Intermediaries and cross-examination resilience in children: The development of a novel experimental methodology

Journal:	*Applied Cognitive Psychology*
Manuscript ID	ACP-21-0013.R1
Wiley - Manuscript type:	Research Article
Keywords:	cross-examination, child witnesses, barristers, Registered Intermediaries, court

Intermediaries and cross-examination resilience in children: The development of a novel experimental methodology

Experimental studies examining child 'witnesses' under cross-examination typically rely on researchers questioning children using a 'barrister's script'. In the current research, experienced barristers used a defence statement from a mock perpetrator (who committed a theft 11 months earlier) to challenge typically developing children's evidence under cross-examination. We also assessed whether Registered Intermediaries (RIs), trained professionals who facilitate communication between vulnerable witnesses and members of the justice system, help children reduce compliance with misleading cross-examination suggestions. Results demonstrated that children (6-11 years) complied with barristers' challenges to a high degree: 94% agreed with at least one of the barristers' seven false suggestions. However, when assisted by an RI, children were significantly less compliant with barrister challenges. These findings, and additional analyses of the nature of child responses and barrister questions, provide novel exploratory evidence for the beneficial role of RIs in tempering the adverse effects of cross-examination style questioning for children.

Keywords: cross-examination, barristers, child witnesses, Registered Intermediaries, court

Introduction

In adversarial justice systems, such as England and Wales, child witnesses in criminal trials provide their evidence-in-chief (direct evidence) via video-recorded Achieving Best Evidence investigative interviews (Ministry of Justice, 2011). Subsequently, they may be questioned on this evidence by the opposing counsel ('cross-examination'), who has an interest in undermining this evidence. This can mean that witnesses, "having first been questioned by someone who wants them to say one thing…are then cross-examined by another person who wants to make them say the opposite" (Spencer, 2012, p.1). Here, we report the development of a novel experimental methodology to investigate cross-examination performance in typically developing children. We also assess whether providing child witnesses with a 'Registered Intermediary' (RI; a trained professional who facilitates communication between vulnerable witnesses and members of the justice system, Ministry of Justice, 2020a) improves the quality of children's evidence, by reducing compliance with barrister challenges about false information.

Recommendations of the Pigot Commiteee (Home Office, 1989) led to legislation in England and Wales that enabled, with the agreement of the court, vulnerable and intimidated witnesses to benefit from 'special measures' (Youth Justice and Criminal Evidence Act, 1999). These included: screens (preventing the witness from seeing the defendant); live links (enabling the witness to give evidence during the trial from outside the court room via a televised link); the removal of wigs and gowns (by judges and barristers); pre-recorded video evidence-in-chief and cross-examination; use of aids for communication (enabling questions or answers to be communicated to or from the witness); and examination of the witness assisted by an RI. Although most of these recommendations have since been fully

implemented in England and Wales (the jurisdiction relevant to the current study), live-link cross-examinations were retained[1].

Improving the quality and reliability of children's evidence under cross-examination is an urgent international priority given serious concerns about how child witnesses are treated in criminal courts (e.g., Andrews, Lamb, & Lyon, 2015a; Spencer, 2012; Zajac, O'Neill, & Hayne, 2012). Studies of court transcripts (e.g., Australia, England, New Zealand, Scotland, USA) highlight that large proportions of questions posed to children during cross-examination are inconsistent with best practice guidelines and developmental level, with heavy reliance on closed, option-posing, suggestive (leading), repeated, and complex questions (e.g., Andrews & Lamb, 2016; Andrews et al., 2015a; Andrews, Lamb, & Lyon, 2015b; Evans, Lee, & Lyon, 2009; Hanna & Henderson, 2018; Hanna, Davies, Crothers, & Henderson, 2012; Henderson & Lamb, 2019; Henderson, Andrews, & Lamb, 2019; Klemfuss, Quas, & Lyon, 2014; Zajac, Gross, & Hayne, 2003; Zajac & Cannan, 2009). Suggestive questions are particularly problematic, as the likelihood of errors increases with their use (Lamb, Malloy, & La Rooy, 2011). Such questions "should only be used as a last resort and only when necessary (e.g., to immediately safeguard a person)" (Bull, 2010, p. 9), yet they are commonly *recommended* to advocates to maintain control of the discourse (Hanna et al., 2012). This illustrates the conflict between the aims of cross-examination (to test evidence) and best practice guidelines (to elicit evidence) (Zajac et al., 2012). Indeed, some have called cross-examination "a virtual 'how not to' guide to investigative interviewing" (Henderson, 2002, p. 279), directly violating methods that promote

[1] In 2014, a pilot programme of video-recorded live-link cross-examinations in England was trialled (Baverstock, 2016), involving pre-trial Ground Rules Hearings (which can place restrictions on traditional cross-examination practices to improve witness experiences) and video-recorded cross-examinations (to reduce delays between giving initial evidence and cross-examination in court). The scheme has now been rolled out to all Crown Courts across England and Wales. Henderson et al. (2019) and Henderson and Lamb (2019) evaluated cases with and without pre-trial Ground Rules Hearings prior to pre-recorded children's cross examination. With these measures, fewer suggestive questions were asked, and question complexity was reduced.

completeness and accuracy (Zajac et al., 2012) and exploiting children's vulnerabilities (Henderson et al., 2019). Almost 90% of witnesses under 11-years do not understand questions they are asked at court (Plotnikoff & Woolfson, 2009). Further, almost 95% of cross-examination transcripts of child sexual abuse cases reveal inconsistencies, largely between what is said in police interviews relative to subsequent cross-examination (Pichler et al., 2020). Worryingly, a comparative study of child sexual abuse case transcripts in Australia found no improvements in the format of questions used over the past 60 years (leading questions still predominated), with *more* questions asked, which were *more* likely to be complex (Zajac, Westera, & Kaladelfos, 2018).

Empirical studies of cross-examinations support these findings, noting that high numbers of children change their responses following questioning. In children of 4-11 years, 70-98% changed at least one aspect of their testimony when challenged (e.g., Bettenay, Ridley, Henry, & Crane, 2014; Righarts, Jack, Zajac, & Hayne, 2015; Zajac & Hayne, 2003, 2006; Zajac, Jury, & O'Neill, 2009). Most previous empirical studies employed researchers challenging witnesses by asking scripted cross-examination questions, although occasionally trainee legal professionals have been used (e.g., Bettenay et al., 2014). Yet, it is more realistic to allow barristers free reign to tackle cross-examinations in the way they see fit. In the present study, an unscripted approach was used to assess cross-examination compliance in children, enabling barristers to adapt according to the way a child responded, and to press points more emphatically if they were making headway, which is not possible using a script.

The study also investigated whether one of the special measures, the Witness Intermediary Scheme (available in England and Wales since 2004), would help reduce child witnesses' compliance with barrister challenges about false information. The role of RIs is wide-ranging but includes assessing the communication abilities of vulnerable witnesses and offering impartial and specific advice on posing best practice questions by accommodating

each individual child's language and communication needs. The aim is to facilitate communication between the child and relevant professionals to ensure it is complete, coherent and accurate (Collins & Krahenbuhl, 2020; Krahenbuhl, 2019; Cooper & Wurzel, 2014). Several other international jurisdictions (e.g., Northern Ireland, New Zealand, Norway, New South Wales, Australia) have adopted intermediary schemes, although details of the schemes vary (see Cooper & Mattison, 2017; Cooper & Wurzel, 2014; Taggart, 2021). Feedback on the RI scheme has been generally positive (Collins & Krahenbuhl, 2020; Ministry of Justice, 2020a; Plotnikoff & Woolfson, 2015), and mock juror studies suggest that the presence of an RI does not have a negative impact on perceptions of child witnesses (e.g., Krahenbuhl, 2019). However, further empirical evidence in relation to RI use during mock cross-examinations is needed and the current study offers exploratory evidence in this regard.

The current study forms part of a broader research programme examining child witness performance during all stages of a mock criminal investigation: initial statements (Blinded for peer review); investigative interviews (Blinded for peer review); identification line-ups (Blinded for peer review); and cross-examinations (presented here). Children viewed a staged event involving a minor mock crime (in which one man 'stole' another man's phone or keys) and were cross-examined on this evidence approximately 11 months after undergoing initial investigative interviews (representing close to the average delay of eight months for a case to go to trial in England and Wales at the time of the study; Plotnikoff & Woolfson, 2012). Qualified, experienced barristers took on the role of the defence barrister and were presented with a defence statement with which to question the children, allowing the barrister to adopt an unscripted approach.

The first primary research question was whether, and to what extent, children would comply with the barrister's challenges on seven elements of false information in the

statement. A second primary research question considered whether providing child witnesses with RI assistance reduced compliance with the barrister's challenges on this false information (a proportion of our sample was assisted by a fully qualified, experienced RI at all stages of giving formal evidence). Given the lack of previous empirical evidence, predictions were tentative. We hypothesised that: (1) children would comply to a large degree with barrister challenges on false information; and (2) a beneficial effect of RI assistance on compliance with false information on cross-examination challenges would emerge, as RIs facilitate communication, for example, rephrasing questions in a developmentally appropriate manner in line with an individualised communication assessment. Two subsidiary research questions were also addressed: (3) in RI assisted cross-examinations, would children's responses show less compliance (and more resistance) to challenges on false information?; and (4) in the RI condition would barristers change the style and nature of questions in line with the recommendations given for questioning (based on each child's communication assessment and according to best practice for interviewing young children)? We tentatively predicted that children in the RI condition would be less likely to comply with, *and* more likely to resist, challenges on false information; and that barristers would ask more questions in the RI condition consistent with best practice. The broader research programme included a control interview condition (Best-Practice) and two other interview conditions (Sketch-Reinstatement of Context and Verbal Labels). We did not expect the two other interview conditions to differ from the Best-Practice condition in terms of cross-examination resistance or nature of responses/questions.

Method

Participants

A total of 202 typically developing children were recruited from mainstream primary schools in London and the Southeast of England, but three were excluded: one had a full-

scale IQ in the intellectual disability range; and two were unavailable for the investigative interview (see Henry et al., 2017a, for further details). Of the remaining 199 children, 177 (84 boys, 93 girls) were available for cross-examination 11 months later (range 8-13 months). At this stage, one further child (a girl) was excluded because she did not respond to any cross-examination questions. The remaining 176 children ranged in age from 6 years 7 months to 11 years 3 months (mean = 8 years 6 months, SD = 1 year 2 months) at the time of the initial investigative interview; and 7 years 7 months to 12 years 3 months (mean = 9 years 5 months, SD = 1 year 2 months) at the cross-examination stage. See Table 1 for details.

[insert Table 1 about here]

Materials and Procedure

As described, this research was part of a wider project exploring the performance of child witnesses across different stages of the criminal justice process (children on the autism spectrum were included, but we were unable to cross-examine enough children to ensure reliable findings with this group). Of relevance to the current paper, were three phases.

Phase 1 – Staged event and evidence gathering statements ('Brief Interviews'). Children watched a staged event (either live or on video[2]) of two men delivering a short talk about what school was like a long time ago. As well as telling the children a series of facts about Victorian schooldays and showing them some equipment (e.g., an abacus, a slate), a minor theft occurred in which one of the men 'stole' the other's keys/phone[3]. For ethical reasons this was a mild minor crime event. Immediately after the event, the children were

[2] 144 children saw the event live and 32 children saw it via video. A *t*-test on number of correct details recalled in the brief evidence-gathering statement across these two groups was non-significant: Mean live = 33.82 (SD = 14.84); Mean video = 38.94 (SD = 14.17), $t(174) = 1.78$, $p = .08$. Nevertheless, we ran our primary analyses on both the full sample and the live-only sample to ensure this variable did not affect the findings.

[3] Two versions of the event differed slightly in terms of names used (Alex/Adam, Max/Mark), objects shown (abacus/slate), and prop 'stolen' (keys/phone). No differences emerged in the number of correct details recalled in the brief evidence gathering statement across these two versions for the current sample: Mean Version A (n=87) = 34.03 (SD = 12.94); Mean Version B (n=89) = 35.45 (SD = 16.49), $t(174) = .63$, $p = .53$. Nevertheless, we controlled for this variable in our primary analyses.

questioned individually about what they saw, in a brief evidence gathering statement that began with the open question: "Tell me what you remember about what you just saw" and was followed (if necessary) by prompts asking about who was there, what the people looked like, when it happened and where it happened (see Henry et al., 2017a, for further information).

Phase 2 – Investigative Interviews. Approximately one week later, children took part in one of four types of investigative interview.

Best-Practice. Based on Achieving Best Evidence principles (Ministry of Justice, 2011), this interview comprised seven key phases: (1) greet and personalise the interview; (2) rapport building (chatting to the child about areas of interest); (3) truth and lies exercise (e.g., determining whether the child correctly responds to a statement along the lines of 'that lady is wearing a blue jumper' when it is red); (4) explain the purpose of the interview; (5) free recall (recall attempt 1 – 'Tell me everything you can remember about what you saw'); (6) questioning (recall attempt 2 – using open questions based upon what the child had already recalled); and (7) closure.

Registered Intermediary (RI). Here, children were supported by one of two experienced, practising RIs. Prior to the interview, the RI individually assessed each child and there was a meeting between the RI and each interviewer to discuss recommendations for the interview and to flag any individual needs. RIs advised the interviewers to follow the protocol for the Best-Practice interview, with some adaptations (e.g., simplifying the verbal instructions given to the children, and recommending the use of visual aids that were provided by the RIs). At all times, the RI was present to facilitate communication between the child and the interviewer. As the interviewer proceeded through the Best-Practice interview protocol, the RI intervened when appropriate to facilitate effective communication (verbally or by suggesting the use of suitable props).

Verbal Labels. This followed the procedure for the Best-Practice interview except that, following phase 5 (free recall), witnesses received 'tell me more' prompts in relation to four key areas (adapted from Brown & Pipe, 2003): (1) the people in the event; (2) the setting where the event took place; (3) the objects that were involved and what happened with them (actions); and (4) what the people said.

Sketch-Reinstatement of Context (Sketch-RC). This followed the procedure of the Best-Practice interview except that, prior to phase 5 (free recall), witnesses were instructed to think about the event and draw whatever reminded them about it, as well as what happened. Witnesses were asked to explain to the interviewer what they were drawing. After finishing their sketch, children were asked to give a free recall account of what happened (as per the Best-Practice interview) and were told they could use their drawing to point out or explain things (Dando, Wilcock, & Milne, 2009).

