederborg et al., 2021; Lawrie et al., 2021). Finally, the use of mock interviews with standardized yet ecologically valid scenarios enhances both the realism and reliability of the evaluation.

Methodology

Participants

Participants ($N = 21$) were experienced investigative interviewers from a Sexual Assault Squad of an Australian police service, who underwent specialist investigative interview training. Inclusion criteria were (a) active duty with the police force, and (b) active at the sexual assault squad. Participation was voluntary. One officer withdrew prior to T2 due to reassignment to another division, resulting in a final sample of 20 participants. All participants completed a demographic survey upon enrolment, capturing gender, age, education level, years of service, and prior specialist interview training.

The 20 participants were predominantly female (70%), with 20% aged 20–29 years, 40% aged 30–39 years, and 40% aged 40 years or older. Educational backgrounds ranged from completion of high school (60%) to college-level qualifications (20%) and bachelor's degrees (10%); two participants (10%) reported no qualification beyond compulsory schooling. On average, participants had served 12.3 years in the police force ($SD = 9.1$, range = 0.5–32). Most participants (90%) had received prior interview training between 2014 and 2021,

most commonly specialist training in vulnerable witness or child interviewing (70%), followed by general investigative interviewing (20%). Taken together, these characteristics reflect an experienced cohort.

An a-priori power analysis was conducted using GPower 3.1 (Faul et al., 2009). Based on Benson and Powell (2015) we expected a large effect size ($\eta^2 = .44$) and a correlation of $r = .3$ between pre- and post-training measures for a repeated measures design for one group and two measures (open-ended and specific questions). The power analysis revealed that at least eight participants were required to have a power of $1 - \beta = .95$. The study was approved by the Human Research Ethics Committee at Griffith University, as well as the executive board of the police organisation.

Procedure

The study involved three measurement points over approximately six months. Following recruitment, all participants completed a 20-minute pre-training baseline mock interview, with a professional actor playing the role of an adult sexual assault complainant. These interviews were used to establish a baseline measure of interview competence (T1: Pre-training). Next, participants completed a self-paced 20-hour specialist investigative interviewing training, developed and delivered by the Centre for Investigative Interviewing, Griffith University. The course focused on core principles of evidence-based interviewing and consisted of a 6-module online training.

The first module covered best practice guidelines of investigative interviewing. At the start of the module, a film was shown about an allegation of sexual abuse. This film was

used throughout the course as the 'staged event' for the mock witnesses in the training materials. Then, the module continued with 5 exercises, and a reading. The exercises were a practice interview with a colleague to experience the role of the interviewee, followed by a reflection on the effect of different question types on the interviewee, feedback on different question styles, a multiple-choice quiz on the question types in an interview transcript, and a multiple-choice quiz on regarding a reading.

The second module was aimed at defining the questions. In this module, trainees coded three interview transcripts. After each coding, they received feedback about the correct code for each question and completed a reflection exercise on their coding errors (i.e., what misunderstanding or confusion led to the error). The reflection was followed by information on the most common coding errors per interview and an explanation of why they may occur. The module concluded with a quiz on the coding of question types and a film depicting an interview with a mock witness accompanied by commentary on the questioning approach.

The third module contained exercises for practicing the right questions. By means of a computer simulated exercise, the trainee practiced interviewing with an avatar as many times as possible within a 1- to 2-week period. The module continued with information on question stems and an exercise to get the trainees familiarized with using the question stems. After learning the question stems, the trainees listened to 3 interviews accompanied by commentaries, read an article on the limitations of investigative interviewers' use of open-ended questions, and concluded with a quiz.

The fourth module consisted of a workshop on interviewing in the context of repeated events. During the workshop the participants attended a lecture on repeated events, followed by an interactive group discussion. In the fifth module, trainees assessed their own performance by conducting a mock interview with a course tutor trained at playing the role of an interviewee, transcribing and coding this interview, and reflecting on their performance. Directly after the interview, the interviewee received feedback from the tutor. A reflection survey was conducted in module 6, as well as another mock interview, followed by expert feedback. Participants took 2 to 3 months to finish their training.

Directly after the training, all participants completed a post-training mock interview (T2: Post-training), approximately four months after baseline. A third assessment mock interview was conducted at T3 (Follow-up), ten to twelve weeks after T2, to assess skill retention.

All interviews were conducted online through MS Teams, video-recorded, and set to a maximum duration of 20 minutes. To enhance ecological validity, three mock interview scenarios with different victim behaviour-types were based on actual adult sexual assault police reports selected by detectives. Scenarios were comparable in complexity, and all consisted of a single experience (but this was unknown to participants, who were expected to elicit information about frequency in all mock interviews).

Three female actors were trained by the first author to ensure consistency in their responses. The actors were instructed to respond to open-ended questions with two or

three relevant details, specific cued-recall questions (Wh-questions) were answered with one relevant detail, and forced choice questions (option-posing questions) or yes/no-questions were to be answered with one of the options posed or a simple yes or no without additional information.

