

## The impact of multiple interviews on the accuracy and narrative coherence of children's memories

Zsafia A. Szojka, Annabelle Nicol, Non Davis Frenguelli, David La Rooy & Hayden Henderson

To cite this article: Zsafia A. Szojka, Annabelle Nicol, Non Davis Frenguelli, David La Rooy & Hayden Henderson (2021): The impact of multiple interviews on the accuracy and narrative coherence of children's memories, Psychology, Crime & Law, DOI: [10.1080/1068316X.2021.1984479](https://doi.org/10.1080/1068316X.2021.1984479)

To link to this article: <https://doi.org/10.1080/1068316X.2021.1984479>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 18 Oct 2021.



[Submit your article to this journal](#)



Article views: 93




[View related articles](#)



[View Crossmark data](#)

## The impact of multiple interviews on the accuracy and narrative coherence of children's memories

Zsofia A. Szojka <sup>a\*</sup>, Annabelle Nicol<sup>b</sup>, Non Davis Frenguelli<sup>a\*\*,</sup>, David La Rooy<sup>a#</sup> and Hayden Henderson<sup>c</sup>

<sup>a</sup>Royal Holloway, University of London, London, UK; <sup>b</sup>Abertay University, Dundee, UK; <sup>c</sup>University of Southern California, Los Angeles, CA, USA

### ABSTRACT

This study investigated the accuracy and narrative coherence of children's accounts of a staged event across two interviews in comparison to a control condition to discern between the effects of repeated recall and delay between interviews. Seventy-six 8–11-year-olds took part in a first aid training session. Half of the children were randomly assigned to be interviewed using open-ended questions twice, one week after the event and five weeks after the event, whilst the other half were interviewed only once, five weeks after the event. Supporting the hypotheses, children reported more details over the course of two interviews than in a single interview either 1-week or 5-weeks after the event, and details that remained consistent across the two interviews were more accurate than reminisced details. The increased completeness of children's accounts in two interviews was accompanied by an increase in the use of markers of causal-temporal connectedness. The hypothesis regarding the negative effect of delay on the accuracy of children's testimony was partially supported, as details reported in the first, 1-week interview were more accurate than details in the single 5-week interview. Results demonstrate that multiple interviews can increase the narrative coherence of children's testimony without decreasing their accuracy.

### ARTICLE HISTORY

Received 12 June 2020  
Accepted 20 July 2021

### KEYWORDS

Children's memory; multiple interviews; narrative coherence; accuracy


Contrary to *Achieving Best Evidence* (ABE) guidelines (Ministry of Justice, 2011) stating that the video-recording of interviews should minimise the need for witnesses to repeat their accounts, many child witnesses in the United Kingdom are interviewed more than once (Plotnikoff & Woolfson, 2001; Szojka et al., 2020; Waterhouse et al., 2016). Frequent reasons underlying the decision to conduct further interviews include non-disclosure in the initial interview, the disclosure of new details after the interview and additional evidence surfacing that appears to contradict the child's statements (Waterhouse et al.,

**CONTACT** Zsofia A. Szojka  szojka.zsofia@gmail.com  University of Southern California Gould School of Law, 699 W Exposition Blvd, Los Angeles, CA 90089, USA

\*Present address: University of Southern California, Los Angeles, CA, USA

\*\*Present address: Swansea University, Swansea, UK

#Present address: Department of Psychological Medicine, University of Otago, Dunedin, New Zealand

 Supplemental data for this article can be accessed <https://doi.org/10.1080/1068316X.2021.1984479>.

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

2016). Early research on children's memory generally viewed multiple interviews as a method of suggestion, and therefore, a risk factor potentially decreasing the accuracy of children's testimony (Ceci & Bruck, 1993). However, many of the key studies concluding that multiple interviews decrease the accuracy of children's accounts have relied on a combination of multiple interviews with other methods of suggestion, such as leading questions, social pressure or misleading physical props (e.g. Cassel & Bjorklund, 1995; Leichtman & Ceci, 1995; Ornstein et al., 1992). For instance, in Leichtman and Ceci's (1995) 'Sam Stone' study, the accuracy of children's accounts remained high across multiple interviews in the control condition where no suggestive methods were used.

Reviews of research on the impact of multiple interviews on children's testimony have concluded that children's recall across interviews can remain highly accurate when interviewers rely on open-ended questioning methods (Goodman & Quas, 2008; La Rooy et al., 2009). Furthermore, findings from both experimental (e.g. Knutsson et al., 2011; La Rooy et al., 2005) and field research (e.g. Hershkowitz & Terner, 2007; Katz & Hershkowitz, 2012; Leander, 2010; Waterhouse et al., 2016) have shown that multiple interviews frequently allow children to disclose new details that were not reported in the initial interview. However, the addition of new details and omission of previously mentioned details decreases the consistency of multiple interviews, which might compromise the credibility of children's testimony (Brewer et al., 1999; Stromwall & Granhag, 2005; Szojka et al., 2017). Using a narrative coherence framework to integrate measures of the completeness, consistency and connectedness of children's testimonies, Szojka et al. (2020) found that multiple forensic interviews affected the narrative coherence of testimonies positively, as they increased the completeness of witnesses' testimonies, without compromising their consistency and causal-temporal connectedness. However, due to researchers' ignorance of 'ground truth' in field studies, the accuracy of children's recall could not be measured. As previous findings on the relationship between the narrative coherence and accuracy of children's recall have been inconsistent (Chae et al., 2016; Kulkofsky et al., 2008; Wang et al., 2015), field studies need to be complemented by controlled experimental research where the impact of multiple interviews on the narrative coherence and accuracy of children's recall can be established.

## The accuracy of children's testimony

Research suggests that schoolchildren and even pre-schoolers are able to provide accurate descriptions of events they experienced, however, the accuracy of children's recall is influenced by a wide range of factors, including their developmental level, the type of event they are asked to recall and the style of questioning used by the interviewer (for a review see Lamb et al., 2008). Children's accounts are most accurate in response to open-ended questions, whilst suggestive and overly specific questions compromise accuracy (e.g. Lamb et al., 2007b; Leichtman & Ceci, 1995), especially for pre-schoolers, who appear more susceptible to the effects of suggestion and closed questioning (Ceci & Bruck, 1993). However, these rules are not absolute; in some studies, pre-schoolers were remarkably resistant to misleading questioning (Goodman et al., 1991; Saywitz et al., 1991), indicating that children can monitor the accuracy of their recall from a very young age when details to be remembered are distinctive (Ghetti et al., 2002). On the other hand, even truthful and overall accurate statements elicited through open-

ended questioning can include a substantial number of inaccuracies (Brubacher et al., 2019). Among other mechanisms, inaccuracies appear in children's statements through intrusions, resulting from a failure to inhibit event-irrelevant thoughts, and through conjunctions, details of other events falsely reported as relating to the current event due to source monitoring errors (Brubacher et al., 2019; Johnson et al., 1993). Price et al.'s (2016) research has shown that children who experienced repeated instances of the same event or similar events are particularly susceptible to the latter type of error (Price et al., 2016).

Reviews of studies using open-ended, non-suggestive questioning cast doubt on the view that multiple interviews inherently compromise the accuracy of children's testimony (Goodman & Quas, 2008; La Rooy et al., 2009). For instance, Peterson (2015) found that children's accounts of an injury requiring a hospital stay remained accurate in multiple interviews, even when the delay between interviews was several years long. Furthermore, some studies found that multiple interviews can improve the accuracy of children's testimony, either through an 'inoculation effect' resulting from memory refreshment interviews across long delays (Brown et al., 2015), or through providing children with an opportunity to recall events in an open-ended interview following suggestive questioning in a previous session (Melinder et al., 2010). However, regardless of their reliance on suggestive or open-ended methodologies, only a small proportion of studies compared the accuracy of recall in accounts obtained in multiple interviews with the recall of a control group interviewed a single time, matched for delay (Brown et al., 2015; Goodman et al., 1991; Knutsson et al., 2011; Quas et al., 2007). Without the inclusion of a delayed control group, it is not possible to distinguish the effects of delay from the effects of multiple recall occasions. Thus, in field studies investigating the effect of multiple interviews on the accuracy of children's testimonies, researchers cannot control for the effects of delay. By definition, more time will have elapsed since the events children are asked about by the second interview than the first one, therefore, potential changes in the accuracy of testimonies may result from forgetting due to delay, rather than from multiple recall efforts. However, experimental studies can overcome this limitation by involving a delayed control condition; a group of participants interviewed on single occasion at the same time as the second or third interview of the group assigned to the multiple interviews condition. Although the relationship between delay and memory accuracy may seem straightforward, studies investigating this relationship have reported remarkably inconsistent findings. For example, delays on the scale of weeks did not exert a negative impact on accuracy in some studies (Hubbard et al., 2016; Knutsson et al., 2011), whilst other studies found a decrease in accuracy after comparable delays (Dietze et al., 2013; Salmon et al., 2012).