Phase 3 - Cross-examination. Prior to the cross-examination, children were 'refreshed' on their evidence as per Achieving Best Evidence guidance (Ministry of Justice, 2011) and the Registered Intermediary Procedural Guidance Manual (Ministry of Justice, 2015). This is standard practice for witnesses in advance of cross-examination within courts in England and Wales. Therefore, as in real-life, cross-examination performance may draw upon original memories of the event and recent memories of the refreshed interview. The researcher visited the child to explain that, in the next day or so, they would be speaking to a barrister who would ask them some questions about the staged event they previously saw. The researcher explained that the child would be listening to the audio of their interview[4], to remind them of the event and what they had said. After refreshing of the evidence, the researcher again reminded the child about the forthcoming cross-examination.

[4] We did not have permission to video all children, although we did have permission to audio record all children, therefore, audio recordings were used to refresh children on their evidence.

A team of six barristers was recruited for the cross-examinations, comprising four men and two women. Five were currently practising barristers, whilst one was no longer practising but had their own legal business. Barristers had between 5-21 years of criminal law experience (mean=15.2 years).

Cross-examination – a new methodological approach. For the cross-examination, a 'defence statement' was developed for each version of the staged event, which the barristers were asked to put to the children. This created a more realistic situation in which the barrister was representing a defendant in relation to a charge of theft. The defence statement (and the cross-examination protocol) was developed with the advice and guidance of an experienced barrister. The first two items in the statement included correct information designed to set the scene, establish rapport with the child witness, and make them feel at ease. The remaining points contained an element of untruthfulness (except for points 6 and 7, which were included so children did not feel that they were disagreeing with all the points the barrister was raising). Table 2 provides a sample defence statement for one version of the event.

[insert Table 2 about here]

Barristers were asked to challenge the child on all seven of the 'false' points (e.g., *"I think you've got a little bit mixed up because it wasn't the phone that Adam put in his pocket, it was the keys, wasn't it?"*) a maximum of four times (a decision, in consultation with one of the barristers, to avoid ethical concerns). As there was variability in this (based on barrister judgement), scores only reflect whether a child complied immediately, following challenge/s, or not at all. If the child complied with the challenge on first time of asking, they received a resistance score of 0; if they complied with a challenge on the second or subsequent time of asking, they received a resistance score of 1; and if they did not comply at all, they received a maximum resistance score of 2. Average resistance scores on each of the seven false points could range from 0-2, with higher scores indicating higher cross-examination resilience (i.e.,

lower compliance with false statements). On a few occasions, barristers judged that it was not necessary to pose all challenges to the children. In real life, barristers make judgements about how much/little to press a witness and do not take a fixed approach, so the present study aimed to reflect this. Therefore, mean resistance scores were calculated for each child based on the total number of challenges given.

We were careful to code the child's original recall of information pertaining to each of the seven false points (taken from the investigative interview), so this score could be controlled in the analyses. These 'memory trace' scores were allocated for full (3), moderate (2), partial (1) or no (0) knowledge about six of the false points in terms of degree of information recalled in the investigative interview. For one other point (false point 5), this was a complete confabulation about something that did not happen at all in the event, therefore, a score of 0 was allocated for all children because it was not possible to code this item in terms of original recall of information (Maximum memory trace score=18: see Table 1 for mean memory trace scores and Supporting Information for full details of the coding scheme). Fifteen percent of the transcripts were independently coded by a second rater for memory trace scores and intra-class correlations for information pertaining to each of the challenges ranged from .89 to 1.00, indicating excellent inter-rater reliability.

Cross-examination – the study protocol. One special measure available to support vulnerable witnesses in courts in England and Wales is the 'live link'. The child is not present in the courtroom with the barristers, judge or jury, but is in a separate room. Those in the courtroom see the child via a television screen, and the child can see the judge or barrister on his/her screen. To mimic this, cross-examinations were performed using video conferencing software (Skype). A female researcher was in a room with the child at their school and partially took on the role of 'judge'. We could not entirely replicate the judge role as we had no facility for the child to view the judge only via the screen - and for ethical reasons the

researcher had to be with the child - so this aspect of the study must be viewed as approximate to real-life. There was a brief 'ground rules hearing' between the judge and the barrister prior to each individual cross-examination (with or without an RI) where the judge explained any important considerations to the barrister (e.g., age of child, any additional needs they had). As a prelude to the cross-examination, the judge explained to the child that they: (1) needed to tell the truth – must not guess or make anything up; (2) could say that they 'don't know' or 'cannot remember'; (3) should say if they do not understand something the barrister says; (4) could tell the barrister if they get something wrong; and (5) should say if there is a problem of any kind (as per the Judicial College Bench Checklist: Young Witness Cases, 2012). The judge also described the role of the barrister, explaining that they would be asking the child questions about what happened during the staged event. The judge added that the job of the barrister was to test the evidence, so they may ask questions that challenge what the child has said, but all the child needed to do was tell the truth about what they could remember or say if they did not know the answer. Whilst judges are advised to explain how often breaks are planned, and to inform the child that the judge can always see them via live link (even if they cannot see the judge), these elements were not incorporated in the instructions as: (a) the cross-examinations were short, and breaks would not be needed; (b) the judge was already in the room with the child.

Once the child and barrister were introduced, they listened to the child's audio of their investigative interview together, so everyone could hear it (barristers were provided with a transcript of the children's testimony, as well as basic demographic information, in advance of the cross-examination, to enable them to prepare their questions; in real-life, they would have access to the child's evidence-in-chief in advance of the refreshing of the evidence). The barrister then began questioning the child, with the only stipulations being that they were to cover all points on the defence statement (unless the child appeared to show any signs of

distress), and that – for ethical reasons – they were not to excessively challenge the child on their testimony (no more than four challenges per point).

At three time points (before, during and after the cross-examination), children were presented with a ten-point visual analogue rating scale. This enabled us to monitor how worried or anxious the children were (1 = no anxiety; 10 = high anxiety) and to offer additional support or reassurance if their responses highlighted that they were affected by the cross-examinations. Note that these anxiety ratings were not study variables but introduced for ethical reasons. Most children were not highly anxious at any point. Before the cross-examination, 7 children (4%) had scores at the top end of the anxiety scale (8, 9, 10); during the cross-examination this figure was 9 children (5%); after the cross-examination nearly all (171 children, 97%) had the lowest anxiety scores of 1, 2, or 3 (and the remaining 5 children had moderate scores of 4, 5 or 6). Cross-examinations were, on average, 8.56 minutes long (SD = 2.24 minutes, range 3.53 minutes to 16.25 minutes).

Cross-examination protocol – the RI condition. The protocol for the cross-examinations was the same across three interview conditions (Best-Practice, Sketch-RC and Verbal Labels), but there were some differences for the RI condition. As per recommendations for best practice in England and Wales at the time of the study (Registered Intermediary Procedural Guidance Manual, Ministry of Justice, 2015), children received RI assistance both at their initial interview and again at cross-examination. Of the 33 children in the RI condition, 18 were assisted by the same RI at both stages, which is also recommended best practice, and 15 had a different RI at cross-examination (although using exactly the same protocol). In real cases there is also likely to be some variability in whether the same RI is available for both stages. RI assistance involved the following: Prior to the cross-examination, all children were re-assessed by the RI to ensure that information about the child's communication needs (originally collected 8-13 months previously) was up-to-date

and accurate. This re-assessment took place at least a week before the cross-examination and consisted of: (1) re-establishing rapport with the children; (2) explaining what would happen in the cross-examination; (3) checking the children could say they 'don't know' or 'cannot remember', and could state whether the barrister (adult) was wrong or right; (4) checking the children could respond to questions beginning with, for example, 'when' or 'how'; and (5) preparing simplified instructions for the judge to present during the preamble before the cross-examination (to make them easier to follow and remember). The barristers and RIs also met together for a dedicated 'ground rules hearing' (see Cooper, Backen & Marchant, 2015, for further details) prior to all RI cross-examinations, in which the RIs explained what their role was and discussed their recommendations with the barristers. In real-life, ground rules hearings would take place for each individual child. However, the RIs noted that many of their recommendations would be the same for most children in the study, so one overall ground rules hearing was conducted (with RIs flagging individual cases where necessary). [Note that this was in addition to the 'short' ground rules hearing for each individual child just before the cross-examination (regardless of interview condition).]

At the ground rules hearing, RIs discussed the principles of questioning and gave barristers a written summary of their suggestions. The summary included advice to: practice the live link prior to the child coming into room; use a short and simple preamble; be careful about references to do with time (e.g., when, how long), or questions requiring a number in the answer (e.g., how many); use a slow pace; allow thinking time; use short sentences with only one point per question; use basic vocabulary and sentence structure; and use names the child knows people by. Question types were discussed and RIs recommended avoiding questions that: were negatively phrased; were statements with a questioning intonation; were tagged (e.g., 'Max forgot his coat, *didn't he*?'); had an answer implied; and were repeats of already asked questions. The RIs additionally: reviewed each barrister's list of cross-

examination questions and highlighted the specific needs of individual children prior to cross-examination sessions (discussions by phone or email); reminded barristers that visual materials were available if needed to support expressive language (drawing materials, small world figures/furniture) and sequencing of events (post-it notes, timelines); and brought along calming objects so they were available to the children if necessary. Importantly, RIs did not intervene about the content of the questions but rather the format (Ministry of Justice, 2015), for example "[Barrister's name], could that question be rephrased, as you know it's a tagged question" or if they thought the child would not understand the question, for example, "I am not sure [child's name] will understand that complex question". In the RI condition, an RI was present alongside each child for every cross-examination, simplified the instructions given to the children by the judge, and made interventions during the cross-examinations as required. For example, if the barrister moved away from planned questions or began to use statements with tags, the RI would remind the barrister of best practice. The RI also intervened if the child appeared not to understand or follow the questioning.

Coding child responses and barrister questions. Children's responses were coded into mutually exclusive categories reflecting whether they complied, resisted, did not respond, responded with an open question, or sought clarification (see Table 3). When a child responded with an acknowledgement (e.g., 'okay'), this was not coded as a response to the question. If the child said they were not sure, this did not mean they had complied: children were instructed to say 'don't know' if this was the case, so they were resisting the barrister's attempts to get them to agree with them.

[insert Table 3 about here]

Barrister questions were coded into one of seven overarching mutually exclusive primary categories (see Table 4 for details). All questions (as well as non-content-based utterances which were given the code 'other') were coded separately, even if they occurred,

sequentially, e.g. "*That's really helpful, thank you very much* (code=other). *Okay, now they talked to you about Victorian schools* (code=assertion, true). *Did they tell you lots of things about what happened in Victorian times?* (code=invitation closed, true)" would attract three codes as indicated. Barrister questions were additionally coded for each instance of 17 other secondary features (see Table 5), which were not mutually exclusive categories, i.e., a question could challenge credibility as well as contain a tag. The coding systems were developed by looking at guidance on questioning available at the time (May 2015) in The Advocate's Gateway (Toolkit 6, 2015), the Judicial College Bench Checklist: Young Witness Cases (2012), and the Equal Treatment Bench Book (Judicial College, 2013). We also used an iterative process of discussion and reflection on the coding process to capture all question types in one overarching primary code, yet additionally reflect other relevant question features within the secondary codes. The classification system was designed to be as comprehensive and informative as possible, although it could not capture more subtle features such as intonation.

Reliability of coding. To establish coder agreement, 10% of scripts were coded independently by a second coder. Overall percentage agreement was 91% (range 86-100%) for the child codes, 89% (range 82-92%) for the barrister primary codes and 88% (range 81-100%) for the barrister secondary codes, all of which represented moderately high agreement.

[insert Tables 4 and 5 about here]

Control measures. Around the time that the children took part in Phases 1 and 2 of the study, several cognitive measures (intelligence, language, memory, attention) were administered to ensure factors that may affect eyewitness recall and cross-examination were controlled or matched between interview groups (see Table 1 for differences between conditions that were controlled for statistically). ***Intelligence.*** Two subtests (Vocabulary and Matrix Reasoning) of the second edition of the Wechsler Abbreviated Scale of Intelligence

(WASI-II; Wechsler & Zhou, 2011) were used to provide an assessment of intellectual ability and to establish suitability for entry into the study. ***Language.*** The British Picture Vocabulary Scale Third Edition (BPVS-3; Dunn, Dunn, & Styles, 2009) was used to provide a measure of receptive vocabulary. Two subtests (Sequencing, and Grammar and Syntax) of the Expressive Language Test 2 (ELT-2, Bowers, Huisingh, LoGiudice, & Orman, 2010) assessed narrative ability and grammatical morphology, respectively. Two subtests (Recalling Sentences and Formulated Sentences) of the Clinical Evaluation of Language Fundamentals, 4th edition (CELF-4 UK; Semel, Wiig, & Secord, 2006) provided an assessment of the ability to recall and formulate grammatically correct, meaningful sentences. ***Memory.*** Subtests from the Test of Memory and Learning 2 (TOMAL-2; Reynolds & Voress, 2007) were used to provide a composite memory measure, comprising both verbal ('Memory for Stories' and 'Paired Recall') and non-verbal ('Facial Memory' and 'Visual Sequential Memory') memory. ***Attention.*** The Test of Everyday Attention for Children (Tea-Ch; Manly, Robertson, Anderson, & Nimmo-Smith, 1999) was used to assess a range of relevant attention skills: selective/focused attention (the 'Sky Search' subtest); sustained attention (the 'Score!' subtest); and sustained-divided attention (the 'Sky Search Dual Task' subtest).

General procedure

Ethical approval was obtained from the relevant university Research Ethics Committee. Prior to participation, written consent was obtained from parents, and children also gave their own written assent to participate. At the start of Phase 1, children viewed the staged event and immediately took part in the Brief Interviews (Blinded for peer review). Phase 2, Investigative Interviews (Blinded for peer review) and Identification Lineups (blinded for peer review), took place around one week later. Cognitive testing also took place around this time, which was split over several sessions to fit in with school timetables and to ensure children remained engaged with tasks. Phase 3, the cross-examinations, took

place 8-13 months (Mean = 11.06 months, SD 1.69 months) after viewing the staged event. As some variability in this delay emerged across conditions (see Table 1) due to timing of school holidays and availability of RIs/barristers, we controlled for delay in the primary statistical analyses. All children were refreshed on their evidence in one session with the researcher, before the researcher returned at least one day later to conduct the cross-examination with the barrister. Children in the RI condition were re-assessed in a session prior to the refreshing of their evidence (on a different, earlier day). The RI was always present at the cross-examination and, beforehand, used a visual aid to explain to the child that they should only say what really happened, that if the barrister got something wrong, they could tell them, and equally that it was OK to say that the barrister 'got it right'. In addition, the children were told, using the visual aid, that it was OK to say 'I don't know', 'I can't remember', or 'I don't understand'.