The participants were also pseudo-randomly allocated to one of three behaviour-types (reluctant, defiant, emotional) at each of the three time points, such that each participant interviewed complainants displaying all three behaviours and the behaviours were equally represented across time points. This counterbalancing prevented learning effects based on familiarity with content. While not directly analysed in this study, these profiles will be the focus of a subsequent study on natural use of support behaviours. The roleplaying actors were blind to the research design and hypotheses.

Coding

Question Type

Each interview was video recorded, transcribed, and coded by the first author, following a standardised protocol. During coding, the coder was blind to timepoint. The coding scheme categorized questions into four broad categories, consistent with literature (Powell et al., 2016): open-ended questions, specific questions, leading questions, and minimal encouragers. Four types of open-ended questions (invitation, breadth, depth, descriptive), and four types of specific questions (cued-recall, yes/no, verification, forced-choice) were coded. Table 1 presents an overview of coding categories, and Appendix A provides a full overview of question type coding. For each interview, the proportion of question type relative to the total number of questions was calculated.

Table 1. Overview of Question Type Coding Categories

Category	Subtypes and Examples	Coding Rule or Definition
Open-ended questions	Invitations ("Tell me everything that happened...")	Do not constrain content; invite narrative.
	Breadth ("What happened next?")	
	Depth ("You said X – tell me more about X" [action])	
	Descriptive ("Tell me about [object/person]")	
Specific questions	Cued recall (Wh-questions, e.g., "Where did this happen?")	Direct the interviewee toward a specific detail; typically elicit short responses.
	Yes/no questions	
	Verification (confirming prior detail)	
	Forced choice (e.g., "Was it before or after...?")	
Leading questions	E.g., "Tell me more about when he kissed you" (if not previously disclosed)	Introduce new, undisclosed or suggestive content. Can be any question type.

Minimal encouragers	"Uh-huh", "Mm-hmm" (standalone only)	Short utterances that encourage continuation without adding content.
---------------------	--------------------------------------	--

Interviewer Behaviour

Since best practice interviewing involves the application of multiple effective behaviours beyond questioning approaches, the presence or absence of five behaviours were coded dichotomously: The elicitation of the complainant's narrative (i.e., full account obtained in terms of lead up, offence, and safe place), a frequency question (to determine if the event that is discussed happened one time or more than one time), variety in the use of open-ended question stems, the use of ground rules (e.g., "If you don't know an answer do not guess" or "Remember, I wasn't there"), and the use of minimal encouragers (e.g., "uh-huh" or "mm-hmm").

Narrative elicitation was coded as present if the interviewer elicited all three parts of an experience (the lead up, the offence itself, and the resolution) without using a series of specific questions. This coding procedure meant that coders only considered the open-ended questions and their immediate responses. The use of a 'frequency question' to ask about whether there were repeated similar incidents was coded as present when

asked appropriately ("Did X happen one time or more than one time?" or "Has this happened before?"). It was coded as absent if no questions about frequency were asked, or if the interviewer asked an inappropriate question such as, "How many times has it happened?". Variety of open-ended questions was coded as effective interviewing behaviour because these questions are more effective when varying stems are used (e.g. "Tell me more. What else happened. And then what happened.") as opposed to repeating the same stem. If the interviewer used at least 3 different open-ended question stems, variety was coded as present. If not, or if they used the same question stem more than 3 times in a row, variety was coded as absent. Ground rules were counted throughout the interview. If the interviewer used one or more ground rules it was coded as present. Finally, minimal encouragers were coded in the same manner as the ground rules. These utterances encourage the interviewee to continue providing information via non-verbal gestures (e.g., "uh-huh" or "mm-hmm") or repeating the last word. Table 2 presents an overview of coding categories, and Appendix B provides a full overview of interviewer behaviour coding.

Table 2. Overview of Interviewer Behaviour Coding Categories

Behaviour	Coding Rule or Criteria
Narrative elicitation	Interviewer elicits all three parts of the narrative (lead-up, offence, resolution) using open-ended prompts.
Frequency question	Asked appropriately: "Did this happen one time or more than one time?" (coded 0 if absent or inappropriate, e.g., "How many times...?").
Variety of open-ended stems	≥ 3 unique stems used and no more than 3 repetitions in a row.
Ground rules	At least one instruction clarifying rules (e.g., "Don't guess", "Correct me if I make a mistake").
Minimal encouragers	At least 3 standalone encouragers (e.g., "uh-huh"), not immediately followed by a question.

Reliability

In addition to the first author, two trained research assistants who were blind to the study and the hypotheses double-coded 20% of the transcripts for reliability purposes. Interrater reliability was high, with Cohen's kappa (κ) values of 0.968 (question type) and 0.792 (interviewer behaviours) between research assistant 1 and the first author and 0.975 (question type) and 0.846 (interviewer behaviours) between research assistant 2 and the first author. All remaining interviews were coded by the first author, who was blind to timepoint during coding.