### **Narrative coherence and accuracy**

In cognitive psychology, the concept of narrative coherence is generally used to describe the extent to which details in a story are connected either at the 'local' level of individual clauses (Kulkofsky et al., 2008) or at the 'global' level characterising the entire narrative (Peterson et al., 2014). However, in legal contexts, narrative coherence is a broader construct, encompassing not only the extent to which details are connected, but also the completeness and consistency of the story presented (Pennington & Hastie, 1992). In line with the predictions of the story-telling model (Pennington & Hastie, 1992), previous

research has shown that each of the three dimensions of narrative coherence, including completeness (Kyriakidou et al., 2014), consistency (Brewer et al., 1999; Connolly et al., 2008; Stromwall & Granhag, 2005) and causal-temporal connectedness (Mugno et al., 2016; Olaguez & Klemfuss, 2020; Voss et al., 1999) influence the perceived credibility of the testimony which determines the likelihood of charging, prosecution and conviction. However, the relationship between each element of narrative coherence and the accuracy of children's testimony is controversial and contingent upon other characteristics of the testimony, including the delay between interviews.

From the three elements of narrative coherence, consistency has been the most widely researched in relation to the accuracy of children's testimony (Baugerud et al., 2014; La Rooy et al., 2005; Price et al., 2016). Most studies found that details that remained consistent across multiple interviews were more accurate than reminisced details in subsequent interviews (e.g. La Rooy et al., 2005; Pipe et al., 1999; Salmon & Pipe, 2000). However, laboratory research suggests that whilst consistency correlates with accuracy at the level of individual details, the overall consistency of witness statements does not predict their accuracy (Baugerud et al., 2014; Gilbert & Fisher, 2006). Furthermore, the relationship between consistency and accuracy in multiple interviews is influenced by delay (La Rooy et al., 2005). In La Rooy et al.'s (2005) study examining the effect of delay on the accuracy of reminisced details, 92% of newly recalled information was accurate when the delay between the event to be recalled and the interview was short, and interviews were conducted in less than a day apart. However, the accuracy of information decreased to 72% when the interviews were conducted less than a day apart but 6 months after the event, and to 56% when the interviews were separated by a 6-month delay (La Rooy et al., 2005). In contrast, when children were asked about a real-life stressful event (removal from their families by the Child Protective Services) in Baugerud et al.'s study (2014), newly recalled information remained highly accurate both 1 week and 3 months after the event. Based on these results, Baugerud et al. (2014) suggest that low consistency across multiple interviews does not necessarily implicate low accuracy.

In contrast to consistency, the relationship between connectedness and accuracy has only been studied in single interviews so far. Narrative structure theory predicts that establishing causal-temporal connections allows children to organise and represent memory details more effectively, leading to both increased completeness and increased accuracy of recall (Nelson & Fivush, 2004). However, in Kulkofsky et al.'s (2008) study examining 3- to 5-year olds' recall of a play event, the accuracy of children's accounts was negatively correlated with the extent to which their narratives included causal-temporal connections. Kulkofsky et al. (2008) suggest that this tendency resulted from a quantity-accuracy trade-off, in that a high degree of causal-temporal connectedness allowed children to recall more details, but increased completeness led to an increase in the proportion of incorrect information. Consistent with this theory, when the completeness of children's accounts was controlled for, increased connectedness was associated with improved accuracy (Kulkofsky et al., 2008). Kulkofsky et al. (2008) also examined the potential effect of delay on the relationship between narrative coherence and accuracy by manipulating the delay between the play session children participated in and the interview to test children's recall either 1 week or 1 month after the event. The researchers expected that accuracy would decrease and the negative correlation between the connectedness and completeness of children's account and their accuracy would be enhanced after the

longer delay, due to the increased influence of forgetting and memory reconstruction (Kulkofsky et al., 2008). However, in contrast to their hypothesis, the length of delay between the event and the interview had no effect on the accuracy of children's testimony, although narratives were less connected in the monthlong delay condition.

Other studies reported a more straightforward relationship between narrative coherence and accuracy, finding that both completeness and causal-temporal connectedness contribute to increased accuracy in children's descriptions of staged events (Chae et al., 2016; Wang et al., 2015). Chae et al.'s (2016) found that preschool-aged children who provided more complete and more connected accounts about a staged conflict also provided more accurate information both in their free recall and their responses to direct questions. This overall positive relationship between narrative coherence and accuracy is consistent with narrative structure theory (Nelson & Fivush, 2004). Contrasting findings regarding the relationship between components of narrative coherence and the accuracy of children's accounts may be partially accounted for by the types of event used. Children may put more emphasis on monitoring and controlling the accuracy of their accounts when reporting a real-life event with potential consequences than in their descriptions of a play event (Chae et al., 2016).

Overall, research suggests that separate components of narrative coherence may exert different, and sometimes contrasting effects on the accuracy of children's accounts. Whilst causal-temporal connectedness appears to improve the accuracy of children's accounts, increased completeness may, under some circumstances, lead to decreased accuracy. Additionally, while consistent accounts are sometimes found to be more accurate, this relationship is contingent upon the type of event children are asked to remember and the delay between the event and the interviews.

## The present study

The present study aims to follow up Szojka et al.'s (2020) findings suggesting that multiple interviews increase the narrative coherence of children's testimony through investigating the effect of multiple interviews on the narrative coherence and accuracy of children's recall of a first aid training event. Whilst previous studies have examined the impact of multiple interviews on the completeness, consistency and accuracy of children's accounts (e.g. Baugerud et al., 2014; La Rooy et al., 2005; Price et al., 2016), the extent of causal-temporal connections across multiple recall occasions has received less attention, despite past research showing a positive relationship between connectedness and accuracy (Chae et al., 2016). Previous findings indicate that elements of narrative coherence develop substantially across middle childhood (Habermas & de Silveira, 2008; Reese et al., 2011), therefore, the present sample included children between the ages of 8 and 11 years.

When conducting multiple interviews, investigators can use the cumulative amount of information recalled in the two interviews rather than only the contents of the second interview. Therefore, when assessing the completeness, accuracy and connectedness of information available to interviewers after the second interview, children's overall recall across the two interviews was compared with their recall in the first interview one week after the event, and the control condition consisting of a single interview 5 weeks after the event. Completeness was operationalised as the number of correct details in children's narrative, while accuracy refers to the proportion of correct details from all of the details

each child reported. Connectedness was measured through linguistic markers of simple and complex temporal connections, causal relations and optional states.

We hypothesised that (1) information provided over the course of the two interviews would be more complete than information provided in a single interview, either 1 week or 5 weeks after the event, however, (2) both information provided over the course of the two interviews and information provided in the single 5-week interview would be less accurate than information provided in the 1-week interview due to the increase in inaccurate details over the 5-week delay. We further expected that (3) consistent details would be more accurate than reminisced details in the overall recall of the experimental group. Due to the scarcity of research on the effect of multiple interviews and delay on the connectedness of children's testimony, our analyses assessing the frequency of markers of connectedness were exploratory. Finally, we explored whether children's age and gender affected the completeness, consistency, connectedness and accuracy of their recall.

## Method

### Participants

Ninety-seven children between the ages of 8 and 11 ( $M = 9.34$ ,  $SD = .98$ ) were recruited for the study. Eighteen children who took part in the first aid training were not present at the date of the interviews, therefore, no data was collected from them. Data was erroneously collected from three children who were younger than 8-years-old, but their data was not included in the analyses. The final sample of 76 children included boys ( $N = 34$ ) and girls ( $N = 42$ ), recruited from an after-school club ( $N = 21$ ) and a primary school ( $N = 55$ ).

### Design and procedure

Permission to conduct the study was granted by the university's ethics committee ahead of the start of data collection. Children in the experimental group ( $N = 41$ ) were interviewed 1 and 5 weeks after the event to-be-remembered whilst children in the delayed control group ( $N = 35$ ) were only interviewed 5 weeks after.

Initial contact with schools and after school clubs was made via an information pack sent by post, which was followed up by a personal meeting if gatekeepers expressed interest in the study. Based on interest and availability, two locations were chosen for the study: a primary school in Essex and an after-school club in Surrey. Once schools' participation in the study was confirmed, teachers were asked to distribute the information leaflet and consent form to parents. The consent form was returned by 97 of the 130 eligible children at the two locations. At the start of the first aid training and the interviews, the activities involved in the study were explained to the children, and their verbal assent was obtained. Each location was visited three times, first to deliver the first aid session, then one week later to interview the children about their experiences, and finally five weeks later for the follow-up interview.

