The Impact of Multiple Interviews on the Narrative Coherence, Accuracy and Credibility of Children's Testimonies

Zsofia Szojka

Royal Holloway, University of London

June 2020

This thesis is submitted for the degree of Doctor of Philosophy.

Declaration of Authorship:

I, Zsofia Szojka, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signature:

Date:

Acknowledgements

First, I would like to thank my supervisors, Dr David La Rooy and Dr Rosie Meek for their continuous support and valuable advice throughout the PhD process. Thank you, David, for involving me in exciting projects and introducing me to inspiring researchers at moments when I was questioning my place in the academic community. I would also like to express my gratitude to the School of Law for providing the funding that allowed me to undertake this project.

Completing the three studies in this thesis would not have been possible without the help of the teachers and pupils who volunteered their time to take part in the experiment and the friends and strangers who responded to my survey. Non Frenguelli and Annabelle Nicol deserve special recognition for their generosity in helping me collect and analyse vast amounts of data in their free time.

Finally, I would like to thank my friends and family, who kept me sane and gave me the motivation to keep going. The advice of the ‘testing team’ has been essential for correcting glitches in the pilot version of my survey. Thank you, Ádám, for the brainstorming sessions that helped me see my research in a different light. Thank you, Ancsi and Peti for providing me with much-needed distraction at the Castle. Mum, thank you for inspiring me to keep learning and occasionally reminding me not to take myself too seriously. Thank you, Khalid, for everything.

Abstract

The lack of forensically meaningful measures of the quality of children's testimony that can be used in both field and laboratory studies prevents reaching strong conclusions about the impact of multiple interviews. Drawing on theory and research from cognitive science and forensic psychology, a new model of narrative coherence, consisting of the completeness, consistency and connectedness of children's recall, was developed to assess the quality of testimonies elicited in multiple interviews.

The thesis describes three studies evaluating the impact of multiple forensic interviews on the narrative coherence, accuracy and credibility children's testimony. Study 1 revealed that multiple interviews impact the narrative coherence of forensic testimonies provided by child witnesses positively, through increasing their completeness without decreasing their consistency or connectedness. Children's age was associated with the completeness and connectedness of their recall. To overcome the limitations of field research, children's recall of a staged experiment was analysed in Study 2. Children interviewed twice provided equally accurate descriptions as participants interviewed a single time in a control group matched for delay, however, the 1-month delay between interviews resulted in a slight but significant decrease in accuracy. Study 3 aimed to determine whether the benefits of multiple interviews found in Study 1 and Study 2 affect witness credibility. Mock jurors rated the witness in a fictional child abuse case as more reliable and competent and were more likely to convict the defendant when presented with two interviews rather than one. Credibility ratings and verdict decisions were also associated with jurors' gender and the consistency of the witness testimony.

In conclusion, results support the view that multiple interviews provide an effective method of increasing the completeness and credibility of children's testimony. The accuracy

of children's testimony was found to be compromised by delay, rather than multiple interviews.

Table of Contents

Chapter 1. Introduction 10
1.1 Key Memory Concepts in Research on Multiple Interviews with Children 14
1.2 Fundamental Questions in the Study of Multiple Interviews with Child Witnesses 16
1.3 Developing an Integrated Measure of the Quality of Children's Testimony 42
1.4 An Overview of the Present Studies 55

Chapter 2. Narrative Coherence in Multiple Forensic Interviews with Child Witnesses Alleging Physical and Sexual Abuse 60
2.1 Introduction 60
2.2 Method 69
2.3 Results 74
2.4 Discussion 94

Chapter 3. The Impact of Multiple Interviews on the Accuracy and Narrative Coherence of Children's Descriptions of a Staged Event 103
3.1 Introduction 103
3.2 Method 110
3.3 Results 119
3.4 Discussion ..... **Error! Bookmark not defined.**

Chapter 4. The Impact of Narrative Coherence and Multiple Interviews on Mock Jurors' Ratings of the Credibility of Child Witnesses 141
4.1 Introduction 151
4.2 Method 168
4. 3 Results 171
4.4 Discussion 185