Results

Assumption Checks and Analytic Plan

All analyses were conducted using SPSS (version 30.0). Before inferential analyses, data were examined for missing values and evaluated against relevant statistical assumptions. No missing data were observed. Both Shapiro-Wilk tests ($p < .05$) and visual assessments of Q-Q plots indicated significant deviations from normality in the proportions of open-ended and specific

question types at both pre-training (T1) and immediate post-training (T2) phases. Interestingly, at follow-up (T3), data for open-ended questions conformed to a normal distribution. Due to these initial normality violations, and in accordance with best-practice recommendations for small and non-normally distributed samples (De Vocht, 2023), non-parametric Friedman's two-way analysis of variance by ranks was applied.

The primary analytic approach involved within-subject comparisons across the three measurement points. Friedman tests assessed changes in the proportion of question types, with subsequent pairwise Wilcoxon signed-rank tests applying Bonferroni adjustments for post-hoc analysis. Additionally, interviewer behaviours—such as narrative elicitation, frequency questions, variety in open-ended questions, use of ground rules, and minimal encouragers—were analysed using Cochran's Q tests, suitable for binary data assessed at multiple time points (De Vocht, 2023).

Statistically significant outcomes were further explored with Bonferroni-adjusted McNemar tests. The McNemar test is the repeated measures equivalent of the Pearson

chi-square test and should be used to assess non-random change in dichotomous paired variables across two measurement points (Adedokun & Burgess, 2012). An alpha of .05 was set for all tests. Effect sizes were reported using Kendall's W, categorised as small (.10), medium (.30), and large (.50) (Cohen, 1988).

Question Types

Table 3 presents the mean scores and standard deviations for the proportion of

each question type used across the three measurement points (T1: pre-training, T2: post-training, T3: follow-up). Overall, participants demonstrated a clear shift towards recommended questioning techniques following the training.

Table 3. Means and Standard Deviations for Proportion of Question Types Across Timepoints.

Question type	Pre-commencement		Post-training		Follow-up	
	M	SD	M	SD	M	SD
Open-ended	.26^a	.13	.69^b	.16	.61^b	.16
Invitation	.02 ^a	.02	.03 ^a	.02	.03 ^a	.02
Breath	.11 ^a	.05	.23 ^b	.09	.19 ^b	.09
Depth	.14 ^a	.11	.32 ^b	.13	.30 ^b	.12
Descriptive	.05 ^a	.05	.12 ^b	.08	.09 ^b	.08
Specific	.68^a	.13	.31^b	.16	.39^b	.16
Cued recall	.41 ^a	.09	.17 ^b	.09	.18 ^b	.11
Yes/No	.20 ^a	.10	.10 ^b	.09	.12 ^b	.08
Verification	.04 ^a	.04	.04 ^a	.06	.04 ^a	.04
Forced choice	.02 ^a	.03	<.01 ^b	.01	<.01 ^b	.02

Note. Means within a row sharing the same superscript do not differ significantly. Different superscripts indicate significant differences based on pairwise Wilcoxon signed-rank tests with Bonferroni correction, following a significant Friedman's test ($p < .05$).

To formally assess these trends, a series of non-parametric Friedman tests was conducted on the question type variables to examine longitudinal changes in the use of open-ended questions across the three time points. A significant main effect was found for open-ended question use, $\chi^2(2) = 24.70$, $p < .001$, indicating that the proportion of open-

ended questions changed over the course of the training program. Pairwise comparisons (Wilcoxon signed-rank tests) with a Bonferroni-adjusted alpha of .017 revealed that trainees used a significantly higher proportion of open-ended questions at T2 (post-training, $Z = -3.88$, $p < .001$), and at T3 (follow-up, $Z = -3.73$, $p < .001$), compared to T1

(pre-training). There was no statistically significant difference between T2 and T3, $Z = -2.28$, $p = .023$.

The results further confirmed that the training had a significant effect on most question types. Specifically, there were statistically significant changes over time in the use of open-ended breadth [$\chi^2(2) = 14.40$, $p < .001$, $W = .36$], depth [$\chi^2(2) = 14.80$, $p < .001$, $W = .37$], with medium effect sizes, and descriptive questions [$\chi^2(2) = 9.10$, $p = .011$, $W = .23$], with a small to medium effect size. Likewise, significant reductions were observed in the use of specific question types, including cued recall [$\chi^2(2) = 26.66$, $p < .001$, $W = .67$], yes/no [$\chi^2(2) = 14.10$, $p < .001$, $W = .35$], forced-choice [$\chi^2(2) = 16.35$, $p < .001$, $W = .41$], and overall specific questioning [$\chi^2(2) = 18.92$, $p < .001$, $W = .47$]. Effect sizes were medium to large for specific question types. No significant differences were found for invitations [$\chi^2(2) = 4.42$, $p = .087$, $W = .11$], though the trend was in the expected direction.