### First aid training

Children took part in pairs in a 15-minute first aid training session held by the third author. First aid training has been used as the event-to-be-remembered in previous research

(Brown et al., 2018) and was chosen for the present study due to its value for the children and schools involved. The first aid training session was developed using an altered version of the 'Life. Live it' workshops developed by The British Red Cross specifically for children. The teaching materials for the workshops are available on The British Red Cross website and permission was obtained from The British Red Cross to use the materials for the purpose of this study. The first aid training session involved four stages of different activities. In the first stage, the research assistant introduced the children to the content of a first aid box and explained how to treat minor injuries. During this activity, children tried to put a plaster and an eye pad on themselves. The second stage consisted of a sorting activity, where children were asked to discern between minor injuries and emergencies requiring urgent medical care. In the third stage, children watched two videos demonstrating how to help someone who is choking or unconscious. In the final stage of the session, children practised calling an ambulance through role play. Following the first aid session, children were rewarded with a certificate.

## Interviews

Children were pseudorandomly assigned to the single or multiple interviewing condition by compiling separate alphabetical lists of girls and boys in each age group and selecting every second child for the experimental condition. One week after the training session, the first author interviewed children assigned to the experimental group individually about their experiences of the first aid training. The interview protocol was developed based on the National Institute of Child Health and Development (NICHD) protocol. The NICHD protocol is an effective method of eliciting detailed and accurate testimonies from child witnesses (Lamb et al., 2007a) and the protocol has been adopted by investigators in several countries (La Rooy et al., 2015). The interviews were audio recorded. Each interview started with rapport building and episodic memory training, followed by the discussion of the ground rules for the interview (encouraging children to use 'I don't know' and 'I don't understand', to correct the interviewer if she makes a mistake and asking them to promise to tell the truth). The substantive part of the interview started with an open-ended invitation to 'Tell me everything that happened when you were learning about first aid' and followed by further invitations and open-ended follow-up prompts. No closed-ended or suggestive questions were used by the interviewer. When children indicated they could not recall any new information, they were asked to 'Think back and try to tell me everything that happened from the beginning to the end'. At the end of the interview, children were thanked for their contribution and given a chance to ask questions from the interviewer.

Five weeks after the first aid session, all children ( $N = 79$ ) were interviewed by the same interviewer. Second interviews were conducted in the same manner as the first one. At the end of the interview, children were thanked again and given a small gift (an eraser in the shape of an animal) for their participation.

## Coding

The interviews were transcribed verbatim by the first author. Transcripts ( $N = 117$ ) were coded by the first author according to the completeness, consistency, connectedness and accuracy of children's responses.



## Completeness

The first aid training event was specifically designed to include 35 target details referring to participants, objects and actions. The completeness of children's recall was measured as the number of correct target details they reported. Each detail was counted once, even if the child mentioned it again in the interview. Target details in each training phase are listed in [Table 1](#).

## Consistency

In the repeated condition, the consistency of each correct detail was coded as repeated, reminisced or omitted. Repeated details were correctly mentioned in each interview, reminisced details were missing from the first interview but correctly remembered in the second interview and omitted details were correctly described in the first interview but missing from the second interview. Contradictions were defined as the presence of both a correct and an incorrect version of the same detail in the same interview or across the two interviews.

**Table 1.** List of Target Details by Training Phase.

Phase of first aid training	Target details
<b>Introduction</b>	Partner Teacher
<b>Inspecting the first aid kit</b>	First aid kit Guessing what is inside Looking inside Plasters Put (plasters) on Eye pad Wrap (eye pad) around head Icepack Pop (icepack) Shake (icepack)
<b>What is an emergency?</b>	Emergency or not Cards Sorting (the cards) Example of emergency Example of not emergency
<b>How to help</b>	Video about choking Hit back Five times Check object Repeat Video about fainting Turn them on their side Tilt their head Listen for breathing
<b>Calling for help</b>	Call for help 999 Instruction sheets Ambulance Type of accident Phone number Location Directions Certificate

## Accuracy

Incorrect details were defined as demonstrably untrue details related to the first aid training mentioned by the children. Details referring to events other than the first aid training (e.g. children's family or school) were ignored, as the veracity of these details could not be determined. The accuracy of children's testimony was calculated as the proportion of correct details from all details mentioned. Incorrect details by training phase are listed in Table 2.

## Causal-temporal connectedness

Markers of connectedness were coded using a modified version of the linguistic cohesion coding scheme developed by Kulkofsky et al. (2008).

1. *Simple temporal markers.* Phrases signifying chronological order.
2. *Complex temporal markers.* Phrases placing an event in time with relation to other events.
3. *Markers of causal relations.* Phrases describing cause-and-effect relationships between details.
4. *Markers of optional states.* Phrases referring to conditional events.

The number of linguistic markers present in each interview was counted. Linguistic markers in false starts (*Q: Tell me everything that happened during the first aid training. A: So ... we put some plasters on.*) and fillers (*We watched the videos, then ... then ... then ... I can't remember.*) were not coded, as these do not represent a temporal or causal relationship between two details.

## Inter-rater reliability

The first author coded all of the transcripts, and a random selection of 20% of the transcripts ( $N = 24$ ) were coded for inter-rater reliability by the second author. The second coder was blinded to the hypotheses and conditions of the study. Cohen's  $\kappa$  was used to assess agreement between the main coder and the second coder for the presence or absence each target detail category (e.g. *first aid kit, plasters*) and linguistic marker (e.g. *then, because*).

**Table 2.** Incorrect Details by Training Phase.

Phase of first aid training	Incorrect detail type	Examples
<b>Inspecting the first aid kit</b>	Incorrect object in the first aid kit	<i>Wet wipes, cold compression</i>
	Incorrect action with eye pad	<i>Put bandage on arm</i>
	Incorrect action with icepack	<i>Pierce icepack</i>
<b>How to help</b>	Incorrect emergency in video	<i>Person not breathing</i>
	Incorrect response to choking	<i>Tap shoulder</i>
	Incorrect response to fainting	<i>Shake person</i>
<b>Calling for help</b>	Incorrect emergency number	<i>911, 991</i>
	Incorrect activity	<i>Put cards in order</i>
	Incorrect service	<i>Police</i>

There was perfect agreement ( $\kappa = 1.00$ ) between the coders regarding 42% of correct detail categories ( $N = 15$ ) and 89% of incorrect detail categories ( $N = 8$ ). Reliability was very good for 29% of correct detail types ( $N = 10$ ), with Kappas ranging between .82 and .92. There was substantial reliability for 29% ( $N = 10$ ) of correct detail types and 11% ( $N = 1$ ) of incorrect detail types, with Kappas that ranged between .75 and .78. Reliability was very good for markers of connectedness,  $\kappa = .95$ ,  $SE = .01$ , 95%  $CI$  [.93, .97].

## Analysis plan

Descriptive statistics were calculated to examine the completeness, consistency, connectedness and accuracy of children's accounts in each condition. Preliminary analyses were conducted to assess whether children's age and gender were associated with the outcome measures. Three linear mixed models (LMMs) were conducted to examine whether the fixed effects of interview condition, children's age and gender significantly affected the completeness, accuracy and connectedness of children's recall. In order to address the hypotheses, children's recall across the two interviews of the experimental condition was used as a baseline compared to their recall in the 1-week interview of the experimental condition and the 5-week interview in the delayed control condition. By-subject random effects for child were included to account for individual differences. A negative binomial distribution was used for the connectedness GLMM to correct for the overdispersion. No overdispersion was detected for the other models. Following these analyses, a further LMM examined whether the accuracy of consistent, forgotten and reminisced details differed in the overall recall of the experimental group.

Analyses were performed using the *glmer* function in the R package *lme4* with the *bobyqa* optimiser (Bates et al., 2015) for Poisson outcome measures (completeness, connectedness) and the *lmer* function (using Satterthwaite's method for *t*-tests) in *lme4* for continuous proportion outcomes (accuracy). Model-fitting was computed using the *anova* function in the R stats package (R Core Team, 2013). Pairwise comparisons with Tukey HSD and adjusted means (i.e. estimated marginal means) were computed using the *emmeans* function in the R package *emmeans* (Lenth, 2020). Estimated marginal means are a more accurate way to compute adjusted means that are unbiased by imbalanced datasets by giving equal weight to each cell (Lenth, 2020). All statistical output for models can be found in the Appendices. The results from the best-fit models are reported below alongside the unstandardised fixed effect estimates ( $B$ ), standard errors of the estimates ( $SE$ ), and estimates of significance ( $Z$  and  $p$  values for GLMMs,  $t$  and  $p$  values for LMMs).