Results

The key outcome measures for the primary research questions concerned: (1) children's cross-examination resistance scores on seven cross-examination challenges pertaining to false elements from the defence statement; and (2) whether RI assistance during cross-examinations reduced children's compliance with these challenges on false information.

Table 6 shows mean resistance scores (SDs). Ten children resisted all seven challenges on false information that the barrister put to them (5.7%), meaning that 94.3% of children complied with at least one challenge. Five children complied with all seven challenges on false information (2.8%).

[insert Table 6 about here]

Hierarchical multiple regression was used to examine whether cross-examination resistance scores on the seven false information challenges differed between children in the

RI condition versus other conditions (note that we had no reason to expect cross-examination differences for the Sketch-RC and Verbal Labels conditions as they involved adaptations to investigative interview protocols). At step 1, three background variables showing differences between interview conditions (see Table 1 for details), namely age at cross-examination, IQ, and Verbal Memory, were controlled (BPVS scores also differed between interview conditions, but IQ and BPVS scores were highly correlated, r=.66, so only IQ was controlled). Three additional control variables included: memory trace scores (concerning relevant information pertaining to the false information challenges) as children in the RI condition had higher memory trace scores (they had benefitted from RI intervention at the investigative interview stage) (Henry et al., 2017b); event version (A or B); and length of delay before cross-examination (this differed across condition – see Table 1). At step 2, three dummy-coded interview condition variables were included to test for differences between conditions in cross-examination resistance. Best-Practice was the reference (control) group to which the other three conditions were compared: RI, Sketch-RC and Verbal Labels. The dependent variable was average cross-examination resistance score (see Table 6). With nine predictor variables in total, Green (1991) would recommend a sample size of at least 122, thus for the current regression our sample size exceeded the minimum numbers recommended. Key statistical checks (multicollinearity, Durbin-Watson, tolerance and VIF statistics, Cook's and Mahalanobis distances, standardised DFbetas, leverage values, plots of standardised residuals and predicted standardised values, standardised residuals, partial plots) were within acceptable limits (Field, 2013).

Table 7 gives details of the regression. The full regression model was significant, $F(9, 166) = 5.37$, $p<.001$, accounting for 22.5% (18.3% adjusted) of the variance in cross-examination resistance scores. Step 1 was significant (R^2 change = 7.7%; $F(6, 169) = 2.35$, p=.03), indicating that the six control variables accounted for a small proportion of the

variance when entered on their own (although only memory trace was significant when inspecting standardised Beta values, Beta = .16, p=.04). Crucially, Step 2 was also significant (R^2 change = 14.8%; F(3, 166) = 10.61, p<.001), indicating interview condition differences in cross-examination resistance. Inspection of the standardised Beta-values at Step 2 showed that only the contrast between the RI and Best-Practice interview conditions was significant (Beta = .47, p<.001). As tentatively predicted, children in the RI condition were less compliant with cross-examination challenges than children in the Best-Practice condition, with higher resistance scores (an average of .63 out of 2 higher with a 95% CI of .37-.88), once all other variables had been accounted for. All other variables were non-significant predictors at Step 2. To check whether initial viewing of the event live or via video affected the findings, this regression was repeated with only children who had seen the event live (n=144). The results were identical in all respects, except that memory trace score at Step 1 just missed significance (p = .055)[5].

[insert Table 7 about here]

Children's responses

The first subsidiary research question had two components: first, whether the numbers of compliant responses by children to barrister challenges on false information would be lower in RI interviews; and second, whether the numbers of resistant responses by children to barrister challenges on false information would be higher in RI interviews. Whilst children gave, on average, 46.40 (SD = 14.31) responses across the cross-examination, this differed across interview conditions, F(3, 172) = 3.10, p=.03, partial η^2 =.05. Bonferroni corrected paired comparisons indicated that children gave significantly more responses in the RI

[5] Results were similar when barrister was included as a further control variable – the only significant predictor at Step 2 was the contrast between the RI and Best-Practice interview conditions (p<.001). At Step 1 memory trace (p=.03) and barrister (p=.01) were significant predictors. However, this analysis is only exploratory because not all barristers were evenly spread across conditions.

condition (mean = 52.82, SD = 13.05) than in the Best-Practice condition (mean = 43.82, SD = 11.40) (p=.02), but no other comparisons were significant. Given this, subsequent analyses were carried out on proportional scores (proportions of each type of response in relation to total number of responses for each child). Table 6 includes mean proportions of the seven types of responses.

Proportional data were not all normally distributed, so Kruskal-Wallis tests were used to explore whether there were differences between interview conditions for each type of response, with a Bonferroni adjusted significance level of p<.007 (for seven tests). Bonferroni corrected follow-up paired comparisons were used to explore any differences between interview conditions. Values of η^2 represent large (>.14), medium (.06-.14) or small (.01-.06) effect sizes.

Two analyses were of relevance to predictions as follows. For *Complies (with false information)* responses, a significant interview condition effect was present, $H(3) = 34.04$, p<.001, η^2 =.18. Follow-up comparisons indicated that, as predicted, proportions of Complies (false) responses were lower in the RI condition than in all other conditions: Best-Practice (z = 5.39, p<.001); Verbal Labels (z = 4.94, p<.001); and Sketch-RC (z = 4.22, p<.001). For *Resists (false information)* responses, no significant interview condition effect was present, contrary to predictions, $H(3) = 3.09$, p=.38, η^2=00.

We did not have specific predictions for the other five question types, but we present these analyses here, for completeness. For Complies (with true information) responses, a significant interview condition effect was present, $H(3) = 18.33$, p<.001, η^2=.09: proportions of Complies (true) responses were lower in the RI condition than in other conditions: Best-Practice (z=4.02, p<.001); Verbal Labels (z=3.54, p=.002); and Sketch-RC (z=3.05, p=.014). For Open responses, a significant interview condition effect was present, $H(3) = 21.96$, p<.001, η^2=.11: proportions of Open responses were higher in the RI condition than in other

conditions: Best-Practice (z=-3.48, p=.003); Verbal Labels (z=-4.48, p<.001); and Sketch-RC (z=-3.51, p=.003). No other interview condition effects reached significance for child responses: Resists (true information), $H(3) = 10.65$, p=.014, η^2=.04; No Response, $H(3) = 5.86$, p=.12, η^2=.02; and Seeks Clarification, $H(3) = 4.44$, p=.22, η^2=.01.

Barrister questions

A second subsidiary research question concerned whether, in the RI condition, the barristers' questions might be more consistent with best practice guidance for cross examination or re-examination. Table 8 shows mean numbers of questions per cross-examination, as well as proportions of each of the seven primary overarching types of questions for each interview condition. Overall, barristers asked an average of 61.39 (SD =18.78) questions per child. A one-way analysis of variance (data were normally distributed) showed a significant effect of interview condition, $F(3, 172) = 3.89$, p=.01, partial η^2=.06. Bonferroni corrected paired comparisons indicated that barristers asked significantly more questions in the RI condition (mean =71.09, SD =17.87) than in the Best-Practice (mean =58.92, SD =16.75) (p=.01) and Sketch-RC conditions (mean =58.26, SD =16.43) (p=.02). [This is consistent with real cross examinations: to simplify questions, asking two questions rather than one is often necessary.] The RI and Verbal Labels (mean =60.35, SD =22.42) conditions did not differ significantly (p=.08). Given these differences, further analyses on barrister questions were performed using proportional scores: the total number of questions in each question-type category were divided by the total number of barrister questions asked per child. These proportional data were not all normally distributed, so Kruskal-Wallis tests were used to explore whether there were interview condition differences on each question type, with a Bonferroni adjusted significance level of p<.007 (for seven tests). Bonferroni corrected follow-up paired comparisons were used to explore any differences between interview conditions.

Invitation Open questions differed significantly across interview condition, $H(3) = 45.24$, $p<.001$, $\eta^2=.25$. Proportions of Invitation Open questions were higher in the RI condition than in other conditions: Best-Practice ($z=-5.63$, $p<.001$); Verbal Labels ($z=-6.18$, $p<.001$); and Sketch-RC ($z=-5.03$, $p<.001$).

Invitation Closed (true information) questions differed significantly across interview condition, $H(3) = 39.91$, $p<.001$, $\eta^2=.22$. Proportions of Invitation Closed (true) questions were higher in the RI condition than in other conditions: Best-Practice ($z=-3.80$, $p=.002$); Verbal Labels ($z=-5.91$, $p<.001$); and Sketch-RC ($z=-5.00$, $p<.001$). A difference between Verbal Labels and Best-Practice also emerged ($z=2.87$, $p=.02$).

Assertion (true information) questions differed significantly across interview condition, $H(3) = 48.78$, $p<.001$, $\eta^2=.27$. Proportions of Assertion (true) questions were lower in the RI condition than in any other condition: Best-Practice ($z=5.41$, $p<.001$); Verbal Labels ($z=6.49$, $p<.001$); and Sketch-RC ($z=5.51$, $p<.001$).

Assertion (false information) questions differed significantly across interview condition, $H(3) = 16.71$, $p<.001$, $\eta^2=.08$. Proportions of Assertion (false) questions were lower in the RI condition than in the Verbal Labels condition ($z=3.64$, $p=.001$) and the Sketch-RC condition ($z=2.81$, $p=.03$); and that they were higher in the Verbal Labels condition than in the Best-Practice condition ($z=-2.80$, $p=.03$).

Option-posing questions differed significantly across interview condition, $H(3) = 11.49$, $p=.009$, $\eta^2=.05$. Proportions of option-posing questions were lower in RI than in Best-Practice interviews ($z=2.65$, $p=.049$). No other paired comparisons were significant.

Invitation Closed (false information) questions ($p=.10$) and *Other* questions ($p=.03$) showed no significant interview condition differences.

[insert Table 8 about here]

Table 9 includes breakdowns of barrister questions into 17 secondary features. These are presented as proportions (i.e. divided by the total number of barrister questions), but will not add up to one given the categories are not mutually exclusive (any question could be classified in one or more ways). [Note: no instances of the barrister saying the child was 'lying' were found; similarly, mean proportions for use of idiom were less than 1%; so these data were excluded.] These proportional data were not all normally distributed, so Kruskal-Wallis tests were used to explore interview condition differences for each question feature, with a Bonferroni adjusted significance level of $p<.003$ (for 15 tests). Bonferroni corrected follow-up paired comparisons were used to explore any differences between interview conditions.

Eight secondary question features showed significant interview condition differences.

Tags, $H(3) = 53.71$, $p<.001$, $\eta^2=.29$. Proportions of Tags were lower in the RI condition than in any other condition: Best-Practice ($z=5.58$, $p<.001$); Verbal Labels ($z=6.54$, $p<.001$); and Sketch-RC ($z=6.23$, $p=.008$).

Credibility, $H(3) = 30.74$, $p<.001$, $\eta^2=.16$. Proportions of Credibility challenges were lower in the RI condition than in other conditions: Best-Practice ($z=5.46$, $p<.001$); Verbal Labels ($z=3.92$, $p=.001$); and Sketch-RC ($z=3.03$, $p=.01$).

Repetition, $H(3) = 22.54$, $p<.001$, $\eta^2=.11$. Proportions of Repeated questions were higher in the RI condition than in the Best-Practice ($z=-4.65$, $p<.001$) and Verbal Labels ($z=-3.17$, $p=.009$) conditions.

Social Influence of another person, $H(3) = 28.64$, $p<.001$, $\eta^2=.15$. Proportional use of Social Influence was higher in the RI condition than in other conditions: Best-Practice ($z=-5.28$, $p<.001$); Verbal Labels ($z=-3.95$, $p<.001$); and Sketch-RC ($z=-3.67$, $p=.001$).

Possibility, $H(3) = 22.30$, $p<.001$, $\eta^2=.11$. Proportional use of Possibility was lower in the RI condition than in other conditions: Best-Practice ($z=4.71$, $p<.001$); Verbal Labels ($z=3.03$, $p=.014$); and Sketch-RC ($z=2.97$, $p=.018$).

Praise, $H(3) = 26.92$, $p<.001$, $\eta^2=.14$. Proportions of Praise were lower in the RI condition than in other conditions: Best-Practice ($z=4.86$, $p<.001$); Verbal Labels ($z=4.37$, $p<.001$); and Sketch-RC ($z=3.43$, $p=.004$).

Filler questions, $H(3) = 22.90$, $p<.001$, $\eta^2=.12$. Proportions of Filler questions were higher in the RI condition than in the Best-Practice ($z=-4.61$, $p<.001$) and Verbal Labels ($z=-3.67$, $p=.001$) conditions.

Reassurance, $H(3) = 24.63$, $p<.001$, $\eta^2=.13$. Proportions of Reassurance were lower in the RI condition than in other conditions: Best-practice ($z=4.95$, $p<.001$); Verbal Labels ($z=3.15$, $p=.01$); and Sketch-RC ($z=2.73$, $p=.038$).

[insert Table 9 about here]

Discussion

In this paper, a novel experimental methodology for the cross-examination of vulnerable child witnesses has been presented. Experienced barristers questioned children based on a 'defence statement' containing seven false elements, without recourse to a 'script' (as is typically used in experimental research on cross-examination). As predicted, children complied with barristers' challenges on this false information to a high degree: 94% of children complied with at least one cross-examination challenge on false information, consistent with previous experimental studies using scripted questioning in which compliance rates ranged between 70% and 98% (cf. Bettenay et al., 2014; Righarts et al., 2015; Zajac & Hayne, 2003, 2006; Zajac et al., 2009). Our findings underline concerns about whether cross-examination is a reliable method for obtaining best evidence from child witnesses,

given that lawyers try to 'persuade children to change details in their accounts, often by exploiting their developmental limitations' (Andrews & Lamb, 2016, p. 953).