To identify the specific time points at which changes occurred, Wilcoxon signed-rank tests were conducted for all question types with significant Friedman results.

Bonferroni correction was applied ($\alpha = .017$).

The comparisons revealed consistent and statistically significant improvements between T1 and T2 across all variables. In contrast, changes between T2 and T3 were generally non-significant, suggesting that training gains were retained rather than further developed. The differences between T1 and T3 remained significant, reinforcing the sustained impact of the intervention.

Interviewer Behaviours

For the dichotomous variables representing specific interviewer behaviours, Cochran's Q tests were used to assess change over time (see Table 4). Significant differences were found for narrative elicitation [$Q(2) = 28.35$, $p < .001$, $W = .71$], use of the frequency question [$Q(2) = 22.93$, $p < .001$, $W = .57$], variety in open-ended question use [$Q(2) = 20.57$, $p < .001$, $W = .51$], with large effect sizes, and the use of ground rules [$Q(2) = 10.67$, $p = .005$, $W = .27$], with a small to medium effect size. No significant change was observed for the use of minimal encouragers [$Q(2) = 2.00$, $p = .368$, $W = .05$], likely due to ceiling effects, as usage was high across all time points.

Table 4. Cochran's Q Test Results for Interviewer Behaviours Across Timepoints.

	Pre-commencement	Post-training	Follow-up		
Interviewer behaviours	Yes (%)	Yes (%)	Yes (%)	Q	p-value
Narrative elicitation	5 ^a	85 ^b	80 ^b	28.35	<.001
Frequency question	15 ^a	85 ^b	75 ^b	22.93	<.001
OEQ Variety	30 ^a	90 ^b	90 ^b	20.57	<.001
Ground rules	40 ^a	80 ^b	80 ^b	10.67	.005
Minimal encouragers	95 ^a	90 ^a	100 ^a	2.0	.37

Note. Means within a row sharing the same superscript do not differ significantly. Different superscripts indicate significant differences based on pairwise McNemar's tests with Bonferroni correction, following a significant Cochran's Q test ($p < .05$).

Pairwise McNemar's tests were performed as post-hoc analyses for the significant Cochran's Q results. With Bonferroni correction applied ($\alpha = .017$), statistically significant increases were confirmed between T1 and T2 for all four behaviours. No further change was detected from T2 to T3, indicating sustained effects of training. For instance, the proportion of participants eliciting a complete narrative increased significantly from T1 to T2 [$\chi^2(1) = 14.06$, $p < .001$], and this improvement was maintained at follow-up ($p = 1.000$). Similar patterns were observed for variety in open-ended questions and usage of an appropriate frequency question.

Discussion

The current study evaluated the impact of specialist investigative interview training on police officers' interviewing skills in adult sexual assault cases. The main aim was to assess skill acquisition and its maintenance over time following training, with a particular focus on the type of questions used and five effective interviewer behaviours (i.e., narrative elicitation, use of a frequency question, variety in open-ended questions, use of ground rules, and use of minimal encouragers). Overall, the results showed significant improvements in the use of open-ended questions and interview behaviours post-training, with effects largely sustained at follow-up. These findings are reviewed next.

Type of Questions Asked

In line with expectations, there was a substantial increase in the use of open-ended question types post-training. The medium

effect size suggests a practically meaningful improvement in interviewers' use of both breadth and depth questions, which persisted over time. This finding supports our hypothesis and aligns with prior studies demonstrating the effectiveness of incremental child interview training programs that focus intensively on question formulation and narrative facilitation (e.g., Benson & Powell, 2015; Brubacher et al., 2022; Powell et al., 2016). The increase in initial invitations was more modest, which is explained by the fact that interviews about single events typically only need one invitation per interview, meaning they are relatively independent of overall open-ended question use.

The sustainability of the improvements is particularly noteworthy. Performance at long-term follow-up was still significantly better than baseline, suggesting that the combination of didactic instruction, practical application, and performance feedback was associated with lasting behavioural change, at least in an experienced sample of interviewers. This supports broader evidence from longitudinal evaluations of interview training indicating that skills such as open-ended questioning can be retained up to nine months post-training, particularly when reinforced through repeated practice and mock interviews (Brubacher et al., 2022). Nevertheless, the training in the current study was comparatively short and involved just three practice attempts.

We also observed significant reductions in the proportion of specific questions asked. This decline suggests that participants were not only acquiring new skills but also replacing less-effective habits – an essential outcome in adult sexual assault investigations where narrative elaboration and coherence are difficult to obtain but paramount to

progress the case (Zajac et al., 2019). The medium to large effect sizes suggest that training produced practically meaningful improvements to interviewers' questioning practices.