GLMMs combine the properties of linear mixed models (which incorporate random effects) and generalised linear models (which handle non-normal data) and are preferable to traditional analysis of variance (ANOVA) models because they have fewer assumptions, handle response variables from different distributions, and maximise power while simultaneously estimating between-subject variance (Bates et al., 2015). A further advantage of GLMMs over ANOVAs for the present study is that they allow for unequal repetitions (some participants interviewed once, others interviewed twice), reducing the number of statistical tests needed to address the hypotheses.

## Results

### Preliminary analyses

There was no difference in the age of children in the experimental ( $M = 113.20$  months,  $SD = 1.70$ ) and control ( $M = 114.91$  months,  $SD = 1.95$ ) groups,  $t(75) = .67$ ,  $p = .50$ ,  $d = .15$ , 95% CI  $[-3.41, 6.85]$ . There was also no difference in the distribution of gender  $\chi^2(1) = .39$ ,  $p = .53$  across conditions. Preliminary LMM and GLMM analyses were conducted to explore whether children's age and gender affected the completeness, connectedness, consistency or accuracy of their accounts (Appendix 1). Children's age did not significantly affect any of the dependent measures, therefore, it was excluded from further analysis. Children's gender was significantly associated with the completeness of children's accounts and thus was included in the completeness GLMM, but was not associated with their accuracy, consistency or connectedness.

### Completeness

On average, children reported 15.63 ( $SD = 5.00$ ) details in the first interview of the experimental condition, 18.54 ( $SD = 5.30$ ) details over the course of the two interviews, and 13.11 ( $SD = 5.16$ ) details in the delayed control condition (Table 3). The best fit model for the completeness of children's recall included interview condition and child's gender without an interaction. Consistent with the first hypothesis, significantly more details were elicited over the course of the two interviews (adjusted  $M = 17.7$ ,  $SE = .92$ ) compared with the first interview 1 week after the event (adjusted  $M = 14.9$ ,  $SE = .81$ ,  $B = -.17$ ,  $Z = -3.19$   $p = .001$ ), and the delayed control interview 5 weeks after the event (adjusted  $M = 12.6$ ,  $SE = .78$ ,  $B = -.34$ ,  $Z = -4.21$   $p < .001$ ). Post hoc analyses (Appendix 2) comparing estimated marginal means showed no significant difference in completeness between the 1-week interview of the experimental group and the 5-week interview of the control group ( $p = .10$ ). Girls (adjusted  $M = 16.6$ ,  $SE = .79$ ) provided more details than boys (adjusted  $M = 13.4$ ,  $SE = .76$ ,  $B = -.21$ ,  $Z = -2.91$ ,  $p = .004$ ).

### Accuracy

On average, children recalled .61 ( $SD = .80$ ) incorrect details in the 1-week interview, 1.07 ( $SD = .74$ ) incorrect details overall in the experimental condition, and 1.14 incorrect details in the delayed control condition (Table 3). Details reported in the first interview were 96% ( $SD = 6\%$ ) accurate on average, compared with 95% ( $SD = 6\%$ ) over the two interviews and 92% ( $SD = 8\%$ ) in the control condition. The best fit model for the accuracy of children's recall included interview condition only.

**Table 3.** The Number of Correct and Incorrect Details Recalled in Each Interview in the Control and Experimental Conditions.

	Single 5 weeks		Repeated 1 1 week		Repeated 2 5 weeks		Repeated Overall	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Number of correct details	13.11	5.16	15.63	5.00	14.49	5.46	18.54	5.30
Number of incorrect details	1.14	1.12	.61	.80	.73	.74	1.07	.98

In contrast to the second hypothesis, there was no significant difference between the accuracy of children's recall in the 1-week interview (adjusted  $M = 96\%$ ,  $SE = .01$ ) and over the course of the two interviews (adjusted  $M = 95\%$ ,  $SE = .01$ ,  $p = .14$ ). Furthermore, a main effect of interview condition showed that children's recall in the single 5-week interview (adjusted  $M = 92\%$ ,  $SE = .01$ ) was less accurate than their overall recall across the two interviews of the experimental condition ( $B = -.03$ ,  $t = -2.20$ ,  $p = .03$ ), although post-hoc analyses (Appendix 3) comparing estimated marginal means did not replicate this finding ( $p = .08$ ). Consistent with the hypothesis, post-hoc analyses comparing estimated marginal means showed that children's recall in the 1-week interview was significantly more accurate than in the single 5-week interview ( $B = -.04$ ,  $t = -2.70$ ,  $p = .03$ ), but did not significantly differ from the accuracy of their overall recall ( $p = .29$ ).

### Connectedness

Combining simple temporal markers, complex temporal markers, markers of causality and markers of optional states, children used 3.23 ( $SD = 3.69$ ) markers of connectedness on average in the 1-week interview of the experimental condition, which increased to 6.48 ( $SD = 7.19$ ) over the course of the two interviews (Table 4). In the 5-week delayed control interview, 2.84 ( $SD = 3.18$ ) markers of connectedness were reported overall. The best fit model for the number of connectedness markers in children's accounts included an interaction between marker type and interview number. Simple temporal markers (adjusted  $M = 5.95$ ,  $SE = .53$ ) were used significantly more frequently than complex temporal markers (adjusted  $M = 1.41$ ,  $SE = .16$ ,  $B = -1.70$ ,  $Z = -10.68$ ,  $p < .001$ ), markers of optional states (adjusted  $M = 4.48$ ,  $SE = .41$ ,  $B = -.35$ ,  $Z = -2.67$ ,  $p = .008$ ), and markers of causality (adjusted  $M = 1.34$ ,  $SE = .15$ ,  $B = -1.40$ ,  $Z = -9.45$ ,  $p < .001$ ). Post-hoc analyses (Appendix 4) further revealed that markers of optional states were significantly more frequent than complex temporal markers ( $p < .001$ ) and markers of causality ( $p < .001$ ). Children reported significantly more markers of connectedness over the course of the two interviews (adjusted  $M = 4.28$ ,  $SE = .45$ ) than in the 1-week interview of the experimental condition (adjusted  $M = 2.16$ ,  $SE = .24$ ,  $B = -.89$ ,  $Z = -4.56$ ,  $p < .001$ ), and in the delayed control condition (adjusted  $M = 2.05$ ,  $SE = .25$ ,  $B = -.71$ ,  $Z = 5.32$ ,  $p < .001$ ).

Results also revealed a significant interaction between interview number and marker type for the delayed control condition and the overall recall of children in the experimental condition ( $B = .68$ ,  $Z = 2.72$ ,  $p = .006$ ). Post-hoc tests revealed that in the two interviews of the experimental condition, complex temporal markers were used more frequently (adjusted  $M = 1.51$ ,  $SE = .27$ ) than causal markers (adjusted  $M = .77$ ,  $SE = .17$ ,

**Table 4.** The Proportion of Linguistic Markers of Connectedness in Children's Responses in the Experimental and Control Conditions.

	Single 5 weeks		Repeated 1 week		Repeated 2 5 weeks		Repeated Overall	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
All markers	1.25	1.13	1.02	.56	1.11	.68	1.06	.55
Simple temporal markers	.55	.41	.49	.27	.56	.42	.53	.29
Complex temporal markers	.22	.49	.09	.11	.09	.10	.09	.09
Optional markers	.41	.50	.32	.26	.34	.34	.33	.28
Causal markers	.07	.10	.11	.15	.12	.16	.11	.13

$p = .04$ ), but there was no significant difference between these marker types in the 5-week delay condition (complex tm: adjusted  $M = 1.85$ ,  $SE = .30$ , causal: adjusted  $M = 2.49$ ,  $SE = .37$ ,  $p = .34$ ). In addition, simple temporal markers were used more frequently than markers of optional states overall in the experimental condition (simple tm: adjusted  $M = 10.13$ ,  $SE = 1.27$ , optional: adjusted  $M = 7.17$ ,  $SE = 5.57$ ,  $p = .04$ ), but not in the delayed control condition (simple tm: adjusted  $M = 4.17$ ,  $SE = .62$ , optional: adjusted  $M = 3.61$ ,  $SE = 2.68$ ,  $p = .81$ ).