Chapter 5. Implications and Conclusion 195
5.1 The Impact of Multiple Interviews on the Narrative Coherence, Accuracy and Credibility of Children's Testimony 195
5.2 Assessing the Present Model of Narrative Coherence 205
5.3 Strengths and Limitations of the Methodology 212
5.4 Future Research 213
5.5 Evaluating Guidelines for Multiple Interviews 216
5.6 Conclusion 232

References 235

Appendices 258
Appendix A 258
Appendix B 260
Appendix C 263
Appendix D 266
Appendix E 273
Appendix F 277
Appendix G 280
Appendix H 282
Appendix I 302

List of Tables

Table 1 Results of Studies Published After 2008 Relying on Open-ended Questioning methods 31
Table 2 Results of Studies Published After 2008 Relying on Suggestive Questioning Methods 39
Table 3 Examples of the Content Codes Assigned to Children's Responses 72
Table 4 Examples of Linguistic Markers of Connectedness 74
Table 5 The Number of Details Reported in Each Content Category in Interview 1 and Interview 2 77
Table 6 Summary of Analyses of the Content of Children's Responses in Interview 1 and Overall 81
Table 7 Summary of Analyses of the Subjective Content of Children's Responses in Interview 1 and Overall 85
Table 8 Percentage of Responses Including Linguistic Markers of Connectedness in Interview 1 and Interview 2 88
Table 9 Summary of Analyses of the Percentage of Linguistic Markers in Children's Responses in Interview 1 and Overall 90
Table 10 Proportion of Linguistically Connected Responses Elicited by Each Question Type in Interview 1 and Interview 2 92
Table 11 Summary of Analyses of the Proportion of Linguistic Markers in Elicited by Question Type in Interview 1 and Interview 2 93
Table 12 List of Target Details by Training Phase 115
Table 13 Incorrect Details by Training Phase 117
Table 14 The Number of Correct and Incorrect Details Recalled in Each Interview in the Control and Experimental Conditions ..... **Error! Bookmark not defined.**
Table 15 The Proportion of Linguistic Markers of Connectedness in Children's Responses in Each Interview ..... **Error! Bookmark not defined.**
Table 16 Number of Key Events Placed in Correct Temporal Order in Each Interview .**Error! Bookmark not defined.**
Table 17 Analyses of the Completeness of Children’s Recall by Condition, Age and Gender 122
Table 18 Analyses of the Connectedness of Children’s Responses by Condition, Age and Gender 126
Table 19 Analyses of The Effect of Consistency on the Accuracy of Children’s Recall by Age and Gender 130
Table 20 Analyses of the Completeness of Children’s Testimony by Age and Gender in the First Interview and Overall 132
Table 21 Analyses of the Connectedness of Children’s testimony by Age and Gender in the First Interview and Overall 135
Table 22 Correlations between the accuracy, completeness, connectedness and consistency of children's testimony 138
Table 23 Verdict Choices and Confidence Ratings of Male and Female Jurors 172
Table 24 Verdict Choices and Confidence Ratings of Participants in the One Interview and Two Interviews Conditions 172
Table 25 Verdict Choices and Confidence Ratings of Participants in the Consistent and Inconsistent Conditions 173

Table 26 Verdict Choices and Confidence Ratings of Participants in the 8-year old and 12-year-old Witness Conditions 173
Table 27 Summary of Analyses of Verdict choice 174
Table 28 The Influence of Case Factors on Jurors' Verdict Decisions 176
Table 29 Summary of Analyses of the Perceived Importance of Case Factors on Verdict Choices 178
Table 30 The Perceived Reliability, Competence and Credibility of the 8-year old Witness and the 12-year-old Witness 182
Table 31 The Perceived Reliability, Competence and Credibility of the Witness in the Consistent and Inconsistent Conditions 182
Table 32 The Perceived Reliability, Competence and Credibility of the Witness in the One Interview and Two Interviews Conditions 183
Table 33 Male and Female Jurors' Perceived Reliability, Competence and Credibility Ratings of The Witness 183

List of Figures

Figure 1 Children's Age and the Number of New Details in Interview 1 and Overall 78
Figure 2 Children's Age and the Proportion of Consistent Information in Interview 1 and Overall.....87
Figure 3 Children's Age and the Proportion of Connected Responses in Interview 1 and Overall.....89
Figure 4 The Interaction Between Condition and Age Category on the Number of Details.121
Figure 5 The Interaction Between Children's Gender and Interview Number on the Number of Details.....131