Narrative Elicitation

The degree to which the narrative of the complainant was elicited by the interviewer increased significantly after training (large effect size) and remained stable thereafter, supporting our hypotheses. This outcome is particularly salient, as this skill reflects not just question structure but also pacing, listening, and the interviewer's ability to resist prematurely asking a series of specific questions. These findings are consistent with recent evaluations of interviewer training that emphasize sustained attention to the witness's unfolding narrative and the use of open prompts (Cyr et al., 2021; Lawrie et al., 2021). Moreover, the strong association between increased use of narrative-facilitating questions and greater narrative elicitation in this study supports the notion that training-induced behavioural change is meaningfully linked to information gain, not merely stylistic compliance (Zekiroski et al., 2024b).

Frequency Question

As sexual assault reports often represent repeated experiences (Snow et al., 2020), we assessed whether or not the interviewers asked an appropriate question about abuse frequency. Our findings showed significant increases in usage of a frequency question after training with a large effect size. Questioning multiple occurrences one by one assists in enabling the interviewee to move beyond generic memory into episodic recall of individual incidents, increasing the utility of the statement in legal proceedings (See Snow et al., 2020, for discussion on gist-based and verbatim memory retrieval).

Variety in Open-Ended Question Stems

A variety in the use of open-ended question stems, or the use of different open-ended questions throughout the interview, is considered more effective than repetitive use of the same open-ended question stem (Powell & Guadagno, 2008). The significant increase in variety of open-ended question stems directly after training with sustained effects at follow-up thus forms an important finding of this study. It shows a deepening of the impact of training and the effectiveness of questioning in adult sexual assault interview practice by going beyond solely asking open-ended questions instead of specific questions, to effectively deploying these newly acquired skills.

Use of Ground Rules

From prior literature, we know that ground rules – such as encouraging interviewees to admit uncertainty, correct the interviewer, or request clarification – are foundational to establishing the interviewee as the expert and reducing suggestibility and error (Brubacher et al., 2015). Although originally designed for child witness protocols (e.g., ABE, NICHHD), their relevance in adult contexts has been increasingly recognized (Benson & Powell, 2015; Dilevski et al., 2024; Fisher & Geiselman, 2010). In the present study, ground rule usage increased significantly from pre- to post-training and from pre-training to the long-term assessment. Here, the effect size was small to medium, in part because interviewers were already using ground rules to some extent prior to the present training. The findings suggest that such elements can be effectively introduced even to experienced professionals when training includes simulated mock interviews and immediate expert feedback.

Minimal Encouragers

Among the interviewer behaviours examined, minimal encouragers were the only one not showing a significant increase post-training. The proportion of minimal

encouragers used was already high at baseline and remained consistently high at both post-training and follow-up assessments. This pattern may reflect a ceiling effect, possibly due to the participants' prior experience—90% had received previous (specialised) interview training. Minimal encouragers are foundational active listening behaviours and are often internalized early in an interviewer's professional development (e.g., Mann, 2016). As such, their frequent use prior to training is unsurprising. It remains possible, however, that training influenced how minimal encouragers were employed—e.g., in terms of timing, context, or perceived authenticity—though these qualitative aspects were not captured in the present coding scheme. Future research might explore the strategic use of minimal encouragers or assess their function within broader interactional patterns to gain a deeper understanding of their role in effective interviewing.

Implications for Practice and Training

The current findings provide preliminary evidence that targeted, skill-based training – delivered as spaced learning in a modular and interactive online format – can significantly improve the quality of adult sexual assault interviews, even among experienced practitioners. Beyond the statistical significance, the observed changes are practically meaningful: enhanced use of open-ended questions and ground rules are directly linked to improved information quality in investigative contexts (Dilevski et al., 2024; Zekiroski et al., 2024a). Moreover, the sustainability of the effects, observed over a ten-week follow-up, addresses a persistent concern in training literature regarding skill decay (Akca et al., 2021).

Limitations and Future Research

Although the findings are encouraging, several limitations should be acknowledged. First, the study used standardised mock

interviews conducted via MS Teams; while scenarios were ecologically informed, adult sexual assault complainant interviews are typically conducted in person, which may limit ecological validity. Second, the relatively small sample size reduces statistical power to detect small effects and – together with the single geographic setting – may limit generalizability. A post-hoc power analysis for the non-parametric analyses suggested a minimum sample size of 27 would have been needed for at least 80% power, indicating the study may be underpowered. Third, causal inference is limited by the pre-post design without a control/comparison group. Fourth, the measurement of effective interviewer behaviours relied on dichotomous coding, which – while reliable – may not capture subtler gradations in behaviour. Fifth, we did not collect participant satisfaction /acceptability data or a precise count of officers' prior adult sexual assault interviews, which would contextualise perceived relevance and transfer. Despite these limitations, the findings are reflective of other training evaluations that have employed spaced and interactive learning and long-term follow-up assessments (Bos & Brand-Gruwel, 2017; Brubacher et al., 2022; Powell et al., 2016; Powell et al., 2022; Zekiroski et al., 2024a).