### Consistency

On average, 4.05 ( $SD = 2.99$ ) correct details mentioned in the first interview were omitted from the second interview whilst 2.90 ( $SD = 2.36$ ) correct new details were added. Although hypermnesia was not the norm across the sample, 11 children (27%) remembered more details in their second interview than during their first recall. The best fit model for the accuracy of consistent, omitted and reminisced information did not include any interactions. In line with the third hypothesis, consistent details were more accurate (adjusted  $M = 97\%$ ,  $SE = .03$ ) than reminisced details (adjusted  $M = 81\%$ ,  $SE = .03$ ,  $B = -.15$ ,  $Z = -3.61$ ,  $p < .001$ ). There was no significant difference between the accuracy of consistent and forgotten details (adjusted  $M = 90\%$ ,  $SE = .03$ ,  $p = .10$ ). Post-hoc analyses (Appendix 5) replicated the main effects.

### Discussion

Results supported the hypotheses that children would report more details over the course of two interviews than in a single interview either 1 week or 5 weeks after the event, but the 5-week delay would compromise the accuracy of their recall in the control condition. However, contrary to our predictions, information reported overall in the experimental condition was no less accurate than details recalled in the 1-week interview. As expected, reminisced details were less accurate than consistent details in the experimental condition. Exploratory analyses of the effect of interview condition on the connectedness of children's accounts revealed that children used significantly more markers of connectedness across the two interviews of the experimental condition than in a single interview either 1 week or 5 weeks after the event.

### Completeness

Although hypermnesia was only present in approximately one quarter of the experimental group, reminiscence between interviews was the norm, leading to a significantly higher number of correct details reported over the course of two interviews than in a single interview. In forensic investigations, multiple interviews are conducted for a multitude of reasons, including non-disclosure or resistance from the child in the first interview and the emergence of new allegations or new evidence after the first interview (Waterhouse et al., 2016). In these situations, the underlying assumption is that multiple interviews will reveal new details and increase the completeness of children's testimonies. The results of the current experiment, along with findings from previous research (e.g. Hershkowitz & Terner, 2007; Katz & Hershkowitz, 2012; Leander,

2010; Szojka et al., 2020; Waterhouse et al., 2016), suggest that multiple interviews can indeed lead to the recall of new details both in field research and under experimental conditions.

Interestingly, whilst Szojka et al. (2020) found that in forensic transcripts children reported almost as many new details in the second interview as in the first interview, the difference between the number of details reported in the first interview and over the course of two interviews was proportionally smaller in this experiment. This difference cannot be explained by a ceiling effect, as even when correct details in both interviews were considered, participants remembered on average only 18 out of the 35 target details. However, details of a staged event may be less salient than autobiographical memories about stressful or traumatic events (Baker-Ward et al., 1993; Baugerud et al., 2014; Peterson, 2015), which could play a role in the difference between the proportional amount of reminiscence found in field research and experimental studies. Additionally, the present study consisted of a single event which children were asked to recall. Most witnesses in the field study testified about repeated events, therefore, the high proportion of new information surfacing in the second interview might have resulted from shifting focus between different occurrences across interview (Patterson & Pipe, 2009). Finally, analyses of completeness in the current study included only correct details, while field designs do not allow researchers to distinguish between correct and incorrect details. Therefore, some reminisced details in field studies may have contained incorrect information.

## Connectedness

In the present study, the causal-temporal connectedness component of narrative coherence was measured using a modified version of Kulkofsky et al.'s (2008) cohesiveness coding scheme. Children used linguistic markers frequently to locate events in time and place. When all types of markers were considered, including simple temporal markers, complex temporal markers, markers of optional states and markers of causal connections, children used more than one marker, on average, for each detail mentioned. This suggests that in response to open-ended questioning, children between the ages of 8 and 11 years spontaneously attempt to provide a causal-temporal framework for the events they discuss.

When describing the first aid event, children in the present study used markers of simple temporal relations and optional states significantly more frequently than markers of complex temporal relations and causal relationships. Interestingly, in Szojka et al.'s (2020) field study, markers of causal relationships were used most frequently, followed by markers of simple and complex temporal relations and optional states. This might reflect differences in the age range of the sample (8–11 in the present study and 3–14 in the field study), as Habermas and de Silveira (2008) suggest that temporal dimensions of narrative coherence improve substantially between the ages of 8 and 12, whereas causal coherence remains in development until late adolescence. Results also revealed that children used more linguistic markers of causal-temporal connections across the two interviews of the experimental condition than in a single interview either 1 week or 5 weeks after the event, in line with an increase in the number of details across interviews.

## Accuracy

As the accuracy of information cannot be ascertained in field research, the task to establish the impact of multiple interviews on the proportion of correct details in children's testimony falls on experimental studies. Consistent with previous research using a delayed control group (Brown et al., 2015; Quas et al., 2007), in the present study, interviewing children a second time did not decrease the accuracy of their accounts. It is difficult to draw strong conclusions about the impact of multiple interviews on the accuracy of children's recall based on the small number of studies using a delayed control group, but these findings demonstrate that providing children with a second recall opportunity does not inherently increase the proportion of incorrect details in their accounts.

However, children's recall in the 5-week single interview was less accurate on average than the 1-week interview, demonstrating the damaging effect of delay on the accuracy of children's accounts. This suggests that although re-interviewing children may not always have a negative effect on the accuracy of children's testimony, interviewers need to weigh the costs and benefits of conducting further interviews very carefully as the delay between the event children are asked to recall and the interviews increases. While there was no significant difference in the accuracy of the 1-week interview and children's overall recall in the experimental condition, descriptively, there was a small increase in the overall number of incorrect details, partially due to children's tendency to make different errors in the first and second interview. However, in real terms, the decrease in accuracy from the first interview to the cumulative information reported over the two interviews was very small, with less than one additional incorrect detail reported on average.

The cost of a small increase in incorrect details in forensic contexts can differ widely according to the aim of the investigation and the forensic relevance of that specific detail. In some cases, incorrect details with high forensic relevance can mislead investigators or result in false convictions. Even incorrect details with low forensic relevance can compromise children's credibility during cross-examination, casting doubt on more central aspects of their testimony in the eyes of the jury (Burrows & Powell, 2014). However, in other cases, a small increase in the number of incorrect details reported is a relatively small price to pay for gaining valuable new details. One such situation is reported in the case study of a 9-year-old child, who was interviewed in relation to her sister's kidnapping (Orbach et al., 2012). Although subsequent interviews with the victim and the perpetrator have shown that the witness reported some incorrect information related to both central and peripheral details over the interviews, the child provided the key piece of information about the identity of the perpetrator that led the investigators to her sister in the sixth interview.

## Consistency

In contrast to findings from field studies (Baugerud et al., 2014; Szojka et al., 2020), in the present study, most details recalled by children were consistent across the two interviews, with only three details reminisced in the second interview, and four details mentioned in the initial interview but omitted from the second one, on average. The consistency of individual details across the two interviews was associated with their accuracy. Details that



were consistent across the two interviews were 97% accurate, whilst details reminisced in the second interview were only 81% correct. The difference in accuracy between consistent and reminisced details found in the present study supports the results of previous research regarding the relationship between consistency and accuracy in children's recall (La Rooy et al., 2005; Pipe et al., 1999; Salmon & Pipe, 2000). In the present study, the two interviews were separated by a monthlong delay, which might have exacerbated the difference in accuracy between consistent and reminisced details (La Rooy et al., 2005). Interestingly, information mentioned in the first interview but omitted in the second interview was also less accurate, on average, than consistent information. Although these details were still 90% accurate, this difference cannot be explained by the delay between the two interviews and might reflect instead an increased survival rate for correct memories compared to incorrect memories in children's free recall (London et al., 2009).

The difference in accuracy between consistent, omitted and reminisced details suggests that an analysis of the consistency of details across multiple interviews may provide an indication of their accuracy under some circumstances. However, the overall consistency of accounts elicited over the course of the two interviews did not correlate with their overall accuracy in past research (Baugerud et al., 2014; Gilbert & Fisher, 2006). Additionally, laboratory studies differ from real-life forensic interviews in some respects, which may influence the relationship between consistency and accuracy. Specifically, the average amount of reminisced information was much higher in field research than in the current study, potentially due to the increased number and complexity of the events children are asked to describe, and the more directive questioning methods used by interviewers. Therefore, whilst in the current study newly recalled information in the second interview presumably consisted of details children did not remember the first time, in field studies, a large proportion of new information in the second interview related to events or aspects of events that were not at all discussed in the first interview (Patterson & Pipe, 2009; Szojka et al., 2020).

### Children's age and gender

In Szojka et al.'s (2020) field study, the completeness and connectedness of children's recall was positively correlated with age. In contrast to previous research demonstrating substantial gains in narrative coherence during middle childhood (Habermas & de Silveira, 2008; Reese et al., 2011), developmental differences in the narrative coherence of children's testimony were not replicated in the present study. However, preliminary analyses revealed a significant effect of gender on the completeness of children's testimony, with girls reporting more details than boys on average.