Chapter 1. Introduction

In the United Kingdom, one in five adults report having experienced severe maltreatment, including sexual abuse, physical abuse and neglect when they were under 18-years old (Radford et al., 2011). The prevalence of child abuse rates first came to light during the 1970s in the United Kingdom and the United States, leading to widespread academic research and eventually, several key changes in the law affecting child witnesses. High profile child abuse cases in the 1980s and 1990s motivated research on the interviewing methods used in forensic investigations involving children (for reviews, see Faller 2015; Goodman, 2006), demonstrating that whilst suggestive interviews can result in false reports (Bottoms, Shaver, & Goodman, 1996; Leichtman & Ceci, 1995), under experimental conditions involving stressful and otherwise memorable events, children were able to provide complete and accurate testimonies of their experiences (e.g. Goodman, Bottoms, Schwartz-Kenney, & Rudy, 1991; Peterson & Bell, 1996; Saywitz, Goodman, Nicholas, & Moan, 1991). As a result of emerging research on children's legal competence, the 1988 Criminal Justice Act lifted the requirement of corroboration for evidence provided by children in courts in the United Kingdom and removed the need for judges to warn jurors to exercise caution when dealing with children's testimony (Davies, 1991).

To summarise the outcomes of research in the field, experts on child sexual abuse signed the International, Interdisciplinary Consensus Statement in 1993, establishing the key principles of good practice in interviewing alleged victims of child abuse (Lamb, 1994). The main suggestions of the report included ensuring that children are interviewed as soon as possible after they first disclose allegations of abuse and video recording forensic interviews. The Statement also recommended that interviewers rely primarily on open-ended questions when interviewing children, using direct questions and specific questions only after free

recall has been exhausted. Further research on the impact of the types of questions used by interviewers on the quality of children's testimony led to the development of structured protocols, of which the NICHD protocol has been the most influential (Lamb et al., 2009). The NICHD protocol has been validated by international research demonstrating its advantages over unstructured interviews in ensuring that child witnesses of all ages are able to provide complete and accurate testimonies and variations of the protocol have been adopted in a wide range of countries, including Israel, the United States and Nordic countries (La Rooy et al., 2015). In the United Kingdom, current recommendations for interviewing child witnesses are described in the *Guidance on Joint Investigative Interviewing of Child Witnesses in Scotland* (Scottish Government, 2011) and the *Achieving Best Evidence in Criminal Proceedings* guidance in England, Wales and Northern Ireland (Ministry of Justice, 2011).

The established guidelines and protocols for forensic interviews with child witnesses instruct interviewers on gathering information about the allegations in a single forensic interview, however, many children involved in legal proceedings as witnesses in the United Kingdom are interviewed more than once (Plotnikoff & Woolfson, 2001; Waterhouse, Ridley, Bull, La Rooy, & Wilcock, 2016). The *Achieving Best Evidence* guidance acknowledges that multiple interviews are necessary in some situations, including if the scope of allegations cannot be covered in a single interview, if additional information comes to light following the first interview and if a victim discloses further allegations following the first interview (Ministry of Justice, 2011). However, the question of whether multiple interviews with child witnesses should be avoided or encouraged in routine circumstances remains controversial. While some guidelines explicitly claim that when children are interviewed multiple times they are more likely to disclose abuse and share more forensically relevant details with interviewers (e.g. Extended Forensic Interview: Carnes, Wilson, &

Nelson-Gardell, 1999; Sexual Abuse Forensic Evaluation: Azzopardi, Madigan, & Kirkland-Burke, 2014), other agencies suggest that repeated interviews are distressing to children and may increase the likelihood of false allegations (e.g. Child Advocacy Center: Cross, Jones, Walsh, Simone, & Kolko, 2007). This review considers the conclusions of recent psychological research regarding the effectiveness of multiple forensic interviews in eliciting complete and reliable evidence from child witnesses.