Future research should explore the long-term sustainability of these effects beyond three months, as well as investigate the impact of refresher sessions or supervision. Even if initial skill retention appears stable, refresher training and supervisory support may contribute to improved adherence, confidence, and adaptability in practice, and help prevent gradual skill erosion or the development of suboptimal habits over time. Moreover, evaluating these components can inform cost-benefit decisions and guide the design of efficient, targeted training programs. Research should also examine whether similar improvements can be replicated in high-stakes real-life interview settings by examining field interviews

(nevertheless, performance on mock and field interviews is correlated, e.g., Powell et al., 2010). It could be interesting to extend analyses to additional interaction skills (e.g., rapport management, pacing, empathic alignment) in future research. Further, qualitative feedback from participants could illuminate perceived barriers and facilitators to implementing training in operational contexts.

Conclusion

The current study suggests that a brief, modular training is associated with substantive and durable improvements in question formulation and key interviewer behaviours among experienced police officers, particularly in handling adult sexual assault interviews. The demonstrated improvements in narrative elicitation, open-ended question use, frequency question, variety in open-ended question stems, and application of ground rules, reflect substantial advancement in interviewing quality the efficacy of an online training approach. Our findings support growing evidence that quality incremental training activities can improve interviewing skills for vulnerable adult witnesses, which are sustained at least several months post-training.

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. <https://doi.org/10.56645/jmde.v8i17.336>
- Akca, D., Larivière, C. D., & Eastwood, J. (2021). Assessing the efficacy of investigative interviewing training courses: a systematic review. *International Journal of Police Science and Management*, 23(1), 73-84. <https://doi.org/10.1177/1461355721100847>
- 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>
- Bos, N., & Brand-Gruwel, S. (2017). Effectiviteit van Blended Learning. *Tijdschrift voor hoger onderwijs*, 35(1), 5-21. <https://tvho.archief.openjournals.nl/pdf/TvHO%202017%2035%201%202.pdf>
- Brubacher, S. P., & Earhart, B. (2019). Investigative interviewing about repeated experiences. In *Evidence-Based investigative interviewing* (pp. 216-233). Routledge.
- Brubacher, S. P. & Powell, M. B. (2025). Uses of technology to facilitate education in skill training: Introduction to the special issue. *Applied Cognitive Psychology*, Accepted June 8, 2025. <https://doi.org/10.1002/acp.70085>
- Brubacher, S.P., Poole, D. A., & Dickinson, J. J. (2015). The use of ground rules in investigative interviews with children: A synthesis and call for research. *Developmental Review*, 36(6), 15-33. <https://doi.org/10.1016/j.dr.2015.01.001>
- Brubacher, S. P., Sharman, S. J., Westera, N. J., Zekiroski, H., Danby, M. C., & Powell, M. B. (2020). Narrative practice may foster comfort but not enhance cognition in adult witness interviews about a mock sexual assault. *The Journal of Forensic Psychiatry & Psychology*, 31(5), 814-821. <https://doi.org/10.1080/14789949.2020.1799056>
- Brubacher, S. P., Shulman, E. P., Bearman, M. J., et al. (2022). Teaching child investigative interviewing skills: long-term retention requires cumulative training. *Psychology, Public Policy, and Law*, 28(1), 123-136. <https://doi.org/10.1037/law0000332>
- Cederborg, A-C., Alm, C., Da Silva Nisen, D, & Lamb, M.E., (2013). Investigative interviewing of alleged child abuse victims: an evaluation of a new training program for investigative interviewers. *Police Practice and Research. An International Journal*, 14, 242-254. <https://doi.org/10.1080/15614263.2012.712292>
- Cederborg, A. C., Lindholm, T., Lamb, M., et al. (2021). Evaluating the quality of investigative interviews conducted after the completion of a training program. *Investigative interviewing. Research and Practice*, 11(1), 40-52. <https://iirg.org/wp-content/uploads/2021/04/iir-RP-Journal-Vol-11-Issue-1-April-2021.pdf>
- Chenier, K., Milne, R., Smith, K., & Snook, B. (2021). Interviewing adult complainants in sexual assault cases. In: *Criminal Investigations of Sexual Offenses: Techniques and Challenges* (pp. 67-84). Springer. https://doi.org/10.1007/978-3-030-79968-7_6
- Clarke, C. & Milne, R. (2001) *National evaluation of the PEACE investigative interviewing course* (Police Research Award Scheme: Report No. PRAS/149). Institute of Criminal Justice Studies, University of Portsmouth. http://www.researchgate.net/profile/Colin_Clarke3/publication/263127370_National_Evaluation_of_the_PEACE_Investigative_Interviewing_Course/links/53da3b620cf2e38c63366507.pdf