Although gender differences in the completeness of recall are consistent with previous research comparing girls' and boys' narratives (Haden et al., 1997; Lamb & Garretson, 2003), there is an important caveat to interpreting this result. Lamb and Garretson (2003) found that the difference between the recall of girls and boys was more pronounced when they were questioned by female interviewers, as girls provided more detailed responses to directive questions asked by female interviewers, whilst boys provided the same amount of details when responding to male and female interviewers. All children in the current study were questioned by a female interviewer, therefore, the

difference between the number of details reported by girls and boys may result from an interaction between the interviewer's and participant's gender.

### Limitations and further research

This study provided a novel insight into the relationship between narrative coherence and accuracy in multiple interviews, however, limitations of the research design and differences between laboratory studies and real-life investigative interviews limited the extent to which findings can be generalised to forensic contexts. Firstly, in the present study, children were asked to recall details of a first aid training, an event that is clearly not characterised by the secrecy and emotional trauma associated with physical and sexual abuse. Therefore, results of the current study should be interpreted as complementary to findings from field studies, which have the benefit of high external validity but do not allow for analyses involving accuracy. Secondly, whilst in field contexts the proportion of invitations often remains low despite guideline recommendations, the interviewer in the current study relied solely on open-ended questions. Results thus show the relationship between narrative coherence and accuracy in ideal, rather than typical forensic interviews. All of the children in the sample were interviewed by the same researcher, who was not blind to the hypotheses of the study. While the interview guide was followed closely in all of the interviews, the researcher may have nonetheless provided unintentional clues affecting the recall of participants in the experimental and control conditions. Thirdly, the sample size for the present study was limited due to restrictions on access to schools and childcare centres. Power analyses conducted using the *powerSim* function in the R package *simr* showed adequate power (>90%) for the LMMs examining completeness, connectedness and consistency, however, the LMM addressing the second hypothesis was underpowered (74%), suggesting a need for caution when interpreting the effect of interview condition on the accuracy of children's recall.

Fourthly, the study design included a 1-week delay between the event to be remembered and the first interview and a 1-month delay between the two interviews. Previous research has found that the length of delay affects the accuracy and narrative coherence of children's accounts in multiple interviews (Kulkofsky et al., 2008; La Rooy et al., 2005), thus, findings from the current study may not generalise to forensic contexts where delays are much more variable, ranging from less than an hour to several years between interviews (Katz & Hershkowitz, 2012; Szojka et al., 2020; Waterhouse et al., 2016). Finally, the current study failed to find developmental differences in the completeness, accuracy, consistency and connectedness of recall. This suggests that when the difference between age groups is narrow, individual differences may outweigh the impact of age on the narrative coherence of recall. Further research involving a wider range of age groups is needed to explore how the relationship between narrative coherence, accuracy and multiple interviews changes across developmental stages.

### Conclusion

The present study was the first to reveal the impact of multiple interviews on the completeness, consistency, connectedness and accuracy of children's recall. Findings regarding the effects of multiple interviews on the components of narrative coherence were

consistent with field research, as the completeness of children's accounts increased over the course of the two interviews. The increase in the number of details was accompanied by an increase in the use of linguistic markers of causal-temporal connectedness. In the experimental condition, individual details that were consistent across the two interviews were more accurate than reminisced details in the second interview. Although the overall recall of children in the experimental condition was no less accurate than their recall in the first interview, the accuracy of details in the delayed control condition was significantly lower than in the 1-week interview, suggesting that interviewers need to be mindful of the impact of delay when making decisions about the timing of multiple interviews.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Zsafia A. Szojka  <http://orcid.org/0000-0002-4126-1261>

## References

- Baker-Ward, L., Gordon, B. N., Ornstein, P. A., Larus, D. M., & Clubb, P. A. (1993). Young children's long-term retention of a pediatric examination. *Child Development*, 64(5), 1519–1533.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Baugerud, G. A., Magnussen, S., & Melinder, A. (2014). High accuracy but low consistency in children's long-term recall of a real-life stressful event. *Journal of Experimental Child Psychology*, 126, 357–368. <https://doi.org/10.1016/j.jecp.2014.05.009>
- Brewer, N., Potter, R., Fisher, R. P., Bond, N., & Luszcz, M. A. (1999). Beliefs and data on the relationship between consistency and accuracy of eyewitness testimony. *Applied Cognitive Psychology*, 13(4), 297–313. [https://doi.org/10.1002/\(sici\)1099-0720\(199908\)13:4<297::aid-acp578>3.3.co;2-j](https://doi.org/10.1002/(sici)1099-0720(199908)13:4<297::aid-acp578>3.3.co;2-j)
- Brown, D. A., Brown, E.-J., Lewis, C. N., & Lamb, M. E. (2018). Narrative skill and testimonial accuracy in typically developing children and those with intellectual disabilities. *Applied Cognitive Psychology*, 32(5), 550–560. <https://doi.org/10.1002/acp.3427>
- Brown, D. A., Lewis, C. N., & Lamb, M. E. (2015). Preserving the past: An early interview improves delayed event memory in children with intellectual disabilities. *Child Development*, 86(4), 1031–1047. <https://doi.org/10.1111/cdev.12364>
- Brubacher, S. P., Peterson, C., La Rooy, D., Dickinson, J. J., & Poole, D. A. (2019). How children talk about events: Implications for eliciting and analyzing eyewitness reports. *Developmental Review*, 51, 70–89. <https://doi.org/10.1016/j.dr.2018.12.003>
- Burrows, K. S., & Powell, M. (2014). Prosecutors' recommendations for improving child witness statements about sexual abuse. *Policing and Society*, 24(2), 189–207. <https://doi.org/10.1080/10439463.2013.784305>
- Cassel, W. S., & Bjorklund, D. F. (1995). Developmental patterns of eyewitness memory and suggestibility. *Law and Human Behavior*, 19(5), 507–532. <https://doi.org/10.1007/BF01499341>
- Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, 113(3), 403–439. <https://doi.org/10.1037/0033-2909.113.3.403>
- Chae, Y., Kulkofsky, S., Debaran, F., Wang, Q., & Hart, S. L. (2016). Low-SES preschool children's eyewitness memory: The role of narrative skill. *Behavioral Sciences & the Law*, 34(1), 55–73. <https://doi.org/10.1002/bsl.2242>
- Connolly, D. A., Price, H. L., Lavoie, J. A. A., & Gordon, H. M. (2008). Perceptions and predictors of children's credibility of a unique event and an instance of a repeated event. *Law and Human Behavior*, 32(1), 92–112. <https://doi.org/10.1007/s10979-006-9083-3>