The existing literature on the effectiveness of multiple interviews in eliciting detailed and accurate statements from child witnesses was summarised and evaluated by two separate reviews published in 2008 and 2009 (Goodman & Quas, 2008; La Rooy, Lamb, & Pipe, 2009). Evidence reviewed in both papers suggested that multiple interviewing can increase the amount of correct information children report about events they had witnessed or participated in, and that the decreases in accuracy often associated with multiple interviews in earlier studies might be partially accounted for by the characteristics of the events children were asked to recall and the types of questions they were asked rather than the practice of multiple interviewing. The two reviews were highly influential in shifting professional perceptions of multiple interviewing from a clear risk factor to a more nuanced view that considers the role of biased interviewing and long delays as mediators of the potential negative effects of multiple interviewing on children's ability to give evidence. Although the effect of multiple interviewing has been studied in different age groups, the recall of younger and older children was only infrequently compared systematically, preventing strong conclusions about the potential relationships between age, suggestibility, and the accuracy of children's testimony following multiple interviews (La Rooy et al., 2009). Further limitations of the studies reviewed include researchers' reliance on children's memories of staged events rather than autobiographical memories. Despite the wealth of studies published on the topic of multiple interviews, authors of the reviews emphasised that research on the role of

contextual variables remained scarce and called for further exploration to help professionals make informed decisions about whether multiple interviews would confer benefits in individual cases (Goodman & Quas, 2008; La Rooy et al., 2009).

In the decade that passed since the publication of these two reviews, studies investigating the influence of multiple interviews on children's ability to provide evidence have begun to focus on forensic applications and became more diverse both in terms of their aims and their methods. Ten years ago, few studies analysed the accounts of real child witnesses in real forensic interviews involving more than one session (Ghetti, Goodman, Eisen, Qin, & Davis, 2002; Hershkowitz & Terner, 2007), whereas there is now a growing body of research investigating whether findings from the laboratory transfer to real-life multiple forensic interviewing practices (e.g. Azad, Christianson, & Selenius, 2014; Katz & Hershkowitz, 2012; Leander, 2010; Waterhouse et al., 2016). New measures of the quality of children's forensic testimonies were also explored in recent years; in addition to the amount of information children recall and the accuracy of those details, newly published studies emphasised the importance of developing measures with high forensic applicability. For instance, Brown at al. (2015) and Hubbard et al. (2016) categorised details mentioned by the children according to their content and revealed that the impact of multiple interviews is not uniform across content types (Brown, Lewis, & Lamb, 2015; Hubbard et al., 2016).

The aims of this chapter are twofold; firstly, to determine whether the results of experiments published in the last decade still support the conclusion that multiple interviews can increase the amount of correct information children recall without inherently decreasing the accuracy of their accounts (Goodman & Quas, 2008; La Rooy et al., 2009) and secondly, to explore the new topics and methodologies emerging in the field in the last decade, with a focus on forensically relevant questions and measures. The first section of this chapter will provide a brief overview of the key memory concepts that have influenced research on

multiple interviews with children. The second section will update the results of previous reviews through summarising findings published in the past decade relating to the impact of multiple interviews on the completeness and accuracy of children's testimony. Following the review of the literature on multiple interviews, the third section will consider the gap between cognitive psychological and legal-forensic understandings of children's testimony and suggest that the framework of narrative coherence would support the development of integrated measures of the quality of children's testimony. Finally, an overview of the three empirical studies constituting this thesis will be provided.

1.1 Key Memory Concepts in Research on Multiple Interviews with Children

Understanding the concept of reminiscence and the differences between memory processes in adults and children is essential for interpreting theory and research on children's testimony provided over multiple interviews. According to Erdelyi (1996), the impact of multiple recall occasions on memory performance depends on the balance between two fundamental processes in memory: reminiscence and forgetting. The process of forgetting is often demonstrated in terms of the Ebbinghaus forgetting curve which shows the decline of memory performance on multiple recall tests distributed over a period of time (Ebbinghaus, 1885). However, Ballard (1913) demonstrated that memory performance can also increase over multiple recall occasions, especially when participants memorise complex stimuli rather than word lists. Ballard's experiments investigating schoolchildren's recall of poems revealed the phenomenon of reminiscence, the successful recall of details which were not remembered on previous recall attempts. Reminiscence is an everyday phenomenon, experienced for instance when after leaving the supermarket, one immediately remembers a crucial item they forgot to buy only a few minutes ago. Under certain conditions, the extent of reminiscence can exceed forgetting, leading to hypermnesia, an increase in the overall amount of details recalled on a later occasion compared with an earlier attempt (Erdelyi, 1996).