We also tested, in an exploratory way, whether RI assistance, available in England and Wales, might help children to give better evidence by reducing compliance with barristers' cross-examination challenges on false information. As per recommendations for best practice in England and Wales at the time of the study (Ministry of Justice, 2015, see also current Registered Intermediary Procedural Guidance Manual, Ministry of Justice, 2020b), children received RI assistance at their initial interview and again at cross-examination. As tentatively predicted, RI assistance at cross-examination reduced children's compliance with false information, even after controlling for background cognitive factors, other key factors that could have influenced the findings, and memory for relevant details of the original event. Specifically, when children were challenged to agree with evidence that was 'false' (i.e., the barrister was suggesting that the child should agree with something in the defence statement that was 'false' and the child needed to resist this line of questioning), RI assistance made it less likely that children would comply with the barrister's challenges. This finding highlights the importance of using RIs for typically developing children to ensure that they do not give compliant responses to false information or change their responses when pressurised. For a child to accept that it was "possible", for example, that a woman had helped set up the video camera (when no such woman was present), would be enough to be used by the defence lawyer in undermining the evidence given the burden and standard of proof in criminal trials[6]. Overall, these exploratory findings about RIs support current recommendations in the Equal Treatment Bench Book (Judicial College, 2018; 2020) that: "All young witnesses should ideally have an intermediary assessment as, no matter how

[6] Although our study specifically looked at compliant responses to 'false information', which are undesirable, in some cases such responses would be appropriate if the information were true.

advanced they appear, their language comprehension is likely to be less than that of an adult witness" (paragraph 98, page 60). For typical children, RIs also help improve volume of recall in interviews and accuracy of identification in video lineups (Henry et al., 2017b; Plotnikoff & Woolfson, 2012; Wilcock et al., 2018). Overall, therefore, providing RIs for primary age typical children may improve the quality of their evidence.

A subsidiary research question concerned whether, when we broke down children's specific responses to barrister questioning, these responses would be less compliant with *and* more resistant to challenges on false information in the RI condition. As tentatively predicted, significantly lower proportions of 'complies with false information' responses were given by children in the RI condition than in other conditions (5% in the RI condition versus 11%-13% in other conditions): children were less likely to agree with a barrister's false statement in the RI condition. Although the proportions of 'resists false information' responses did not vary with interview condition, as expected, this could be because resisting a false statement is more difficult for a child (i.e., actively saying 'that is not true') than not agreeing with a false statement (possible with more passive responses such as 'don't know' or providing no response at all). Overall, these findings accorded closely with the primary research finding that RI assistance helped children to reduce compliance in response to barrister challenges on false information.

A final subsidiary research question concerned whether barristers would ask questions more aligned with best practice recommendations in the RI condition. In support of this, barristers asked proportionally more Invitation Open questions in the RI condition. Whilst these have been associated with inconsistencies (due to the longer answers they elicit) (Pichler et al., 2020), they are consistent with best practice (Ministry of Justice, 2011), are least likely to lead the witness (Henderson et al., 2019), and are highly valued by practitioners (Magnusson, Ernberg, Landström, & Akehurst, 2020). Invitation Open questions were,

nevertheless, relatively rare, as reported in real cases (e.g., Andrews & Lamb, 2016; Pichler et al., 2020; Zajac et al., 2018). Rates here ranged from 4-5% in non-RI conditions, to 12% in RI cross-examinations. Also consistent with best practice, barristers asked proportionally fewer Assertion questions in the RI condition. Such questions are risky because they present a strong statement that might be difficult to resist and could, thus, lead the witness (Henderson et al., 2019; Judicial College, 2013; The Council of the Inns of Court, 2019). Proportions of Assertions about true information were significantly lower (21%) in RI interviews than in other interviews (range 32%-36%), although proportions of Assertions about false information did not reveal such consistent group differences (RI =6%, other conditions =11%-16%).

Other findings concerning the barrister questions were harder to interpret. Invitation Closed (true information) questions were significantly higher in RI interviews (33%) than in other interviews (range 18-24%), although no group differences emerged for Invitation Closed (false information) questions. In real cases it may not be apparent whether these yes/no style questions are misleading, if the truth is not known. Yes/no questions for 'true' information may be less risky in terms of leading the witness, whereas yes/no questions for false information could be actively misleading. Finally, the small group difference in Option-Posing questions indicated somewhat fewer of these in the RI condition than the Best-Practice condition, but rates of these questions were low (3% or less in all conditions), so this result should be viewed with caution.

Further detailed classification of the features of barrister questions into secondary categories offered some evidence that they were more aligned with best practice recommendations in the RI condition. First, there were reductions in the use of suggestive tag questions (4% versus 19%-28%), supporting existing best practice guidance (Ministry of Justice, 2011; Judicial College, 2013, 2018; The Advocate's Gateway, 2015; The Council of

the Inns of Court, 2019). Second, there were reductions in challenges to the children's credibility (2% versus 5%-7%) and fewer suggestions that something 'possibly' happened (<1% versus 3%-4%). Although these questions were infrequent overall, the lower rates in RI interviews may have increased the child's confidence in themselves as a respondent, particularly as children dislike having their credibility challenged (Plotnikoff & Woolfson, 2012).

More difficult to interpret was the fact that RI interviews showed increases in repetitions compared to most other interviews (12% versus 5%-9%). Question repetition is not recommended as it could confuse or exploit the child into changing answers (Andrews et al., 2015b; Ministry of Justice, 2011; Judicial College, 2013, 2018; The Council of the Inns of Court, 2019). In fact, the RIs removed any repeated questions when checking barrister questions before cross-examination, so it is possible that barristers re-introduced them to help children to follow the line of questioning if they lost track, or because they were unable to diverge from the listed questions if they wanted to press a point. Other differences in RI interviews that were unexpected included the use of 'social influence of another person' being more common (9% versus 3%-4%). This could reflect barristers switching from challenging the children's credibility outright or inferring the 'possibility' of being incorrect, to rely on a gentler approach by suggesting they were affected by social influence of another person instead. It could also reflect a technique to check the child's ability to challenge the barrister (or the defendant) who expresses a different view. There was also less praise and reassurance (4% versus 8%-10%, and <1% versus 2%-4%, respectively) in RI cross-examinations, perhaps because barristers opted to give more praise and reassurance in non-RI interviews to conceal the fact that they were undermining the child's evidence. Finally, there were more irrelevant (filler) questions (although note that the RI vs S-RC comparison here was not significant and the values were low in all cases: RI 2% and other conditions 1% or

less). Overall, despite some areas of uncertainty, these findings suggest that recommendations by RIs regarding the wording of cross-examination challenges could align questioning more closely with best practice recommendations.

The study findings may contribute to internationally available sources of guidance about how lawyers should question children in court, given concerns in this area (e.g., Andrews et al., 2015a). Further training about how to question vulnerable witnesses (e.g., advocates in England and Wales now attend training to acknowledge the '20 Principles of Questioning', The Council of the Inns of Court, 2019), along with pre-trial ground rules hearings as standard (see Henderson et al., 2019), would be useful for all barristers involved in child cases. The Advocate's Gateway provides detailed recommendations for barristers and other legal professionals on questioning a range of vulnerable witnesses, including children (www.theadvocatesgateway.org). Pre-trial guidance aimed at children may also help because practice sessions in responding to cross-examination style questions on an unrelated topic can significantly improve children's overall accuracy during a cross-examination interview (Irvine, Jack, & Zajac, 2016; Righarts, O'Neill, & Zajac, 2013), provided it is given close to the interview date (O'Neill & Zajac, 2013). Future research could investigate a combination of RI assistance and timely pre-trial preparation (perhaps delivered as part of the RI assessment), as combining these interventions may further improve the quality of children's cross-examination evidence.

One area the study was unable to illuminate was whether the RI assistance impacted on the child's responses, the barrister's questioning technique, or both. We are also uncertain about the mechanisms and exact points through which RI assistance operated, but it is important to note that the overarching role of the RI is to support the child's communication needs (e.g., simplifying instructions, using visual aids) and impact the barrister's questioning to ensure it is appropriate. All of this should help the witness more easily understand what

others are saying so that they can communicate better. Further research could unpick the important mechanisms underpinning the interplay between children's responses and barristers' questions. The very nature of cross-examination requires some fluidity in questioning and a good advocate will always be influenced by the child's responses. The exception to this would be to use a rigid script of questions (which is necessary in some extreme cases, but not generally). Otherwise, the barrister will be flexible and adapt in response to the child's answers. This was one of the advantages to our novel approach to assessing cross-examination empirically, which has, to our knowledge, not been addressed in previous empirical work.

There are some limitations to the study that should be acknowledged. One is that the findings are applicable only to defence barristers, as different lines of questioning may be applied by prosecution barristers (Denne, Sullivan, Ernest, & Stolzenberg, 2020). Another is that children in the RI condition, as per best practice guidance (Ministry of Justice, 2015), had already received RI assistance during previous phases of the mock criminal investigation: this was given at the investigative interview stage (which also included an identification lineup). Therefore, the current conclusions can only be applied to children who have had RI assistance throughout a criminal investigation which, in practice, is not always the case (RIs may sometimes only brought in at trial stage, although this is not recommended). A related issue was that children in the RI condition remembered more about the initial witnessed event, as RI assistance was effective in increasing the volume of accurate recall at investigative interview (Henry et al., 2017b). This meant that children in the RI condition started their cross-examination with a recall advantage. We mitigated this by controlling for how well the child had recalled key facts about the false information in the defence statement (memory trace scores). Although memory trace was not a significant predictor of cross-examination resistance in the full regression (and many children did not score highly on this

measure), future research could match on initial memory of the staged event before instigating cross-examinations in groups with and without RI assistance. This method would mean that no children could be included who had previously undergone an investigative interview assisted by an RI, but such a method would provide evidence about the effectiveness of RI assistance brought in only at the trial stage.

Further limitations are as follows. We used a mild minor crime event that took place in a familiar environment (the children's school), so were unable to replicate the anxiety, unfamiliarity and potential trauma of a real court case, which limits generalisation of the findings to real cases. Children were seen by friendly and supportive researchers, and the barristers were also approachable and experienced – they were, partly, chosen on the basis of having previous experience in cross-examining children (for ethical reasons) – again, this might not be so in real-life. Our ground rules hearings for non-RI children were also brief, and more recent guidance now recommends they are included as 'good practice' for all young witnesses (Judicial College, 2018, revisions 2020, Equal Treatment Bench Book, p.64). Finally, the length of the cross-examinations, for ethical reasons was short (average 8.56 minutes) compared to real cases (reported in England and Wales as between 45 minutes and 3 hours, Baverstock, 2016). However, Henderson et al. (2019) reported much shorter video-recorded cross-examinations (16 minutes) in a pilot trial of this special measure in England, and with new advocate training and guidance, cross-examinations are likely to be more limited in length (e.g., Judicial College, 2018). Similarly, although studies of court transcripts in Scotland, California and New Zealand have emphasised the large numbers of questions (ranging from 160-500) posed to children by prosecutors and defence lawyers (Andrews & Lamb, 2016; Andrews et al., 2015a; Klemfluss et al., 2014; Zajac & Cannan, 2009), the number of questions posed during pilot video-recorded cross- and direct-examinations in Henderson et al.'s (2019) study was lower (average=92). Thus, although the

current cross-examinations contained fewer questions (average=61), the overall numbers of questions may be more aligned with the newer pre-recorded cross-examinations in England. Given that long and complex cross-examinations will likely lead to fatigue, worsening the quality of evidence (e.g., Zajac et al., 2018), changes that encourage shorter questioning should be advantageous.

Conclusion. The current study was the first to use a more ecologically valid defence statement as the basis for unscripted empirical cross-examinations. Using this novel method, we found that children complied with a very high number of barrister challenges on false information. However, we also found exploratory evidence that RI assistance reduced children's compliance with barristers' cross-examination challenges on false information. This could be, in part, because the barristers asked questions that were somewhat more aligned with best practice recommendations in the RI condition. These findings extend previous research on the utility of RIs during investigations (evidence-gathering interviews and identification lineups). They provide additional evidence of the importance of using RIs to ensure typically developing young children can give accurate testimony during the final investigative phase (cross-examination).

References

Andrews, S.J. & Lamb, M.E. (2016). How do lawyers examine and cross-examine children in Scotland? *Applied Cognitive Psychology, 30*, 953-971. doi: 10.1002/acp.3286

Andrews, S.J., Lamb, M.E. & Lyon, T.D. (2015a). Question types, responsiveness and self-contradictions when prosecutors and defense attorneys question alleged victims of child sexual abuse. *Applied Cognitive Psychology, 29*, 253-261. doi: 10.1002/acp.3103

Andrews, S.J., Lamb, M.E. & Lyon, T.D. (2015b). The effects of question repetition on responses when prosecutors and defense attorneys question children alleging sexual abuse in court. *Law and Human Behavior, 39*, 559-570. doi: 10.1037/lhb0000152

Baverstock, J. (2016). *Process evaluation of pre-recorded cross-examination pilot (Section 28).* Ministry of Justice, Ministry of Justice Analytical Series.

Bettenay, C., Ridley, A.M., Henry, L.A., & Crane, L. (2014). Cross-examination: The Testimony of Children With and Without Intellectual Disabilities. *Applied Cognitive Psychology, 28*, 204-214. doi: 10.1002/acp.2979

Bowers, L., Huisingh, R., LoGiudice, C., & Orman, J. (2010). *The expressive language test 2*. East Moline, IL: LinguiSystems.

Brown, D. & Pipe, M-E. (2003). Individual Differences in Children's Event Memory Reports and the Narrative Elaboration Technique. *Journal of Applied Psychology, 88*, 195–206. doi: 10.1037/0021-9010.88.2.195

Bull, R. (2010). The investigative interviewing of children and other vulnerable witnesses: Psychological research and working/professional practice. *Legal and Criminological Psychology, 15*, 5–23. doi:10.1348/014466509X440160

Collins, K. & Krahenbuhl, S. (2020). Registered intermediaries' assessment of children's communication: An exploration of aims and processes. *Evidence & Proof, 24*(4), 374-395. doi: 10.1177/1365712720952335

Cooper, P., Backen, P. & Marchant, R. (2015). Getting to grips with ground rules hearings: A checklist for judges, advocates and intermediaries to promote the fair treatment of vulnerable people in court. *Criminal Law Review, 2015*, 6, 420-435.