- Dietze, P. M., Sharman, S. J., Powell, M. B., & Thomson, D. M. (2013). Does free recall moderate the effect of mental context reinstatement instructions on children's cued recall? *Psychology, Crime & Law*, 19(10), 881–891. <https://doi.org/10.1080/1068316X.2012.700312>
- Ghetti, S., Qin, J., & Goodman, G. S. (2002). False memories in children and adults: Age, distinctiveness, and subjective experience. *Developmental Psychology*, 38(5), 705–718. <https://doi.org/10.1037/0012-1649.38.5.705>
- Gilbert, J. A. E., & Fisher, R. P. (2006). The effects of varied retrieval cues on reminiscence in eyewitness memory. *Applied Cognitive Psychology*, 20(6), 723–739. <https://doi.org/10.1002/acp.1232>
- Goodman, G. S., Bottoms, B. L., Schwartz-Kenney, B. M., & Rudy, L. (1991). Children's testimony about a stressful event: Improving children's reports. *Journal of Narrative & Life History*, 1(1), 69–99. <https://doi.org/10.1075/jnlh.1.1.05chi>
- Goodman, G. S., & Quas, J. A. (2008). Repeated interviews and children's memory: It's more than just how many. *Current Directions in Psychological Science*, 17(6), 386–390. <https://doi.org/10.1111/j.1467-8721.2008.00611.x>
- Habermas, T., & de Silveira, C. (2008). The development of global coherence in life narratives across adolescence: Temporal, causal, and thematic aspects. *Developmental Psychology*, 44(3), 707–721. <https://doi.org/10.1037/0012-1649.44.3.707>
- Haden, C. A., Haine, R. A., & Fivush, R. (1997). Developing narrative structure in parent-child reminiscing across the preschool years. *Developmental Psychology*, 33(2), 295–307. <https://doi.org/10.1037/0012-1649.33.2.295>
- Hershkowitz, I., & Terner, A. (2007). The effects of repeated interviewing on children's forensic statements of sexual abuse. *Applied Cognitive Psychology*, 21(9), 1131–1143. <https://doi.org/10.1002/acp.1319>
- Hubbard, K., Saykaly, C., Lee, K., Lindsay, R. C. L., Bala, N., & Talwar, V. (2016). Children's recall accuracy for repeated events over multiple interviews: Comparing information types. *Psychiatry, Psychology and Law*, 1–14. <https://doi.org/10.1080/13218719.2016.1256015>
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3–28. <https://doi.org/10.1037/0033-2909.114.1.3>
- Katz, C., & Hershkowitz, I. (2012). Repeated interviews with children who are the alleged victims of sexual abuse. *Research on Social Work Practice*, 1049731512467511. <https://doi.org/10.1177/1049731512467511>
- Knutsson, J., Allwood, C. M., & Johansson, M. (2011). Child and adult witnesses: The effect of repetition and invitation-probes on free recall and metamemory realism. *Metacognition and Learning*, 6(3), 213–228. <https://doi.org/10.1007/s11409-011-9071-y>
- Kulkofsky, S., Wang, Q., & Ceci, S. J. (2008). Do better stories make better memories? Narrative quality and memory accuracy in preschool children. *Applied Cognitive Psychology*, 22(1), 21–38. <https://doi.org/10.1002/acp.1326>
- Kyriakidou, M., Zalaf, A., & Blades, M. (2014). Prosecution reliance on the quantity rather than the quality of evidence for children's testimonies. *The Cyprus Journal of Sciences*, 12, 3–16.
- Lamb, M. E., & Garretson, M. E. (2003). The effects of interviewer gender and child gender on the informativeness of alleged child sexual abuse victims in forensic interviews. *Law and Human Behavior*, 27(2), 157–171. <https://doi.org/10.1023/A:1022595129689>
- Lamb, M. E., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2008). *Tell me what happened: Structured investigative interviews of child victims and witnesses*. John Wiley & Sons Inc. <https://doi.org/10.1002/9780470773291>
- Lamb, M. E., Orbach, Y., Hershkowitz, I., Esplin, P. W., & Horowitz, D. (2007a). Structured forensic interview protocols improve the quality and informativeness of investigative interviews with children: A review of research using the NICHD investigative interview protocol. *Child Abuse & Neglect*, 31(11–12), 1201–1231. <https://doi.org/10.1016/j.chiabu.2007.03.021>
- Lamb, M. E., Orbach, Y., Hershkowitz, I., Horowitz, D., & Abbott, C. B. (2007b). Does the type of prompt affect the accuracy of information provided by alleged victims of abuse in forensic interviews? *Applied Cognitive Psychology*, 21(9), 1117–1130. <https://doi.org/10.1002/acp.1318>

- La Rooy, D., Lamb, M. E., & Pipe, M.-E. (2009). Repeated interviewing: A critical evaluation of the risks and potential benefits. In K. Kuehnlé & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 327–361). John Wiley & Sons Inc.
- La Rooy, D. L., Brubacher, S. P., Aromäki-Stratos, A., Cyr, M., Hershkowitz, I., Korkman, J., Myklebust, T., Naka, M., Peixoto, C. E., Roberts, K. P., Stewart, H., & Lamb, M. E. (2015). The NICHD protocol: A review of an internationally-used evidence-based tool for training child forensic interviewers. *Journal of Criminological Research, Policy and Practice*, <https://doi.org/10.1108/JCRPP-01-2015-0001>
- La Rooy, D. L., Pipe, M.-E., & Murray, J. E. (2005). Reminiscence and hypermnesia in children's eye-witness memory. *Journal of Experimental Child Psychology*, *90*(3), 235–254. <https://doi.org/10.1016/j.jecp.2004.11.002>
- Leander, L. (2010). Police interviews with child sexual abuse victims: Patterns of reporting, avoidance and denial. *Child Abuse & Neglect*, *34*(3), 192–205. <https://doi.org/10.1016/j.chiabu.2009.09.011>
- Leichtman, M. D., & Ceci, S. J. (1995). The effects of stereotypes and suggestions on preschoolers' reports. *Developmental Psychology*, *31*(4), 568–578. <https://doi.org/10.1037/0012-1649.31.4.568>
- Lenth, R. (2020). Emmeans: Estimated marginal means, aka least-squares means. R package version 1.5.0. <https://cran.r-project.org/package=emmeans>
- London, K., Bruck, M., & Melnyk, L. (2009). Post-event information affects children's autobiographical memory after one year. *Law and Human Behavior*, *33*(4), 344–355. <https://doi.org/10.1007/s10979-008-9147-7>
- Melinder, A., Alexander, K., Cho, Y. I., Goodman, G. S., Thoresen, C., Lonnum, K., & Magnussen, S. (2010). Children's eyewitness memory: A comparison of two interviewing strategies as realized by forensic professionals. *Journal of Experimental Child Psychology*, *105*(3), 156–177. <https://doi.org/10.1016/j.jecp.2009.04.004>
- Ministry of Justice. (2011). *Achieving best evidence in criminal proceedings: Guidance on interviewing victims and witnesses, and guidance on using special measures*. <http://www.justice.gov.uk/downloads/victims-and-witnesses/vulnerablewitnesses/achieving-bestevidence-criminal-proceedings.pdf>
- Mugno, A. P., Klemfuss, J. Z., & Lyon, T. D. (2016). Attorney questions predict jury-eligible adult assessments of attorneys, child witnesses, and defendant guilt: Credibility of attorneys and child witnesses. *Behavioral Sciences & the Law*, *34*(1), 178–199. <https://doi.org/10.1002/bsl.2214>
- Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: A social cultural developmental theory. *Psychological Review*, *111*(2), 486–511. <https://doi.org/10.1037/0033-295X.111.2.486>
- Olaguez, A. P., & Klemfuss, J. Z. (2020). Differential effects of direct and cross examination on mock jurors' perceptions and memory in cases of child sexual abuse. *Psychiatry, Psychology and Law*, *27*(5), 778–796. <https://doi.org/10.1080/13218719.2020.1742239>
- Orbach, Y., Lamb, M. E., Rooy, D. L., & Pipe, M.-E. (2012). A case study of witness consistency and memory recovery across multiple investigative interviews. *Applied Cognitive Psychology*, *26*(1), 118–129. <https://doi.org/10.1002/acp.1803>
- Ornstein, P. A., Gordon, B. N., & Larus, D. M. (1992). Children's memory for a personally experienced event: Implications for testimony. *Applied Cognitive Psychology*, *6*(1), 49–60. <https://doi.org/10.1002/acp.2350060103>
- Patterson, T., & Pipe, M.-E. (2009). Exploratory assessments of child abuse: Children's responses to interviewer's questions across multiple interview sessions. *Child Abuse & Neglect*, *33*(8), 490–504. <https://doi.org/10.1016/j.chiabu.2008.12.012>
- Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology*, *62*(2), 189–206. <https://doi.org/10.1037/0022-3514.62.2.189>
- Peterson, C. (2015). A decade later: Adolescents' memory for medical emergencies. *Applied Cognitive Psychology*, *29*(6), 826–834. <https://doi.org/10.1002/acp.3192>
- Peterson, C., Morris, G., Baker-Ward, L., & Flynn, S. (2014). Predicting which childhood memories persist: Contributions of memory characteristics. *Developmental Psychology*, *50*(2), 439–448. <https://doi.org/10.1037/a0033221>
- Pipe, M.-E., Gee, S., Wilson, J. C., & Egerton, J. M. (1999). Children's recall 1 or 2 years after an event. *Developmental Psychology*, *35*(3), 781–789. <https://doi.org/10.1037/0012-1649.35.3.781>