La Rooy's review of studies documenting children's recall over multiple interviews demonstrated that reminiscence is frequently present under conditions resembling forensic investigations, although hypermnesia is rare (La Rooy et al., 2009). Thus, children generally recall fewer details in later interviews than in the initial session due to the omission of details they reported previously, but they also recall some new details, which may be highly useful in forensic investigations. However, legal professionals are often concerned about the accuracy of newly recalled details, believing reminiscence to be the result of suggestive external influences (Krix, Sauerland, Lorei, & Rispens, 2015). Consistent with this view, Loftus demonstrated in a series of studies that the passage of time between an event and a recall opportunity can lead to the weakening of memory traces, resulting in a potential increase in suggestibility to misinformation (for a review see Loftus, 2005). Increased suggestibility can lead to acquiescence to misleading suggestions but also to source monitoring errors, the mistaken attribution of information from external inputs or guessing to genuine memory traces (Johnson, Hashtroudi & Lindsay, 1993). However, an important confound that research investigating the influence of multiple forensic interviews on suggestibility must overcome is the interrelated effects of the passage of time and multiple recall occasions. Contrary to the views expressed by legal professionals (Krix et al., 2015), reminisced details in interviews separated by relatively short delays (from a few minutes to several weeks) can be highly accurate in the testimonies of both adults (Gilbert & Fisher, 2006; Krix et al., 2015) and children (La Rooy, Pipe, & Murray, 2005).

Differences in Memory Processing Between Adults and Children

While the fundamental processes of memory are active at all stages of cognitive development, children's memory functioning differs from that of adults in several important ways. Although pre-schoolers are already able to provide informative descriptions of salient events, Flin et al. showed that young children's recall performance decreases more sharply

with the passage of time than that of adults and older children (Flin, Boon, Knox & Bull, 1992). Children also appear to be more susceptible to misleading suggestions, although this might be the result of their tendency to comply with adults' suggestions, rather than the weakness of their memory traces (Cassel & Bjorklund, 1995; Gudjonsson, Vagni, Maiorano, & Pajardi, 2016). Forensic interviewers need to be aware of the differences between the way adults and children remember and describe their memories and adjust their questioning techniques to children's social-cognitive developmental level when questioning young witnesses. Concerns about children's suggestibility were amongst the key factors which motivated the development of structured interview protocols and the recommendation that specific questions be avoided when questioning child witnesses (Lamb, Hershkowitz, Orbach, & Esplin, 2008). Similarly, the possibility that repeated questions across multiple interviews could be understood by child witnesses as implicit suggestions has historically been one of the main arguments against interviewing children more than once (Ceci & Bruck, 1993, but see Andrews & Lamb, 2014). The following section will examine the conclusions of past research regarding the impact of multiple interviews on children's suggestibility and their ability to provide detailed descriptions of the events they experienced.

1.2 Fundamental Questions in the Study of Multiple Interviews with Child Witnesses

Over ten years ago, two reviews systematically analysed over 80 studies investigating the effect of multiple interviews on children's memory (Goodman & Quas, 2008; La Rooy et al., 2009). In the past decade, a further 21 studies have been published that investigated the impact of multiple interviews on children's ability to provide detailed and accurate descriptions of past events (see Table 1 and Table 2). Researchers followed the suggestions of earlier reviews and began to explore the role of contextual variables, such as the type of event children are asked to recall, children's age and the delay between the event and the

recall tests influencing the impact of multiple interviews on children's testimony, which led to the development of a more nuanced view on the benefits and risks of multiple interviews.

Previous reviews categorised research on multiple interviewing based on three broad factors; the type of event that children were questioned about by interviewers, the presence or absence of a delayed control condition, and whether or not the interviews were biased or suggestive.