Cooper, P., & Mattison, M. (2017). Intermediaries, vulnerable people and the quality of evidence: An international comparison of three versions of the English intermediary model. *The International Journal of Evidence & Proof*, *21*, 351–370. doi: 10.1177/1365712717725534

Cooper & Wurzel (2014). Better the second time around? Department of Justice Registered Intermediaries Schemes and lessons from England and Wales. *Northern Ireland Legal Quarterly*, *65*, 39-61.

Dando, C., Wilcock, R., & Milne, R. (2009). The cognitive interview: The efficacy of a modified mental reinstatement of context procedure for frontline police investigators. *Applied Cognitive Psychology, 23*, 138–147. doi:10.1002/acp.1451

Denne, E., Sullivan, C., Ernest, K., & Stolzenberg, S.N. (2020). Assessing children's credibility in courtroom investigations of alleged child sexual abuse: Suggestibility, plausibility, and consistency. *Child Maltreatment, 25*(2), 224-232. doi: 10.1177/1077559519872825

Dunn, L.M., Dunn, D.M., & Styles, B. (2009). *British Picture Vocabulary Scale (3rd ed.).* London, UK: GL Assessment.

Evans, A.D., Lee, K. & Lyon, T.D. (2009). Complex questions asked by defense lawyers but not prosecutors predicts convictions in child abuse trials. *Law and Human Behavior, 33*, 258-264. doi: 10.1007/s10979-008-9148-6

Field, A.P. (2013). *Discovering statistics using IBM SPSS Statistics: and sex and drugs and rock 'n' roll* (fourth edition). London: Sage publications.

Green, S.B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research, 26*, 499–510. https://doi.org/10.1207/s15327906mbr2603_7

Hanna, K., Davies, E., Crothers, C. & Henderson, E. (2012). Questioning child witnesses in New Zealand's criminal justice system: Is cross-examination fair? *Psychiatry, Psychology and Law, 19*, 530-546. doi:10.1080/13218719.2011.615813

Hanna, K. & Henderson, E. (2018). '[Expletive], that was confusing wasn't it?' Defence lawyers' and intermediaries' assessment of the language used to question a child witness. *The International Journal of Evidence and Proof, 22*, 411-427. doi: 10.1177/1365712718796527

Henderson, E. (2002). Persuading and controlling: The theory of cross-examination in relation to children. In H.L. Westcott, G.M. Davies & R.H.C. Bull (Eds), *Children's Testimony: A handbook of psychological research and forensic practice* (pp. 279-294). Wiley: Chichester.

Henderson, H.M., Andrews, S.J. & Lamb, M.E. (2019). Examining children in English High Courts with and without implementation of reforms authorized in Section 28 of the Youth Justice and Criminal Evidence Act. *Applied Cognitive Psychology, 33,* 252-264. doi: 10.1002/acp.3472

Henderson, H.M. & Lamb, M.E. (2019). Does implementation of reforms authorized in Section 28 of the Youth Justice and Criminal Evidence Act affect the complexity of the questions asked of young alleged victims in court? *Applied Cognitive Psychology, 33,* 201-213. doi: 10.1002/acp.3466

Henry, L.A., Crane, L., Nash, G., Hobson, Z., Kirke-Smith, M., & Wilcock, R. (2017a). Verbal, visual, and intermediary support for child witnesses with autism during

investigative interviews. *Journal of Autism and Developmental Disorders, 47,* 2348-2362. doi: 10.1007/s10803-017-3142-0.

Henry, L.A., Messer, D.J., Wilcock, R., & Crane, L. (2017b). Do measures of memory, language, and attention predict eyewitness memory in children with and without autism spectrum disorder? *Autism and Developmental Language Impairments, 2,* 1-17. doi: 10.1177/2396941517722139

Home Office (1989). *Report of the advisory group on video-recorded evidence*. Chairman His Honour Judge Thomas Pigot, QC.

Irvine, B., Jack, F. & Zajac, R. (2016). Preparing children for cross-examination: do the practice questions matter? *Psychology, Crime & Law, 22*, 858-878. doi: 10.1080/1068316X.2016.1197224

Judicial College (2013). Equal Treatment Bench Book. https://www.sentencingcouncil.org.uk/wp-content/uploads/equal-treatment-bench-book-2013-with-2015-amendment.pdf. Accessed 20th May 2015.

Judicial College (2018). Equal Treatment Bench Book. https://www.judiciary.uk/wp-content/uploads/2018/02/equal-treatment-bench-book-february-v6-2018.pdf Accessed 13th February 2020. Also, Equal Treatment Bench Book (March 2020 revision). https://www.judiciary.uk/wp-content/uploads/2018/02/ETBB-February-2018-amended-March-2020.pdf Accessed 13th April 2021.

Judicial College Bench Checklist: Young Witness Cases (2012). Available at: https://www.judiciary.uk/publications/jc-bench-checklist-young-wit-cases/ Accessed 20 May 2015.

Klemfuss, J.Z., Quas, J.A. & Lyon, T. (2014). Attorneys' questions and children's productivity in child sexual abuse criminal trials. *Applied Cognitive Psychology, 28*, 780-788. doi: 10.1002/acp.3048

Krahenbuhl, S. (2019). Mock jurors' perceptions of a child witness: The impact of the presence and/or absence of a registered intermediary during cross-examination. *Psychology, Crime & Law, 25*:7, 713-728. doi:10.1080/1068316X.2018.1543421

Lamb, M. E., Malloy, L. C., & La Rooy, D. J. (2011). Setting realistic expectations: Developmental characteristics, capacities, and limitations. In M. E. Lamb, D. J. La Rooy, L. C. Malloy & C. Katz (Eds.), *Children's testimony: A handbook of psychological research and forensic practice* (pp. 15–48). West Sussex, UK: Wiley-Blackwell.

Magnusson, M., Ernberg, E., Landström, S. & Akehurst, L. (2020). Forensic interviewers' experiences of interviewing children of different ages. *Psychology, Crime & Law, 26*:10, 967-989. doi:10.1080/1068316X.2020.1742343

Manly, T., Robertson, I.H., Anderson, V. & Nimmo-Smith, I. (1999). *The test of everyday attention for children*. London: Pearson.

Ministry of Justice (2011). *Achieving best evidence in criminal proceedings: Guidance on interviewing victims and witnesses, and using special measures*. London: Her Majesty's Stationary Office.

Ministry of Justice (2015). The Registered Intermediary Procedural Guidance Manual. Available at: https://www.theadvocatesgateway.org/images/procedures/registered-intermediary-procedural-guidance-manual.pdf Accessed 20th May 2015.

Ministry of Justice (2020a). *The witness intermediary scheme: Annual report 2019-20*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/919858/witness-intermediary-scheme-annual-report-2019-2020.pdf Accessed 13th October 2020.

Ministry of Justice (2020b). Registered Intermediary Procedural Guidance Manual. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/955316/registered-intermediary-procedural-guidance-manual.pdf Accessed 13th April 2021.

O'Neill, S. & Zajac, R. (2013). Preparing Children for Cross-Examination: How Does Intervention Timing Influence Efficacy? *Psychology, Public Policy, and Law, 19*, 307-320. doi: 10.1037/a0031538

Pichler, A.S., Powell, M., Sharman, S.J., Zydervelt, S., Westera, N., & Goodman-Delahunty, J. (2020). Inconsistencies in complainant's accounts of child sexual abuse arising in their cross-examination. *Psychology, Crime & Law.* doi: 10.1080/1068316X.2020.1805743

Plotnikoff, J, & Woolfson, R (2009): *Measuring Up? Evaluating implementation of Government commitments to young witnesses in criminal proceedings*. London: NSPCC and Nuffield Foundation.

Plotnikoff, J. & Woolfson, R. (2012). 'Kicking and screaming': The slow road to best evidence. In J. R. Spencer, & M. E. Lamb (Eds.), *Children and cross-examination: Time to change the rules?* (pp. 21–41). Oxford: Hart Publishing.

Plotnikoff, J. & Woolfson, R. (2015). *Intermediaries in the Criminal Justice System: Improving Communication for Vulnerable Witnesses and Defendants.* Bristol, UK: Policy Press.

Reynolds, C., & Voress, J.K. (2007). *Test of Memory and Learning: Second Edition.* Austin, TX: Pro-Ed.

Righarts, S., O'Neill, S., & Zajac, R. (2013). Addressing the Negative Effect of Cross-Examination Questioning on Children's Accuracy: Can We Intervene? *Law and Human Behavior, 37*, 354–365. doi: 10.1037/lhb0000042

Righarts, S., Jack, F., Zajac, R. & Hayne, H. (2015). Young children's responses to cross-examination style questioning: The effects of delay and subsequent questioning, *Psychology, Crime & Law, 21*, 274-296, doi: 10.1080/1068316X.2014.951650

Semel, E., Wiig, E.H. & Secord, W.A. (2006*). Clinical Evaluation of Language Fundamentals Fourth UK Edition*. London: Harcourt Assessment.

Spencer, J.R. (2012). Introduction. In J.R. Spencer & M.E. Lamb. *Children and Cross-Examination: Time to Change the Rules?* (pp. 1-20). Oxford, UK: Hart Publishing.

The Advocate's Gateway (2015*). Toolkit 6: Planning to question a child or young person.* Available at: https://www.theadvocatesgateway.org/images/toolkits/6-planning-to-question-a-child-or-young-person-141215.pdf (Accessed 20 May 2015).

The Council of the Inns of Court (2019). *The 20 Principles of Questioning: A guide to the cross-examination of vulnerable witnesses.* Available at: https://www.icca.ac.uk/wp-content/uploads/2019/05/20-Principles-of-Questioning.pdf (Accessed 13 February 2020).

Wechsler, D. & Zhou, X. (2011). *Wechsler Abbreviated Scale of Intelligence – Second Edition*. Bloomington, MN: Pearson.

Wilcock, R., Crane, L., Hobson, Z., Nash, G., Kirke-Smith, M., & Henry, L.A. (2018). Supporting child witnesses during identification lineups: exploring the effectiveness of Registered Intermediaries. *Applied Cognitive Psychology, 32*, 367-375. doi: 10.1002/acp.3412

Wilcock, R., Crane, L., Hobson, Z., Nash, G., Kirke-Smith, M., & Henry, L.A. (2019). Brief Report: Eyewitness identification in child witnesses on the autism spectrum. *Research in Autism Spectrum Disorders, 66*, 101407. doi: 10.1016/j.rasd.2019.05.007

Youth Justice and Criminal Evidence Act (1999). Available at http://www.legislation.gov.uk/ukpga/1999/23/contents (Accessed: 20 May 2015).

Taggart, J. (2021). 'I am not beholden to anyone…I consider myself to be an officer of the court': A comparison of the intermediary role in England and Wales and Northern Ireland. *The International Journal of Evidence & Proof*, 1–22. doi: 10.1177/13657127211002291

Zajac, R., & Cannan, P. (2009). Cross-examination of sexual assault complainants: A developmental comparison. *Psychiatry, Psychology & Law*, *16*, S36-S54. doi: 10.1080/13218710802620448

Zajac, R., Gross, J., & Hayne, H. (2003). Asked and answered: Questioning children in the courtroom. *Psychiatry, Psychology and Law, 10*, 199–210. doi: 10.1375/132187103322300059

Zajac, R., & Hayne, H. (2003). I don't think that's what really happened: The effect of cross-examination on the accuracy of children's reports. *Journal of Experimental Psychology: Applied, 9*, 187–195. doi: 10.1037/1076-898X.9.3.187

Zajac, R., & Hayne, H. (2006). The negative effect of cross-examination style on children's accuracy: Older children are not immune. *Applied Cognitive Psychology, 20*, 3-16. doi: 10.1002/acp.1169

Zajac, R., Jury, E., & O'Neill, S. (2009). The role of psychosocial factors in young children's responses to cross-examination style questioning. *Applied Cognitive Psychology, 23*, 918–935. doi: 10.1002/acp.1536

Zajac, R., O'Neill, S., & Hayne, H. (2012). Disorder in the courtroom? Child witnesses under cross-examination. *Developmental Review, 32*, 181–204. doi: 10.1016/j.dr.2012.06.006

Zajac, R., Westera, N., & Kaladelfos, A. (2018). The "good old days" of courtroom questioning: Changes in the format of child cross-examination questions over 60 years. *Child Maltreatment, 23*, 186-195. doi: 10.1177/1077559517733815

Table 1. Mean (SD) scores on cognitive variables for children in each interview condition, together with relevant differences (these variables were controlled in the regression analysis).

***Variables*:**	**Best-Practice (n = 65)**	**Verbal Labels (n = 40)**	**Sketch-RC (n = 38)**	**Registered Intermediary (n = 33)**	**Group differences[3] (in bold)**
Age at cross-exam. (months)	114.95 (12.84)	110.55 (12.24)	108.92 (13.96)	118.73 (14.44)	$F(3, 172) = 4.13$, $p = .007$** **RI > S-RC**
WASI-II[1] (IQ)	109.89 (12.97)	106.50 (12.42)	109.39 (13.99)	100.94 (14.20)	$F(3, 172) = 3.70$, $p = .01$* **RI < BP**
TOMAL-2 Composite[1] (Memory)	113.95 (15.63)	111.75 (14.18)	112.53 (12.47)	108.73 (16.25)	$F(3, 172) = .93$, $p = .43$
TOMAL-2 Verbal[1] (Verbal Memory)	114.43 (16.17)	113.95 (15.29)	110.47 (14.29)	104.94 (16.74)	$F(3, 172) = 3.05$, $p = .03$* **RI < BP**
TOMAL-2 Non-verbal[1] (Non-verbal memory)	110.20 (18.18)	106.73 (15.42)	111.68 (13.68)	110.76 (20.11)	$F(3, 172) = .63$, $p = .60$
BPVS-3[1] (Receptive vocabulary)	95.65 (13.02)	94.73 (12.79)	94.87 (13.17)	87.52 (15.30)	$F(3, 172) = 2.95$, $p = .03$* **RI < BP**
ELT-2 Sequencing[1] (Narrative ability)	109.83 (9.06)	107.70 (9.27)	112.11 (8.43)	109.12 (6.91)	$F(3, 172) = 1.76$, $p = .16$
ELT-2 Grammar & Syntax[1] (Grammatical morphology)	106.92 (10.40)	106.97 (10.42)	108.82 (9.45)	103.79 (11.49)	$F(3, 171) = 1.40$, $p = .25$
CELF-4-UK Recalling Sentences[2] (Grammatical understanding/ production)	10.58 (3.41)	11.70 (2.19)	11.26 (2.46)	10.85 (3.23)	$F(3, 172) = 1.31$, $p = .27$
CELF-4-UK Formulated Sentences[2] (Sentence formulation /production)	9.28 (3.28)	10.15 (2.81)	10.58 (2.75)	8.94 (3.34)	$F(3, 172) = 2.39$, $p = .07$
TEA-Ch Sky Search[2] (Selective attention)	9.35 (2.64)	9.25 (2.59)	8.89 (2.85)	8.97 (3.45)	$F(3, 172) = .27$, $p = .84$
TEA-Ch Score![2] (Sustained attention)	9.03 (3.36)	8.85 (3.30)	9.32 (3.80)	8.91 (3.53)	$F(3, 172) = .13$, $p = .94$
TEA-Ch Dual Task[2]	6.91	6.80 (3.46)	6.03 (3.81)	5.15 (3.67)	$F(3, 172) =$

(Sustained-divided attention)	(3.57)				2.03, p = .11
Memory trace score (max=18)	4.61 (3.02)	5.78 (2.99)	5.89 (3.25)	6.94 (2.61)	$F(3, 172)$ = 4.79, p = .003** **RI > BP**
Delay to cross-exam. (months)	10.15 (1.78)	11.20 (1.47)	11.34 (1.56)	12.36 (0.55)	$F(3, 172)$ = 16.78, p < .001*** **BP < all others** **RI > all others**

[1]Standardised scores (mean 100, SD 15); [2]scaled scores (mean 10 SD 3); [3] for paired comparisons after Bonferroni corrections.