- Plotnikoff, J., & Woolfson, R. (2001). *An evaluation of child witness support*. The Scottish Executive Central Research Unit. <http://www.scotland.gov.uk/Publications/2001/>
- Price, H. L., Connolly, D. A., & Gordon, H. M. (2016). Children who experienced a repeated event only appear less accurate in a second interview than those who experienced a unique event. *Law and Human Behavior, 40*(4), 362–373. <https://doi.org/10.1037/lhb0000194>
- Quas, J. A., Malloy, L. C., Melinder, A., Goodman, G. S., D’Mello, M., & Schaaf, J. (2007). Developmental differences in the effects of repeated interviews and interviewer bias on young children’s event memory and false reports. *Developmental Psychology, 43*(4), 823–837. <https://doi.org/10.1037/0012-1649.43.4.823>
- R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Reese, E., Haden, C. A., Baker-Ward, L., Bauer, P., Fivush, R., & Ornstein, P. A. (2011). Coherence of personal narratives across the lifespan: A multidimensional model and coding method. *Journal of Cognition and Development, 12*(4), 424–462. <https://doi.org/10.1080/15248372.2011.587854>
- Salmon, K., & Pipe, M.-E. (2000). Recalling an event one year later: The impact of props, drawing and a prior interview. *Applied Cognitive Psychology, 14*(2), 99–120. [https://doi.org/10.1002/\(SICI\)1099-0720\(200003/04\)14:2<99::AID-ACP639>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1099-0720(200003/04)14:2<99::AID-ACP639>3.0.CO;2-5)
- Salmon, K., Pipe, M.-E., Malloy, A., & Mackay, K. (2012). Do non-verbal aids increase the effectiveness of ‘best practice’ verbal interview techniques? *Applied Cognitive Psychology, 26*(3), 370–380. <https://doi.org/10.1002/acp.1835>
- Saywitz, K. J., Goodman, G. S., Nicholas, E., & Moan, S. F. (1991). Children’s memories of a physical examination involving genital touch: Implications for reports of child sexual abuse. *Journal of Consulting and Clinical Psychology, 59*(5), 682–691. <https://doi.org/10.1037//0022-006x.59.5.682>
- Stromwall, L. A., & Granhag, P. A. (2005). Children’s repeated lies and truths: Effects on adults’ judgments and reality monitoring scores. *Psychiatry, Psychology and Law, 12*(2), 345–356. <https://doi.org/10.1375/pplt.12.2.345>
- Szojka, Z. A., Andrews, S. J., Lamb, M. E., Stolzenberg, S. N., & Lyon, T. D. (2017). Challenging the credibility of alleged victims of child sexual abuse in Scottish courts. *Psychology, Public Policy, and Law, 23*(2), 200–210. <https://doi.org/10.1037/law0000117>
- Szojka, Z. A., Nicol, A., & La Rooy, D. (2020). Narrative coherence in multiple forensic interviews with child witnesses alleging physical and sexual abuse. *Applied Cognitive Psychology, 34*(5), 943–960. <https://doi.org/10.1002/acp.3673>
- Voss, J. F., Wiley, J., & Sandak, R. (1999). On the use of narrative as argument. In *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 235–252). Lawrence Erlbaum Associates Publishers.
- Wang, Q., Bui, V.-K., & Song, Q. (2015). Narrative organisation at encoding facilitated children’s long-term episodic memory. *Memory (Hove, England), 23*(4), 602–611. <https://doi.org/10.1080/09658211.2014.914229>
- Waterhouse, G. F., Ridley, A. M., Bull, R., La Rooy, D., & Wilcock, R. (2016). Dynamics of repeated interviews with children: Repeated interviews with children. *Applied Cognitive Psychology, 30*(5), 713–721. <https://doi.org/10.1002/acp.3246>

## Appendices

**Appendix 1. Preliminary Analyses of the Effects of Children's Age and Gender.**

Dependent measure	Fixed effect	<i>B</i>	<i>SE</i>		<i>p</i>
Completeness	Gender: Female	.22	.08	<i>Z</i> = 2.94	.003
	Age	-.01	.04	<i>Z</i> = -.28	.78
Accuracy	Gender: Female	.03	.02	<i>t</i> = 1.68	.10
	Age	-.01	.01	<i>t</i> = -.86	.40
Connectedness	Gender: Female	.14	.16	<i>Z</i> = .91	.36
	Age	-.11	.08	<i>Z</i> = -1.32	.19
Consistency	Gender: Female	.07	.04	<i>t</i> = 1.81	.07
	Age	-.01	.02	<i>t</i> = -.59	.56

**Appendix 2. Results of GLMM Analysis and Post-hoc Tests Exploring Effects of Interview Condition and Gender on Completeness.**

	Fixed effect	<i>B</i>	<i>SE</i>	<i>Z</i>	<i>p</i>
GLMM results	Interview: 1-week (Exp.)	-.17	.05	-3.19	= .001
	Interview: 5-week (Con.)	-.34	.08	-4.21	< .001
	Gender: Female	.21	.07	2.91	= .004
Contrasts		Ratio	<i>SE</i>	<i>z</i> ratio	<i>p</i>
	Overall (Exp.) / 1-week (Exp.)	1.19	.06	3.19	= .004
	Overall (Exp.) / 5-week (Con.)	1.40	.11	4.21	< .001
	1-week (Exp.) / 5-week (Con.)	1.18	.10	2.03	= .10

**Appendix 3. Results of GLMM Analysis and Post-hoc Tests Exploring Effects of Interview Condition on Accuracy.**

	Fixed effect	<i>B</i>	<i>SE</i>	<i>df</i>	<i>T</i>	<i>p</i>
GLMM results	Interview: 1-week (Exp.)	.01	.01	38.64	1.52	= .14
	Interview: 5-week (Con.)	-.03	.02	77.72	-2.20	= .03
Contrasts	Overall (Exp.) / 1-week (Exp.)	-.01	.01	40.1	-1.52	= .29
	Overall (Exp.) / 5-week (Con.)	.03	.02	79.1	2.20	= .08
	1-week (Exp.) / 5-week (Con.)	.04	.02	79.1	2.70	= .03

**Appendix 4. Results of GLMM Analysis and Post-hoc Tests Exploring Effects of Interview Condition and Marker Type on Connectedness.**

Fixed effect	B	SE	Z	p
<b>GLMM results</b>				
Marker: Complex	-1.70	.16	-10.68	< .001
Marker: Optional	-.35	.13	-2.67	= .008
Marker: Causal	-1.40	.15	9.47	< .001
Interview: 1-week (Exp.)	-.71	.13	-5.32	< .001
Interview: 5-week (Control)	-.89	.19	-4.56	<.001
Marker: Complex X Interview: 1-week (Exp.)	.09	.25	.37	= .71
Marker: Optional X Interview: 1-week (Exp.)	-.01	.20	-.07	= .94
Marker: Causal X Interview: 1-week (Exp.)	.03	.23	.11	= .91
Marker: Complex X Interview: 5-week (Con.)	.68	.25	2.72	= .006
Marker: Optional X Interview: 5-week (Con.)	.20	.21	.96	= .34
Marker: Causal X Interview: 5-week (Con.)	-.28	.27	-1.04	= .30
<b>Marker type contrasts</b>				
	Ratio	SE	z ratio	p
Simple / Complex	4.22	.44	13.67	< .001
Simple / Optional	1.33	.11	3.30	= .005
Simple / Causal	4.43	.48	13.66	< .001
Complex / Optional	.32	.03	-10.80	< .001
Complex / Causal	1.05	.13	.39	= .98
Optional / Causal	3.34	.37	10.92	< .001
<b>Interview condition contrasts</b>				
Overall (Exp.) / 1-week (Exp.)	1.98	.17	7.97	< .001
Overall (Exp.) / 5-week (Control)	2.09	.34	4.60	< .001
1-week (Exp.) / 5-week (Control)	1.06	.17	.33	= .94
<b>Marker type contrasts in Overall (Exp.)</b>				
Simple / Complex	5.46	.87	10.68	< .001
Simple / Optional	1.41	.18	2.67	= .04
Simple / Causal	4.07	.60	9.47	< .001
Complex / Optional	.26	.04	-8.34	< .001
Complex / Causal	.75	.13	-1.67	= .34
Optional / Causal	2.88	.43	7.00	< .001
<b>Marker type contrasts in 1-week (Exp.)</b>				
Simple / Complex	4.98	.96	8.37	< .001
Simple / Optional	1.43	.21	2.42	= .07
Simple / Causal	3.96	.71	7.69	< .001
Complex / Optional	.29	.06	-6.34	< .001
Complex / Causal	.80	.18	-1.04	= .73
Optional / Causal	2.77	.51	5.53	< .001
<b>Marker type contrasts in 5-week (Con.)</b>				
Simple / Complex	2.77	.53	5.27	< .001
Simple / Optional	1.16	.19	.88	= .81
Simple / Causal	5.40	1.23	7.39	< .001
Complex / Optional	.42	.08	-4.45	< .001
Complex / Causal	1.95	.49	2.68	= .04
Optional / Causal	4.67	1.07	6.73	< .001



**Appendix 5. Results of GLMM Analysis and Post-hoc Tests Exploring Effects of Consistency on Accuracy.**

	Fixed effect	<i>B</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>
GLMM results						
	Consistency: Forgotten	.07	.04	115	-1.67	= .10
	Consistency: Reminisced	-.15	.04	115	-3.61	<.001
Contrasts						
	Consistent / Forgotten	.07	.04	76.8	1.67	= .22
	Consistent / Reminisced	.15	.04	77.5	3.60	= .001
	Forgotten / Reminisced	.08	.04	78.9	1.92	= .14