Type of event

Studies published since 2008 examined children's recall of a wide array of events, including neutral and stressful autobiographical memories, staged play sessions, filmed events and entirely false memories (see Table 1 and Table 2). The characteristics of the events children are exposed to and asked to recall in research contexts are crucial in determining children's memory performance, because the saliency, emotional valence and uniqueness of experiences affects the processing, storage and retrieval of memory traces (Peterson, 2011; 2015; Peterson & Bell, 1996; Rudy & Goodman, 1991). Rudy and Goodman contrasted the memories of children who participated in games with a control group who observed the games and found that children were more resistant to false suggestions when they were active participants in the event rather than passive witnesses (Rudy & Goodman, 1991). In addition, Peterson's research suggests that children remember stressful or otherwise emotionally meaningful events better after a passage of time than neutral events (Peterson, 2011; 2015; Peterson & Bell, 1996). Recently, studies began to investigate the role of event uniqueness, revealing that the reports of children who experienced repeated events of the same event with some variations included more contradictions (Connolly, Price, Lavoie, & Gordon, 2008) and contained more errors in

response to specific questions (Powell & Roberts, 2002) than the descriptions of the control group.

Researchers' choice of the type of event children would be questioned determines the measures they can use to assess the quality of testimonies and influences the external validity of their methods. Staged events allow the experimenter a high extent of control over the events and therefore provide a very good measure of the accuracy of children's recall. However, the downside of relying on staged events is their lack of resemblance to the stressful and traumatic events children need to recall in forensic contexts. There are a variety of methods used in staged experiments to improve external validity, such as adding an element of innocuous touching to the event children experience, or swearing children to secrecy (Salmon, Pipe, Malloy, & Mackay, 2012). Another way of ensuring the forensic relevance of staged experiments is to use the same methods of questioning as police investigations, for example by training interviewers on using the NICHD protocol (Brown et al., 2015). Some laboratory studies have also tried to model the effect of coaching by adults through questioning children about entirely false events (Otgaar, Verschuere, Meijer, & van Oorsouw, 2012).

In contrast to staged events, questioning children about real-life events allows researchers to study the effects of stressful or traumatic autobiographical memories (Baugerud et al., 2014; Peterson, 2011, 2015). However, as important life events may constitute a frequent topic of conversation in children's families, it is difficult to establish whether the information reported originates in genuine memory traces or has been supported by shared reminiscence with family members (La Rooy et al., 2009). In recent years, research has begun to focus on field studies involving the analysis of transcripts of multiple forensic interviews with alleged child victims of sexual or physical abuse (Katz & Hershkowitz, 2013; Leander, 2010; Waterhouse, Ridley, Bull, La Rooy, & Wilcock, 2016). External validity is

highest in the case of field research, as in these studies, children report their experiences of child abuse under real-life forensic conditions. However, the accuracy of children's testimony cannot be determined in field studies, as researchers do not know the 'ground truth' of the events children are questioned about. Therefore, field studies must rely on presumed correlates of accuracy, such as consistency or the number of contradictions in children's testimonies (Waterhouse et al., 2016). Courtroom materials often suggest that witness consistency is a strong correlate of accuracy, and that therefore, witnesses should be challenged on the truthfulness of inconsistent details in their story (Krix et al., 2015). However, past research has found that whilst the consistency of individual details across interviews can be indicative of their accuracy, the overall consistency of multiple interviews does not predict their overall accuracy (Gilbert & Fisher, 2006; La Rooy et al., 2005).

Control group

Including a delayed control group in the experimental design is essential for distinguishing the effects of multiple interviewing from the influence of forgetting resulting from the delay between the event children are asked to recall and the interviews. Without a control group matched for the length of delay, the source of potential changes in the completeness and accuracy of children's accounts remains unclear. Goodman et al. found that the accuracy of children's testimony was higher in the first interview 2 weeks after a medical procedure than in the second interview 4 weeks after the event (Goodman et al., 1991). However, a comparison of the accuracy of children's recall in the second interview with the reports of a control group interviewed for the first time 4 weeks after the event showed that the latter group performed worse than the multiple interview group did on their second trial. Thus, the decrease in accuracy observed in the test group between the first and second interviews was a result of delay, rather than multiple interviewing. Multiple interviewing, in fact, appeared to have a slight positive influence on the accuracy of children's recall,

potentially due to the reminder interview preserving some of children's memories (Goodman et al., 1991).