Key:
WASI-II Wechsler Abbreviated Scale of Intelligence, second edition.
TOMAL-2 Test of Memory and Learning, second edition.
BPVS-2 British Picture Vocabulary Scale, second edition.
ELT-2 Expressive Language Test, second edition.
CELF-4-UK Clinical Evaluation of Language Fundamentals, fourth edition, UK version.
TEA-Ch Test of Everyday Attention for Children.

Table 2. Sample defence statement from one of the two versions of the event[a] (including the 'truth' and the seven 'false' statements)

	Points from the defence statement	The 'ground truth' – from the event
1	One morning last year, Max and I visited a school to give a talk about the Victorians to the children and their teachers	True.
2	Max was wearing a blue top and has short brown hair. I was wearing a grey top and had long blond hair tied back in a ponytail.	True.
3	When we arrived, a woman helped us by setting up the video camera at the back which recorded the talk.	False item 1 – Adam set up the video camera. There was no woman involved in the event.
4	We told the children some rules that Victorian children had to obey, for instance, we said that boys must learn needlework	False item 2 – whilst the children were told about rules, this specific example is incorrect – the children were told that girls (not boys) had to learn needlework.
5	We showed the children a slate and Max showed them how to write the letters of the alphabet on it with chalk.	False item 3 – the children were shown a slate, but Max wrote a sum on the slate (not the alphabet).
6	Max is very forgetful and during the talk he asked the children to remind him not to forget his phone at the end of the talk.	True.

7	Max then put his phone on the chair in the hall.	True.
8	Max says that I stole his phone by taking it and putting it in my pocket – I did not do this. Max's phone was on the chair the whole time. I did not go near the chair at any time during or after the talk.	False item 4 – Adam did take Max's phone and put it in his pocket.
9	I did borrow Max's keys during the talk and put them in my pocket.	False item 5 – there were no keys involved in the staged event.
10	At the end of the talk, Max forgot his coat.	False item 6 – Max forgot his jumper (which he spoke about at the start of the talk).
11	When Max forgot his coat, I had to go back to get it.	False item 7 – Max (not Adam) returned after he had left, to collect the forgotten item.

[a] Whilst the other version of the event was very similar, points 4-11 on the defence statement differed: for example, there were slightly different names (Mark and Alex) for the key actors; children saw the theft of a set of keys, but the barrister had to put to them that it was, in fact, a phone; and the children were told that boys had to learn technical drawing (with the barristers suggesting to them that this was girls).

Table 3: Types of child responses during cross-examinations with explanations.

Type of response	**Explanation**
Complies (true)	When a child complies with what the barrister has said, in relation to a true (correct) statement.
Complies (false)	When a child complies with what the barrister has said, in relation to a false (incorrect) statement
Resists (true)	When a child has resisted what the barrister has said, in relation to a true (correct) statement
Resists (false)	When a child has resisted what the barrister has said, in relation to a false (incorrect) statement
No Response	The child has not given a response
Open Response	When a child has given a response to a barrister's open question (they can't comply or resist, as the child is given the opportunity to tell their version of events)
Seeks Clarification	The child seeks Clarification (e.g., "I don't know what you mean")

Table 4: The seven overarching primary codes for barrister questions during cross-examinations, with explanations and examples

Type of Question	**Explanation**	**Example**
Invitation Open	A question that invites the witness to offer their account and does not declare the answer (or have a correct answer)	"Who set up the video camera?"
Invitation Closed (true)	A question that invites a yes or no response, or asks for confirmation – includes true (correct) information	"Did Alex set up the video camera?"
Invitation Closed (false)	A question that invites a yes or no response, or asks for confirmation – includes false (incorrect) information	"Did Mark set up the video camera?"
Assertion (true)	Questions in the form of a statement, which is true (correct); or a statement of the child's previous response	"Alex set up the video camera."; "Towards the back, that's really helpful."
Assertion (false)	Questions in the form of a statement, which is false (incorrect)	"Alex didn't set up the video camera?"
Option Posing	Questions in the form of two or more options (that may include the option to choose 'something else')	"Had he got his back to you, front, side, something else?" "Was it blond or brown

		hair?" "Is that a big room or a small room?"
Other	Utterances that were not content-based questions (e.g. signpost, credibility, praise, clarification and reassurance)	"Lovely, thank you so much B."; "Can I ask you some questions about that because that's really helpful?"; "He did, that's fantastic, well done A."

Table 5: Further secondary classifications of features of the barrister's questions during cross-examinations, with explanations and examples.

Classification	***Explanation***	***Example***
Tag	A question asking for confirmation, suggestive as it communicates the expected response	"Mark picked up the keys, didn't he?"
Credibility	A question that challenges the integrity or credibility of the witness, or their memory	"You think they did. You say you think, did you actually see them do it or are you guessing?"
Negatives	A question containing a negative	"Didn't Mark pick up the keys, not Alex?"
Repetition	Repeating the same question, even if interspersed by others	"Did Alex take the keys?" A: "No". "Did Alex take the keys?"
Confirmation	The advocate confirms the answer the child has given, in a best practice way - a permissible and gentle way of checking evidence	"I want to make sure I understand what you said…" "so they showed you the slate but they didn't do any writing, is that what you're saying?"
Clarification	The advocate checks that the answer the child has given is what was intended	"You nodded, so is that a yes, brilliant, thank you very much."

Social influence of another person	The barrister suggests that 'someone else' told them that what the child has said happened didn't really happen	"Alex told me he didn't take the keys"
Possibility	A question that suggests that what the barrister is putting to them might be true (even if the witness is unsure) – possibility is introduced	"And was there maybe a lady helping out?"
Complex	A question that is linguistically complex, because of the large number of instructions contained in it, because of ambiguity or because it has conjunctions making it long-winded	"But I hope that if I ask you some questions, and I know you've, you've gone through what you said in your, um, your interview about it, uh, if I ask you some questions, we might be able to work out together, um, exactly what happened when those two people came to school, okay?"
Idiom	Phrase with a figurative or literal meaning	"now let's go back to square one"
Do you remember…?	Questions asking the witness if they remember what they said on a previous occasion are particularly frowned upon	"do you remember any other adults in the room?" "can you remember that?"

Lying	Directly accuses the witness of lying	Note: an example is not given as there were no examples of accusing the child of lying in the current study.
Signpost	Explaining or signposting changes of subject (includes references to original evidence, e.g., "in your interview, you said that...")	"Now we're going to talk about the other man, the man with the long hair called Adam."
Praise	Thanking or commending the child in an encouraging way	"That's brilliant, thank you for that. I've only got one more thing to ask you…"
Filler	Irrelevant questions	"The men who came to your school, were they funny?"
Name	The advocate uses the child's name	"That's really helpful, you've got a very good memory here N."
Reassurance	The advocate provides reassurance that the child is doing okay	"That's okay, not to worry, so you can't help me with who set it up if you don't remember."

Table 6. Resistance scores for children in each interview condition (highest average resistance score is 2, lowest is 0), total numbers of child responses, and proportional (prop.) scores for different types of responses for each interview condition. Mean proportions (SDs) are given on line 1, medians (ranges) on line 2.

Scores	**Best-Practice (n = 65)**	**Verbal Labels (n = 40)**	**Sketch-RC (n = 38)**	**Registered Intermediary (n = 33)**
Cross-examination resistance score (average over 7 'false' defence statement elements)	**.85** (.49) .71 (.00-2.00)	**.80** (.41) .86 (.00-1.67)	**.94** (.56) .84 (.00-2.00)	**1.42** (.45) 1.43 (.43-2.00)
Total number of child responses across full cross-examination	**43.82** (11.40) 42 (26-71)	**45.35** (16.41) 42 (12-78)	**46.37** (16.2) 41 (16-78)	**52.82** (13.05) 56 (33-76)
Prop. Complies with true statement	**.38** (.10) .37 (.15-.60)	**.38** (.13) .34 (.00-.66)	**.36** (.10) .34 (.21-.53)	**.29** (.12) .24 (.11-.61)
Prop. Complies with false statement	**.13** (.07) .13 (.00-.37)	**.15** (.10) .11 (.03-.39)	**.13** (.11) .12 (.00-.56)	**.05** (.05) .05 (.00-.17)
Prop. Resists true statement	**.12** (.07) .12 (.00-.28)	**.09** (.08) .08 (.00-.33)	**.11** (.08) .09 (.00-.35)	**.15** (.09) .14 (.03-.37)
Prop. Resists false statement	**.18** (.10) .16 (.03-.50)	**.20** (.10) .21 (.00-.41)	**.20** (.10) .18 (.00-.39)	**.20** (.06) .20 (.06-.32)
Prop. 'No Response'	**.03** (.05) .00 (.00-.30)	**.03** (.04) .03 (.00-.16)	**.04** (.05) .02 (.00-.16)	**.01** (.02) .01 (.00-.05)

Prop. Open Response	.**16** (.11)	.**12** (.12)	.**15** (.13)	.**29** (.15)
	.17 (.00-.38)	.06 (.00-.35)	.15 (.00-.45)	.32 (.03-.50)
Prop. Seeks Clarification	.**01** (.02)	.**02** (.04)	.**02** (.02)	.**01** (.03)
	.00 (.00-.09)	.00 (.00-.20)	.00 (.00-.08)	.00 (.00-.12)

Table 7: Summary of the multiple regression predicting average cross-examination resistance.

Step	***B***	***SE B***	***β***	***p***
STEP 1				
Constant	-.70	.67		.29
Age	.005	.003	.14	.08
IQ	.00	.003	.01	.90
Verbal memory	.002	.003	.07	.44
Memory trace	.03	.01	.16	**.04***
Performance version (A or B)	.08	.09	.08	.37
Cross-exam delay (months)	.04	.03	.14	.10
STEP 2				
Constant	-.08	.63		.90
Age	.003	.003	.07	.33
IQ	.004	.003	.10	.25
Verbal memory	.003	.003	.09	.29
Memory trace	.01	.01	.08	.32
Performance version (A or B)	.01	.08	.01	.92
Cross-exam delay (months)	-.02	.03	-.07	.47
Best-Practice-v-Verbal Labels	-.02	.11	-.015	.86
Best-Practice-v-Sketch-RC	.12	.11	.10	.25
Best-Practice-v-RI	.63	.13	.47	**<.001****

Table 8. Total number of barrister questions across the full cross-examination, and proportions (prop.) of each of the seven primary overarching types of questions for each interview condition. Mean proportions (SDs) are given on line 1 and medians (ranges) on line 2.

Scores	**Best-Practice (n = 65)**	**Verbal Labels (n = 40)**	**Sketch-RC (n = 38)**	**Registered Intermediary (n = 33)**
Total number of barrister questions	**58.92** (16.75) 59 (26-105)	**60.35** (22.42) 65.5 (25-117)	**58.26** (16.43) 56.5 (20-90)	**71.76** (18.03) 70.5 (28-109)
Prop. Invitation Open	**.05** (.04) .04 (.00-.22)	**.04** (.04) .02 (.00-.20)	**.05** (.05) .04 (.00-.2)1	**.12** (.05) .12 (.02-.21)
Prop. Invitation Closed (true)	**.24** (.11) .25 (.03-.51)	**.18** (.09) .17 (.04-.38)	**.20** (.10) .20 (.02-.43)	**.33** (.07) .34 (.15-.49)
Prop. Invitation Closed (false)	**.17** (.07) .16 (.05-.38)	**.15** (.05) .15 (.03-.28)	**.13** (.05) .13 (.02-.25)	**.16** (.05) .16 (.04-.29)
Prop. Assertion (true)	**.32** (.10) .32 (.15-.54)	**.36** (.09) .36 (.22-.57)	**.34** (.10) .34 (.17-.58)	**.21** (.06) .20 (.10-.33)
Prop. Assertion (false)	**.11** (.10) .07 (.00-.3)7	**.16** (.11) .15 (.02-.41)	**.14** (.11) .11 (.01-.39)	**.06** (.04) .06 (.00-.16)
Prop. Option-Posing	**.03** (.03) .03 (.00-.12)	**.02** (.03) .00 (.00-.08)	**.02** (.03) .00 (.00-.11)	**.015** (.02) .00 (.00-.06)
Prop. Other	**.08** (.07) .06 (.00-.23)	**.09** (.06) .08 (.01-.23)	**.11** (.07) .10 (.02-.24)	**.11** (.04) .11 (.02-.20)

Table 9. Proportions of features of barrister questions coded into 17 secondary categories for each interview condition. Categories are not mutually exclusive so overall proportions do not add to 1. Mean proportions (SDs) are given on line 1, medians (ranges) on line 2.