- Cyr, M., Dion, J., Gendron, A., Powell, M., & Brubacher, S. (2021). A test of three refresher modalities on child forensic interviewers' post-training performance. *Psychology, Public Policy, and Law*, 27(2), 221.
<https://doi.org/10.1037/law0000300>
- De Vocht, A. (2023). *Basishandboek SPSS 29*. Utrecht: Bijleveld Press
- Dilevski, N., Brubacher, S. P., Ali, M. M., Earhart, B., & Powell, M. (2024). Adults report positive perceptions of ground rule instructions in mock investigative interviews. *Psychiatry, Psychology and Law*, 31(1) 1–21.
<https://doi.org/10.1080/13218719.2024.2404838>
- Faul, F., Erdfelder, E., Buchner, A., Lang, A-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41, 1149–1160.
<https://doi.org/10.3758/BRM.41.4.1149>
- Fisher, R. P., & Geiselman, R. E. (2010). The cognitive interview method of conducting police interviews: Eliciting extensive information and promoting therapeutic jurisprudence. *International Journal of Law and Psychiatry*, 33(5–6), 321–328.
<https://doi.org/10.1016/j.ijlp.2010.09.004>
- Hartwig, M., Dawson, E. C., Wrede, O., & Ask, K. (2012). Interviewing victims of repeated domestic violence: Investigators' beliefs and strategies. *Psychiatry, Psychology and Law*, 19(5), 672–681.
<https://doi.org/10.1080/13218719.2011.615814>
- Lamb, M. E., Brown, D. A., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2018). *Tell me what happened: Questioning children about abuse*. John Wiley & Sons. <https://doi.org/10.1002/9781118881248>
- Lawrie M, Brubacher SP, Earhart B, et al. (2021) Testing the effectiveness of a blended vulnerable witness training for forensic interviewers. *Journal of Family Trauma, Child Custody and Child Development* 18(3), 279–297.
<https://doi.org/10.1080/26904586.2021.1894303>
- 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>
- Mann, S. (2016). Training and Development. In: *The Research Interview*. Palgrave Macmillan.
https://doi.org/10.1057/9781137353368_10
- McMillan, L., & Thomas, M. (2009). Police interviews of rape victims: Tensions and contradictions. In M. Horvath & J. Brown (Eds.), *Rape: Challenging contemporary thinking* (pp. 255–280). Willan.
- Ministry of Justice (2022) *Achieving Best Evidence in Criminal Proceedings: Guidance on Interviewing Victims and Witnesses, and Guidance on Using Special Measures*. Ministry of Justice.
- Powell, M. B. (2013). An overview of current initiatives to improve child witness interviews about sexual abuse. *Current Issues in Criminal Justice*, 25(2), 711–720.
<https://doi.org/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*, 55(6), 645–659. <https://doi.org/10.1111/ap.12468>
- Powell, M. B., Cavezza, C., Hughes-Scholes, C., & Stooze, M. (2010). Examination of the consistency of interviewer performance across three distinct interview contexts. *Psychology, Crime & Law*, 16(7), 585–600.
<https://doi.org/10.1080/10683160902971063>
- Powell, M.B., Brubacher, S.P., and Baugerud, G.A. (2022). An overview of mock interviews as a training tool for interviewers of children. *Child Abuse & Neglect*, 129(2022), 1–11.
<https://doi.org/10.1016/j.chiabu.2022.105685>

Powell, M. B., Fisher, R. P., & Wright, R. (2005). Investigative interviewing. *Psychology and law: An empirical perspective*, 11-42.

Powell, M., Garry, M., & Brewer, N. (2009). Eyewitness testimony. In I. Freckelton. *Expert evidence*. (pp. 1-42). Lawbook Co.

Powell, M. B., & Guadagno, B. (2008). An examination of the limitations in investigative interviewers' use of open-ended questions. *Psychiatry, Psychology and Law*, 15(3), 382-395. <https://doi.org/10.1080/13218710802101621>

Powell, M.B., Guadagno, B., and Benson, M.S. (2016). Improving child investigative interviewer performance through computer-based learning activities. *Policing and Society*, 26(4), 365-374. <https://doi.org/10.1080/10439463.2014.942850>

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>

Snook, B., Luther, K., Quinlan, H., et al. (2012). Let'em talk!" A field study of police questioning practices of suspects and accused persons. *Criminal Justice and Behavior*, 39(10), 1328-1339. <https://doi.org/10.1177/0093854812449216>

Snow, M. D., Malloy, L. C., Brubacher, S. P., & Sutherland, J. E. (2020). Memory for repeated sexual misconduct: Does repetition matter?. In *Memory and Sexual Misconduct* (pp. 42-70).

Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9780429027857-3>

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207-232. [https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9)

Vrij, A., Hope, L., & Fisher, R. P. (2014). Eliciting reliable information in investigative interviews. *Policy Insights from the Behavioral and Brain Sciences*, 1(1), 129-136. <https://doi.org/10.1177/2372732214548592>

Yii, S.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>

Zajac, R., Westera, N. J., Ali, M. M., & Powell, M. (2019). Investigative interviews with adult sexual assault complainants: Challenges and future directions. In *Evidence-based investigative interviewing* (pp. 177-192). Routledge/ Taylor & Francis Group. <https://doi.org/10.4324/9781315160276-10>

Zekiroski, H., Powell, M. B., Brubacher, S. P., & Curtis, P. (2024a). Evaluation of a police training programme designed to enhance open-ended questions with adult witnesses. *International Journal of Police Science & Management*, 26(3), 317-329. <https://doi.org/10.1177/14613557241233857>

Zekiroski, H., Powell, M. B., & Cashman, K. (2024b). Police officers' perceptions of a training course designed to enhance open-ended questions with adult witnesses. *Psychiatry, Psychology and Law*, 31(1), 1-17. <https://doi.org/10.1080/13218719.2024.2346718>

Appendix A. Question type coding overview.

Question type	Brief description	Definition
Open-ended	Invitations	Invitations are the ideal way to begin a narrative, or new topic within the narrative. Invitations do not direct the complainant to talk about any type of specific information, but rather invite them to tell a narrative. E.g., “Tell me everything that happened last night. Start from the beginning.”
	Breadth	Breadth prompts are high quality questions that encourage complainants to provide more information about the event components or sequence of events. E.g., “What happened next?”
	Depth	Depth prompts are high quality questions that encourage complainants to provide more information about previously mentioned details. E.g., “You said X. Tell me more about X”.
	Descriptive	Open-ended descriptive questions are a type of depth prompt that requests further information about objects rather than actions. E.g., “Tell me about Bob.”
Closed	Cued-recall	Cued recall (also known as Wh-and How questions) are a type of specific question. They specify the type of information being sought from an interviewee, which restricts narrative flow and would typically elicit shorter responses from the victim. E.g., “Where did he hit you?”
Closed	Yes/No	Yes-no are questions that seek only “yes” or “no” responses. They are prone to guessing and do not encourage narrative. E.g., “Did he have anything to drink when he was out?”
Closed	Verification	Verification questions are also questions that seek only “yes” or “no” responses. The difference with the previous category is that these questions are asked to verify previous answers or information from the complainant. E.g., “Did you say he grabbed you with his left hand?”
Closed	Forced choice	Forced choice are questions that provide two or more response options. They are highly prone to guessing and do not encourage narrative. E.g., “Was this before or after you were cleaning the house?”

Appendix B. Interviewer behaviour coding overview.

Interviewer behaviour	Brief description	Definition
Narrative elicitation	Did the interviewer elicit the narrative?	When open-ended questions are used appropriately, they should be able to elicit all three parts of the narrative. This includes the lead up (e.g., what happened to any clothing if a sexual offence), the offence itself, and getting the complainant to a safe place (e.g., where the perpetrator is gone). Interviewers only get a 'Yes' if they elicit all three parts without using specific questions. Only look at direct responses to open-ended questions.
Frequency question	Did interviewer ask the recommended frequency question?	Interviewers should check with the complainant as to whether the event they are discussing has happened once or more frequently, so that they can question about multiple occurrences if needed/applicable. "Did X happen one time or more than one time?" or "Has this happened before?" Note: Code as zero if no frequency question is asked. Code as zero if inappropriate: "How many times did this happen?".
Variety of open-ended questions	Did the interviewer use a variety of open-ended questions?	Open-ended questions are more effective when a variety of question stems are used (e.g. "Tell me more. What else happened. And then what happened.") as opposed to repeating the same stem. Does the interviewer use at least 3 different open-ended question stems? If so; code "yes". If not, or if they use the same question stem more than 3 times in a row; code "no".
Ground rules	Were ground rules provided?	Ground rules are conversational expectations that are used in an interview to help clarify the roles of the interviewer and complainant. They include instructions not to guess, to alert the interviewer if (s)he makes a mistake, to tell when the complainant doesn't understand. They help the interviewee to feel more comfortable in the interview and put them in the role of the expert. If at least one ground rule is given, score with a 'yes'.
Minimal encouragers	Were minimal encouragers used effectively?	To be effective, minimal encouragers must be used alone, rather than followed immediately by a question (e.g., "Uh-huh. Then what?"). Interviewers must use 3 minimal encouragers appropriately anywhere in the interview to get a 'Yes'.

INVESTIGATIVE INTERVIEWING

Research & Practice

The official journal of the International
Investigative Interviewing Research Group



iIIRG

INTERNATIONAL
INVESTIGATIVE
INTERVIEWING
RESEARCH GROUP