Unfortunately, and as with the studies reviewed previously (Goodman & Quas, 2008; La Rooy et al., 2009), few recent experiments have included a control condition matched for delay. However, in many newer studies the lack of a delayed control group can be justified, as studies began to move away from evaluating the value of multiple interviews per se towards analysing interactions between multiple interviews and different questioning techniques or contextual variables.

Open-ended and suggestive questioning techniques

One important contribution of the previous reviews has been to draw a line between studies investigating multiple interviews using open-ended questions and those relying on potentially suggestive methods (Goodman & Quas, 2008; La Rooy et al., 2009). Forensic interviewing guidelines designed for questioning child witnesses strongly emphasise the importance of giving children the opportunity to provide a free narrative of the events in their own words before starting to ask more direct questions. Interviewers are trained to use invitations first before moving on to more direct questions and to avoid using option-posing questions unless other questioning methods have been exhausted (Lamb et al., 2009; Ministry of Justice, 2011; Scottish Government, 2011). Although structured interviewing methods are not used in all jurisdictions, and non-adherence is common even amongst interviewers who are trained to use these guidelines (Lamb, 2016), it is important to conduct studies modelling the environment of *ideal* forensic investigations to demonstrate the effect of multiple interviews in 'good' interviews.

Of the 75 studies analysed by La Rooy et al. (2009), 45 were non-suggestive; the remaining studies used suggestive methods, such as misleading questions, props or social

pressure, and sometimes several of them in combination. The studies reviewed by Goodman and Quas (2008) also relied on suggestive questioning techniques, such as the researchers implying false information and repeatedly asking misleading questions. Both reviews concluded that non-suggestive multiple interviews can increase the amount of information children recall about events without decreasing the accuracy of their memory but stated that the literature was inconclusive regarding the question whether the effects of suggestive interviewing are exacerbated by multiple interviews or conversely, whether multiple interviews may help children resist suggestion under certain conditions.

Following the aims and structure of previous reviews (La Rooy et al., 2009) research designs relying on open-ended questioning methods (Table 1) and studies using suggestive designs (Table 2) will be discussed separately.

Paradigms Relying on Free Recall and Open-ended Questions

In the studies reviewed in this section, children were first given the opportunity to provide a free recall narrative and then asked directive questions, sometimes followed by option-posing questions. A summary of the results of all studies published since 2008 that investigated children's recall in multiple interviews involving open-ended questions is presented in Table 1.

Completeness / number of correct details

In most of the studies reviewed by La Rooy et al. (2009), the number of correct details recalled by children decreased in each successive interview (e.g. Cassel & Bjorklund, 1995; Salmon & Pipe, 2000). However, a subset of studies found that the number of correct details remained constant (e.g. Ackil, van Abbema, & Bauer, 2003; Merritt, Ornstein, & Spicker, 1994) or even increased over several interview occasions (e.g. Peterson, Pardy, Tizzard-Drover, & Warren, 2005; Pipe, Sutherland, Webster, Jones, & La Rooy, 2004). Information

about real-life experiences, both stressful and enjoyable, appeared to be better preserved than information about staged events and videos (La Rooy et al., 2009).

New details emerged in later interviews in virtually every study published in the past decade that included measures of the completeness of children's recall across multiple interviews. However, consistent with the findings of earlier reviews (Goodman & Quas, 2008; La Rooy et al., 2009), most laboratory studies found either that the number of details reported decreased in each successive interview, or that the number of details remained constant across interviews (Baugerud, Magnussen, & Melinder, 2014; Peterson, 2011; 2015; Salmon et al., 2012). Remarkably, Peterson's studies investigating children's recall of a major injury revealed that children were still able to recall comparable amounts of information in later interviews 5 or 10 years after the event with what they reported in their first interview (Peterson, 2011, 2015).