Question Feature classification	**Best-Practice (n = 65)**	**Verbal Labels (n = 40)**	**Sketch-RC (n = 38)**	**Registered Intermediary (n = 33)**
Tag	**.19** (.18)	**.28** (.22)	**.24** (.19)	**.04** (.03)
	.12 (.00-.77)	.22 (.04-.70)	.18 (.00-.70)	.05 (00-.11)
Credibility	**.07** (.05)	**.06** (.04)	**.05** (.04)	**.02** (.02)
	.07 (.00-.19)	.05 (.00-.16)	.05 (.00-.16)	.02 (.00-.08)
Negative	**.03** (.04)	**.04** (.05)	**.05** (.06)	**.04** (.04)
	.02 (.00-.14)	.02 (.00-.14)	.03 (.00-.20)	.03 (.00-.16)
Repetition	**.05** (.08)	**.08** (.11)	**.09** (.09)	**.12** (.07)
	.00 (.00-.28)	.00 (.00-.35)	.08 (.00-.35)	.13 (.00-.37)
Confirmation	**.18** (.13)	**.17** (.11)	**.17** (.13)	**.13** (.06)
	.16 (.00-.62)	.16 (.00-.55)	.14 (.00-.62)	.12 (.02-.25)
Clarification	**.02** (.02)	**.02** (.02)	**03** (.04)	**.02** (.02)
	.00 (.00-.10)	.00 (.00-.09)	.02 (.00-.16)	.02 (.00-.08)
Social influence	**.03** (.05)	**.04** (.05)	**.04** (.05)	**.09** (.05)
of another person	.02 (.00-.24)	.02 (.00-.16)	.02 (.00-.15)	.08 (.01-.24)
Possibility	**.04** (.05)	**.03** (.03)	**.03** (.03)	**.007** (.01)
	.03 (.00-.26)	.02 (.00-.11)	.02 (.00-.14)	.00 (.00-.04)
Complex	**.05** (.06)	**.06** (.08)	**.08** (.10)	**.08** (.06)
	.03 (.00-.30)	.02 (.00-.31)	.03 (.00-.40)	.08 (.00-.22)

Do you remember	.**14** (.09)	.**17** (.13)	.**16** (.10)	.**12** (.08)
	.12 (.03-.40)	.12 (.00-.47)	.15 (.03-.42)	.12 (.00-.39)
Signpost	.**13** (.05)	.**11** (.05)	.**13** (.05)	.**13** (.05)
	.13 (.00-.25)	.10 (.00-.23)	.12 (.03-.20)	.12 (.04-.29)
Praise	.**10** (.05)	.**10** (.07)	.**08** (.05)	.**04** (.04)
	.09 (.02-.24)	.08 (.02-.36)	.07 (.01-.20)	.03 (.00-.14)
Filler	.**004** (.01)	**.006** (.01)	.**01** (.02)	.**02** (.02)
	.00 (.00-.05)	.00 (.00-.06)	.00 (.00-.08)	.02 (.00-.07)
Name	.**08** (.05)	.**07** (.04)	.**06** (.04)	.**05** (.04)
	.07 (.00-.21)	.06 (.00-.16)	.06 (.00-.16)	.05 (.00-.12)
Reassurance	.**04** (.04)	.**02** (.03)	.**02** (.02)	**.006** (.01)
	.03 (.00-.16)	.02 (.00-.13)	.02 (.00-.06)	.00 (.00-.04)

Dear Professor Davies,

We are pleased to submit a revision of our manuscript (ID: ACP-21-0013) entitled 'Intermediaries and cross-examination resilience in children: The development of a novel experimental methodology'.

The comments from you and the reviewers were incredibly helpful. We have reflected carefully on all of the comments and suggestions and hope our efforts to address these in a comprehensive and thorough manner have been successful. Most notably, we have carried out a thorough reanalysis of the compliance data, incorporating a range of additional control variables into our regressions as suggested by Reviewer 2.

We have responded to your comments and the reviewers' comments individually below, with adaptations to the manuscript text highlighted in red.

Further, we published the original version of this paper as a pre-print discussion document at the end of March. We received some encouraging feedback from Dr Kevin Smith of the National Crime Agency, so have incorporated this into our revision. Specifically, there were a couple of minor errors spotted and a suggestion to include a comment that the Equal Treatment Bench Book (Judicial College, 2018, 2020) recommends: *"All young witnesses should ideally have an intermediary assessment as, no matter how advanced they appear, their language comprehension is likely to be less than that of an adult witness"* (para. 98, page 60). This is inserted into our discussion (bottom of p. 26-27).

The pre-print is available at: https://city.figshare.com/articles/preprint/Intermediaries_and_cross-examination_resilience_in_children_The_development_of_a_novel_experimental_methodology/13476201. We are happy to follow journal guidance such that, if our paper were to be accepted, we will link this pre-print to the final published version. If you have any further guidance, please just let us know and we will be happy to follow it.

We look forward to hearing from you in due course and very much hope you will find this version of the manuscript suitable for publication in *Applied Cognitive Psychology.*

Kind regards,

The Authors.

Responses to the Editor's Comments

Comment: The reviewers raise a range of issues, but I concentrate here on those most salient for me. I concur that the presence of the same RI at interview as at court is not common practice (though not unknown) and this limits the generalisation of any conclusions that can be drawn from the results. I also agree that this design feature has knock-on effects in terms of interpretation of any facilitation observed: does it arise from

the initial interview or from the cross-examination? And given that the RI instructed the barristers, as well as supporting the children during examination, does the RI impact on the child's answers or the barrister's questioning technique? Or both? It is difficult to determine from the existing data.

Response: Thank you for raising these issues, which we completely agree are of central importance. We respond to each of them in turn, below.

With respect to the same RI assisting communication throughout a criminal investigation, this is an important point and we are wondering whether the situation may differ somewhat for Registered Intermediaries and non-registered intermediaries? For this project we referred to the best practice guidance in the Registered Intermediary Procedural Guidance manual at the time of the study (Section 3.52, p.32, 2015): "The RI who assisted the witness at interview should, whenever possible, continue to assist the witness up to and including the trial." (The same advice is in the latest guidance, published in 2020). However, as you say, although recommended as best practice, we fully acknowledge that this does not always happen in real cases with RIs for a range of practical reasons. In order to clarify this issue further in the paper, we have now provided information that 18 of the 33 children in the RI condition had the same RI at investigative interview and at cross-examination; whereas the remaining 15 had a different RI for the cross-examination phase. In this respect, our study may be reasonably reflective of what happens in real cases. We acknowledge that this aspect of our design (and our rationale for it) should be more prominently highlighted and dealt with in more depth in the paper and have, therefore, explicitly noted this in the Method (page 13):

"As per recommendations for best practice in England and Wales at the time of the study (Registered Intermediary Procedural Guidance Manual, Ministry of Justice, 2015), children received RI assistance both at their initial interview and again at cross-examination. Of the 33 children in the RI condition, 18 were assisted by the same RI at both stages, which is also recommended best practice, and 15 had a different RI at cross-examination (although using exactly the same protocol). In real cases there is also likely to be some variability in whether the same RI is available for both stages."

We have also included a paragraph on the important issue of whether a facilitation effect with RI assistance might arise from the initial interview or from the cross-examination, and stress that this issue needs to be taken into account when interpreting the findings. We further note that using RIs at both points (interview and cross-examination) means that the applications of the current study are limited to real-life cases where this occurs (please see page 31, para 2). Further, those in the RI condition had a recall advantage initially - however, to deal with this we controlled for memory for the initial event, although this was not a significant predictor in the final model (please see same para which continues to page 32).

The final point about whether the RI impacts on the child's answers or the barrister's questioning technique or both is another complex and very interesting issue. We have included a new paragraph in the discussion specifically about this issue and suggested that

further research could unpick the important mechanisms underpinning the interplay between children's responses and barristers' questions. Please see pp. 30-31.

Reviewer 1

Comment: This article describes a rather complicated study on a topic of considerable importance. As noted in the introduction, there has been a great deal of field (and analogue) research on the cross examination of child witnesses (mostly victim witnesses), a body of literature complementing the extant literature on child witnesses' performance in forensic interviews. That literature is quite adequately reviewed in the introduction.

Response: We thank the reviewer for highlighting the importance of our chosen topic. We also accept that the manuscript is complicated. When revising our manuscript, we have endeavoured to ensure that our methods and procedures are as clear as possible for readers to follow, providing additional information where needed for clarity.

Comment: The study reported here involves an analogue situation rather than an examination of cross examinations involving actual alleged victims. That has the theoretical advantage of allowing researchers to assess the accuracy of the children's responses, but this advantage is not relevant because accuracy is not one of the variables explored. Instead, the focus is on the types of questions asked by the barristers and the extent to which the children resisted or acquiesced to suggestive/leading content introduced by the barristers. The key findings were that the barristers asked 'better' questions when the children were assisted by registered intermediaries (RIs) and that the children were more resistant/less suggestible in that condition. Knowing that RIs allow children to give better evidence would be an important finding for the criminal justice system, but I am not convinced that the study achieves this for a couple of reasons.

Firstly, all the children had been forensically interviewed an average of 11 months earlier in a variety of conditions, with 33 of those forensic interviews facilitated by RIs. All of those children, but only those children, were also assisted by an RI, indeed the same RI, during cross examination, creating a confound between RI condition and cross examination condition. Further, ground rules hearings (or a facsimile thereof—there was one hearing for all cases) were held only for those cases in which an RI was involved. These multiple and fatal forms of confounding preclude any conclusions about the effects of cross examination conditions or of the benefits associated with the use of RIs in the courtroom.

Response: The reviewer raises some important points about potential confounds, which we are pleased to be able to respond to and address.

The potential confound between the interview condition and cross-examination condition is very important, thank you for raising this. In this regard, we have replicated best practice recommendations (see earlier comment), in that RIs would assist witnesses at both interview and cross-examination. Best practice would also be to have the same RI for both stages, although as per our earlier comment we acknowledge that this does not always happen. In the current study, 18 of the 33 RI condition children received the same RI at both phases. In real cases, if the original RI becomes unavailable for the trial, a new RI

would be engaged. Further information about this issue is provided in the Method (page 13 – see earlier comment for text). We also emphasise this point again in the second paragraph of our discussion (page 26) when summarising the findings to make this clearer: *"As per recommendations for best practice in England and Wales at the time of the study (Ministry of Justice, 2015, see also current Registered Intermediary Procedural Guidance Manual, 2020), children received RI assistance at their initial interview and again at cross-examination."*

Finally, we include discussion of the fact that providing RI assistance at both investigate interview and cross-examination (although in line with guidance) limits our conclusions to those who have received RI assistance at both interview and cross-examination (pp. 31-32). *"Therefore, the current conclusions can only be applied to children who have had RI assistance throughout a criminal investigation which, in practice, is not always the case (RIs may sometimes only brought in at trial stage, although this is not recommended)."*

Second, the reviewer notes that ground rules hearings were only held for cases in which an RI was involved. We have now explained that there was a brief 'ground rules hearing' between the 'judge' and barrister prior to each individual cross-examination (with or without an RI), where the judge explained any important considerations to the barrister (e.g., age of child, any additional needs they had). This is now explicitly mentioned in the paper (see pages 11-12) and we apologise for not being clear about this originally. *"A female researcher was in a room with the child at their school and partially took on the role of 'judge'. We could not entirely replicate the judge role as we had no facility for the child to view the judge only via the screen - and for ethical reasons the researcher had to be with the child - so this aspect of the study must be viewed as approximate to real-life. There was a brief 'ground rules hearing' between the judge and the barrister prior to each individual cross-examination (with or without an RI) where the judge explained any important considerations to the barrister (e.g., age of child, any additional needs they had)."*

In the discussion we note that the brevity of our ground rules hearings for non-RI children is a limitation, i.e., that more updated guidance published after our study was conducted recommends ground rules hearings as 'good practice' for all young witnesses (see page 32). *"Our ground rules hearings for non-RI children were also brief, and more recent guidance now recommends they are included as 'good practice' for all young witnesses (Judicial College, 2018, revisions 2020, Equal Treatment Bench Book, p.64)."*

We agree that these are important points, but feel that they do not preclude exploratory conclusions about the effects of cross-examination conditions or the benefits associated with the use of the RIs. In order to highlight that our findings should be regarded as 'exploratory', we have adjusted the text in several places to emphasise this: please see the abstract, introduction (p. 5), discussion (p. 26), and conclusion (p. 33).

Comment: Secondly, the 'target event' about which the children were interviewed was a mock crime that they had witnessed (live or on a video). That's not a very good analogue for a traumatic experienced event and it's likely that most children had very little memory of the 'event' before their memories were refreshed by listening to an audio recording of the initial forensic interview. Realistically, the study explored the children's memories of

listening to prior accounts of the 'event' rather than memories of the event per se. Taken together, the study's ecological validity is extremely suspect on multiple grounds: the event was witnessed not experienced, it was minimally salient, it's very unlikely in the real world that children would be assisted by the same RI in those situations, and current court procedures require ground rules hearings in all cases with vulnerable witnesses, not only those with RIs.

Response: We appreciate these insightful comments and have responded to each one below.

Regarding the target event, we hope that we have not tried to claim that our event was an "analogue for a traumatic experienced event", but apologise if this did not come across clearly. We would still argue that our mild but detailed minor crime event remains relevant, because child witnesses do not only appear in court in relation to traumatic events. Children can be victims and witnesses to a range of offences and may be required to give an ABE interview and subsequently attend Court for non-traumatic offending. We have now emphasised much earlier on that this was a mild minor crime event for ethical reasons, so as not to upset the children (please see bottom of page 7: *"For ethical reasons this was a mild minor crime event."*), and return to this important issue on page 32 in the discussion where we note this as a limitation:
"We used a mild minor crime event that took place in a familiar environment (the children's school), so were unable to replicate the anxiety, unfamiliarity and potential trauma of a real court case, which limits generalisation of the findings to real cases."

Regarding the refreshing of the children's evidence, thank you for raising this point as we completely agree that this needs further clarification. In this regard, we have followed current practice as closely as was possible, as memory refreshing occurs for all witnesses (via being able to re-read their statements or watching ABE videos). Rule 4.49 (and following) of the Ministry of Justice's 2011 'Achieving Best Evidence in Criminal Proceedings Guidance on interviewing victims and witnesses, and guidance on using special measures" https://www.cps.gov.uk/sites/default/files/documents/legal_guidance/best_evidence_in_criminal_proceedings.pdf explains the importance of memory refreshing. The CPS guide for Prosecutors explains the role of memory refreshing for witnesses - https://www.cps.gov.uk/legal-guidance/speaking-witnesses-court. Finally, the Registered Intermediary Procedural Guidance Manual from 2015 (the one we referred to for our study) explicitly requires memory refreshing (with the RI present in RI cases), plus a third party, in our case, the researcher acting as 'judge': https://www.theadvocatesgateway.org/images/procedures/registered-intermediary-procedural-guidance-manual.pdf.

Therefore, in order to clarify this important point we have referred to two of the above documents in the paper, and more fully justified our methodological approach, which we hope is now clearer (please see page 9).
"Prior to the cross-examination, children were 'refreshed' on their evidence as per Achieving Best Evidence guidance (Ministry of Justice, 2011) and the Registered Intermediary Procedural Guidance Manual (Ministry of Justice, 2015)."

We have also explicitly noted that we are exploring a combination of the children's memory of the event and their memory of the interview at refreshing (please see page 9): *"Therefore, as in real-life, cross-examination performance may draw upon original memories of the event and recent memories of the refreshed interview."*

With respect to the same RI assisting communication throughout a criminal investigation, please see earlier comment explaining our rationale for the changes and additions to the paper in relation to this important issue (on page 13).

Regarding ground rules hearings, please see our response to the comment preceding this one, thank you.

We hope our responses have explained more clearly that the procedures of the study do have strong ecological validity (e.g., that not all child witnesses directly experience an event or are called as witnesses to traumatic events, that although best practice is to have the same RI at interview and at court this does not always happen (we reflect this in our sample), and that ground rules hearings did take place for all participants – albeit short - and we have emphasised these issues in our discussion as potential limitations.

The reviewer also comments that "it is likely that most children had very little memory of the 'event'". This is also a very important point to consider. Anecdotally, our barristers confirmed that the child witnesses clearly remembered the incident and not just the memory refreshing. However, in the case of one witness, the child clearly remembered nothing of the event and answered no cross-examination questions, so for this reason we specifically excluded this child from our sample (please see participants section, page 7: *"one further child (a girl) was excluded because she did not respond to any cross-examination questions"*).

Comment: Most of the findings were predictable based on the extant literature on interviewing and cross examination and, as indicated above, the design precluded conclusions about the possible effects of RI involvement. As a result, the study does not make a meaningful contribution to the literature.

Response: We do appreciate the reviewer's concerns in this regard. However, we argue that, albeit in an exploratory manner, with some limitations (all of which we have been explicit about – see pages 31-33), the paper provides the first empirical evidence for the use of RIs at the cross-examination stage. As noted by reviewer 2, it is difficult in a single lab-based study to raise all issues of ecological validity but we feel that this paper provides a good basis for future research to build upon.

Reviewer 2

Comment: This is a very important topic. It's also a very involved study—there are lots of data to contend with. I commend the authors for their concerted attempt to nail down the delicate balance between ecological validity and experimental rigour. However, I do have a slight concern that the study has lost a little of the latter—just because it's part of an even

larger study with myriad other manipulations and measures. In that respect, I wondered whether a monograph would have been a better way to present all of the findings together, thereby avoiding a kind of salami-slicing approach (albeit with an enormous salami) where it's difficult to know where to make the cuts, and where we can't guarantee that readers have a full appreciation of what else has gone on for the participants.

Response: We thank the reviewer for acknowledging the importance or the topic and the challenges of conducting research in this area. We also appreciate the suggestion of publishing this work as a monograph, which we would have been keen to adopt had we not been in a situation where most of the studies have been published already.

Major comments:

Comment: While I really admire the authors' attempt to look at a great many issues, I wonder if they attempted too much. The design is deceptively complex and apparently lacks some counterbalancing (event version, live versus video-taped event, investigative interviewing group) that I would expect to see in a more focused study. Although the authors control for a few variables (e.g., the sizable age range, some individual difference measures, a rough measure of memory trace), there are other variables that haven't been controlled for. Where are the controls for delay to cross-examination (there is a pretty big range)? Pre-cross-examination anxiety? Barrister? While at first glance the sample seems large, I'm not convinced the authors have the statistical power to control for everything they really should be controlling for (I couldn't see a power analysis for multiple regression with multiple control variables).

Response: We appreciate that the design appears quite complex and have been careful to ensure that our revision is as clear as possible. With respect to the important issue of counterbalancing, we feel that the counterbalancing for event version was sound, since this was virtually half and half. However, we have now included event version as a further control variable in the primary regression analysis (please see comment on page 7 within the footnote, new analyses on pages 19-20, and revised Table 7). We also acknowledge that we could not counterbalance for live vs video event stringently, but to control for this, we have run our regression analysis with (n=176) and without (n=144) the 'video' children and included a note about this in the results (please see comment on page 7 within the footnote and results on page 20). Importantly, these factors did not affect the findings. Interview condition was intentionally somewhat unbalanced as we wanted to include more children in our reference condition, the Best-Practice interview; otherwise, numbers across different conditions were similar.

We thank the reviewer for suggesting that delay to cross-examination should have been controlled too. We completely agree and apologise for overlooking this. We have now included it as a control measure in our regressions (see new analyses on pages 19-20). We have also included a comment about this on page 18 and added it as a variable to Table 1: *"As some variability in this delay emerged across conditions (see Table 1) due to timing of school holidays and availability of RIs/barristers, we controlled for delay in the primary statistical analyses."*

With respect to barrister, this was not as evenly distributed across interview conditions as we would have liked, but when we included this as a further control variable in an exploratory regression, results were similar and the key finding was unchanged (see footnote on page 20). Therefore, encouragingly, the central results remain unchanged with all of these additional controls.

We have also added more information to justify our sample size and statistical power for the analyses we have conducted (page 19):
"With nine predictor variables in total, Green (1991) would recommend a sample size of at least 122, thus for the current regression our sample size exceeded the minimum numbers recommended."

We will take up the issue of anxiety in a later response.

Comment: It strikes me that the RI could have "worked" in plenty of different ways, and while from an ecological validity standpoint we could argue that it doesn't really matter, I would much rather see research that was better placed to elicit mechanism. For example, what if the whole effect of the RI was somehow due to having them at the investigative interview? Or children's increased contact with an unfamiliar adult? Both are admittedly unlikely, but they are things a good experimental design would allow us to rule out, and this design doesn't.

Response: We completely agree with these comments – in this study it was not possible to elicit RI mechanism/s. We want to stress that this was the first empirical study of RIs, and part of our aim was to ascertain whether we could look at RIs in a laboratory-based study and trial a new way of conducting experimental cross-examinations. We have added a paragraph about this important issue in the discussion and suggest the importance of future work addressing such mechanisms (please see pages 30-31).

Comment: I would like to see an acknowledgement, and analysis, of the bidirectional influences between barristers' questions and children's responses. In a non-scripted situation—where the child's responses could easily be influencing the questions—it seems like the ideal time to examine how the RI could influence both the barrister's reaction to the child and the child's reaction to the barrister. A sequential analysis would be a good way to do this.

Response: We agree about the bidirectional influences between barristers' questions and children's responses. We want to emphasise that the very nature of cross-examination requires some fluidity in questioning and a good advocate will always be influenced by the child's responses. The exception to this would be to use a rigid script of questions (which is necessary in some extreme cases, but not generally). Otherwise, the barrister will be flexible and adapt in response to the child's answers - this was one of the advantages to our novel approach to assessing cross-examination empirically, which has, to our knowledge, not been addressed in previous empirical work on cross-examination. Whilst assessing this relationship further could be useful, we are conscious that this would not be a straightforward analysis in what the reviewer has already described as a "deceptively complex" study. For example, because of our ecologically valid approach in allowing

barristers some flexibility with how they approached questioning, there is a lack of consistency in terms of how questions were posed/elaborated on that makes it difficult to draw definitive statistical conclusions. We have, however, noted in the discussion that an important area for future work would be to more thoroughly consider the interplay between barrister's questions and children's responses (see pages 30-31).

Comment: The focus of the analysis was on children's resistance to false information. Information about how often children went along with false suggestions is interesting, but not particularly telling without good comparative data/analysis on what happens to children's errors under cross-examination. This allows a diagnostic approach to response changes; a complete picture. There's a feeling in this manuscript that resistance is always good and compliance is always bad. What if the RI makes children resistant to any kind of response change? Is that necessarily good?

Response: We thank the reviewer for raising this important point. We agree that resistance is not always good, and compliance not always bad and have included some discussion of this on page 26. We would also add that we deliberately added four questions in our interview protocol that aimed to elicit correct information from the children (i.e., items to which they should agree with the barrister, if their memory was correct). However, the points we analysed from the cross-examination were all based upon challenging the child's account by using demonstrably incorrect facts (e.g., 'a woman helped us by setting up the video camera at the back' – when no woman had been present and the camera had in fact been set up by one of the male actors) so resistance was important. This was not a gentle challenge to the child's recollection but trying to implant a completely false narrative. The more errors that can be highlighted in a witness's evidence, the less reliable they will be seen as being, even when they are then clear and adamant about something salient; one of the main jobs of the person cross-examining is to cast doubts on the reliability of the witness's recollection. Please see comment inserted on page 26:
"For a child to accept that it was "possible", for example, that a woman had helped set up the video camera (when no such woman was present) would be enough to be used by the defence lawyer in undermining the evidence given the burden and standard of proof in criminal trials".

To further address the reviewer's concerns and emphasise the key issues here, we have carefully reviewed the manuscript to ensure that our wording is more balanced, emphasising that when we talk about compliance, this is compliance with 'false information' - so in this instance it is undesirable. We have also explicitly noted that resistance is not always good and compliance bad, but that changes to children's responses as a result of barrister questioning can cause difficulties (please see p. 26 - footnote on this page).
"Although our study specifically looked at compliant responses to 'false information', which are undesirable, in some cases such responses would be appropriate if the information were true."

Minor Comments

Comment: Why did some children see the event on videotape when this wasn't a

manipulation? I'm not sure I buy your argument/t-test that there's no meaningful difference between them; it's a 15% difference.

Response: As we note in the manuscript, this was a pragmatic decision (due to our actors needing to withdraw from the project). We have now run our regressions with and without the children who saw the event on video to deal with this important issue. Results remain unchanged (please see pp. 19-20 in the results).

Comment: Readers are given relatively little information on how barristers framed their challenges—we hear about "credibility challenges" and "possibility" but those kinds of codes don't give us content information. Did RIs intervene at the level of challenge content (as opposed to format)? I think more detail is necessary here.

Response: Apologies for the missing details. We have now clarified that RIs did not intervene about the content of the questions, but rather about question format:
As with all of our methodological decisions, this was to enhance ecological validity: RIs do not restrict what is being asked, but how it is asked (Registered Intermediary Procedural Guidance Manual, 2015). Please see page 15 for comments about this issue and examples: *"Importantly, RIs did not intervene about the content of the questions but rather the format (Ministry of Justice, 2015), for example, "[Barrister's name], could that question be rephrased, as you know it's a tagged question" or if they thought the child would not understand the question (e.g., "I am not sure [child's name] will understand that complex question"."*

Comment: Did RI presence change anxiety scores? (Sorry if I missed this.)

Response: We have now clarified in the paper that the collection of information on child anxiety was not part of the experiment as such. Rather, it was for purposes of ethics: to ensure the children felt comfortable with the cross-examination (before, during and after), so we could monitor their wellbeing and respond appropriately. We had no predictions regarding anxiety and it was not part of the experimental design. As such, we do not feel that it would be appropriate to report these data as though they were part of our study, or to link them to our independent variables. Please see page 13 for a comment about this: *"Note that these anxiety ratings were not study variables but introduced for ethical reasons."*

Comment: The authors state that they used a video conferencing approach to mimic the "live link", but in a real live link, the judge wouldn't be with the child, right? I was confused here.

Response: Apologies for the confusion here. The reviewer is correct in that the child on live link would be in a separate private room to the Court room, so not with the judge (instead, they would be able to see the judge on screen). For pragmatic reasons, the research assistant in the room with the child took on the role of judge (this was the case for ALL interviews, with or without an RI, so would not have affected our experimental manipulation). We have now explained this more clearly in the manuscript (please see pages 11-12, and earlier comment for inserted text).

Comment: I wanted more information about the instances in which the barristers didn't pose all the challenges. What were the reasons, and what were the ns for each reason? (And if it was child anxiety, was that consistent with children's self-report anxiety?)

Response: To enhance ecological validity of the study, we were keen to give barristers flexibility in posing questions to the child witnesses. This decision was made because, in real life, barristers have to make a judgement call about how much/little to press witnesses – barristers do not take a fixed approach, and we wanted to reflect this. We have now added some additional information to the paper to explain this more fully (please see top of page 11):
"On a few occasions, barristers judged that it was not necessary to pose all challenges to the children. In real life, barristers make judgements about how much/little to press a witness and do not take a fixed approach, so the present study aimed to reflect this."

Comment: The memory trace scoring felt a bit vague. Could this be coded in a more objective way, with a measure of inter-rater reliability added? Related question: should errors of commission and omission be treated differently when doing this—if so, how?

Response: We apologise that the reviewer did not see the Supporting Information we uploaded with the submission which gives detail about how memory trace scoring was conducted. In this, we have provided information about our approach memory trace coding (to ensure replicability). However, to allay the reviewer's concern, we have added further information in the text (please see page 11).
"These 'memory trace' scores were allocated for full (3), moderate (2), partial (1) or no (0) knowledge about six of the false points in terms of degree of information recalled in the investigative interview. For one other point (false point 5), this was a complete confabulation about something that did not happen at all in the event, therefore, a score of 0 was allocated for all children because it was not possible to code this item in terms of original recall of information (Maximum memory trace score=18: see Table 1 for mean memory trace scores and Supporting Information for full details of the coding scheme)."

We were also able to get inter-rater reliability for 15% of transcripts. This has been included on page 11:
"Fifteen percent of the transcripts were independently coded by a second rater and intra-class correlations for information pertaining to each of the challenges ranged from .89 to 1.00, indicating excellent inter-rater reliability."

We coded positively here for presence of correct information on six false items and gave no score for the confabulated item (false item 5), so memory trace coding did not capture errors. In order to capture errors, we would have had to use percent accuracy for each item – taking into account correct and incorrect information - which could be considered for future research.