Hypermnesia was found in two staged experiments (Dietze, Sharman, Powell, & Thompson 2013; Knutsson, Allwood, & Johansson; 2011). Interestingly, both experiments studied children's recall of events they witnessed passively, which is less memorable than active participation according to previous research (Baker-Ward, Hess, & Flannagan, 1990; Rudy & Goodman, 1991). Knutsson et al.'s study included a delayed control condition, thus, the authors were able to conclude that children interviewed 4 days and 8 days after having watched a short film clip provided more detailed testimonies that children questioned for the first time 8 days after the event (Knutsson et al., 2011). A field study investigating the reports of children alleging sexual abuse across one to three interview sessions also reported a remarkable extent of hypermnesia; compared with an average of 250 details mentioned in interview one, children recalled an average of 488 details in second interviews (Leander, 2010). The 10 children who were interviewed a third time recalled an average of 308 details,

less than the average recall in the second interview but still hypermnesic compared with their first recall.

Despite the fact that many children who are interviewed about suspected child abuse allege more than one occurrence of abuse, in the past, laboratory research assessing the impact of multiple interviews on children's recall focused almost exclusively on the memories of children who experienced a single instance of the event in question (Connolly & Read, 2006). Price et al. compared the recall of children who took part in 4 play session with a control group of children taking part in a single session in two interviews separated either by 4 months or one year and found that the experimental group forgot less details between interviews (Price, Connolly, & Gordon, 2016). The authors concluded that multiple interviews may be beneficial for children who experienced repeated abuse but costly for those who were only abused once, as these children may forget more correct details and report more incorrect details in successive interviews (Price et al., 2016).

Findings from the laboratory showing that children with repeated experience benefitted more from multiple interviews than children who experienced a single occurrence of an event (Price et al., 2016) were not replicated by Azad et al.'s field study comparing the recall of two groups of children who witnessed homicidal violence (Azad, Christianson, & Selenius, 2013). Children who have witnessed domestic violence on previous occasions reported more details in a single interview than those whose first experience witnessing abuse was the homicide (Azad et al., 2013). When interviewed more than once, children with a single experience were able to provide more detailed testimonies (Azad et al., 2013). Contrasting results may be partially explained by the higher degree of similarity between the events that children experienced in Price et al.'s (2016) study compared with the occurrences of domestic abuse witnessed by children in Azad et al.'s (2013) research, especially as the final occurrence of abuse resulted in homicide.

In addition to investigating whether multiple interviews led to an increase in the number of details recalled, recent studies also analysed the content or forensic relevance of newly recalled details. Most studies used some variation of the *person / action / location / object / time* content code, but no consensus emerged regarding the impact of multiple interviews on children's recall of specific content types (Brown et al., 2016; Hubbard et al., 2016; Peterson, 2011, 2012). Peterson (2011) reported an increase in the frequency of action-related details in the second interview, whilst Hubbard et al. (2016) found a decrease in the number of action-related details and an increase in non-dynamic details.

Other studies measured the forensic relevance of information reported in multiple interviews using a *central / peripheral* (Katz & Hershkowitz, 2013; Peterson, 2011, Waterhouse et al., 2016) or *sexual / sensitive / neutral* (Leander et al., 2010) distinction. Two field studies reported that the first interview included the highest number and proportion of forensically relevant details, although a substantial number of new central details emerged in later interviews (Katz & Hershkowitz, 2013; Waterhouse et al., 2016). In contrast, Leander (2010) found that the number of forensically relevant details increased across multiple interviews, with 45, 100 and 103 sexual details reported in the first, second, and third interview, respectively.

Accuracy / proportion of correct details

Among the 45 non-suggestive studies analysed in previous reviews (Goodman & Quas, 2008; La Rooy et al., 2009), mixed results were found regarding the impact of multiple interviewing on the accuracy of children's testimony. Although many studies reported a decrease in accuracy resulting from an increase in erroneous information, a decrease in correct information or both an increase in errors and a decrease in correct details (e.g. Baker-Ward et al., 1990; Pipe, Gee, Wilson, & Egerton, 1999; Salmon & Pipe, 1997), this effect

was not universal across studies. Several studies reported that accuracy remained constant across interviews (e.g. Greenhoot, Ornstein, Gordon, & Baker-Ward, 1999; Gordon & Follmer, 1994; Salmon & Pipe, 2000), and two experiments showed a decrease in errors in some age groups from the first to the second interview (e.g. Pipe et al., 1999, Experiment 1: 8-11-year-olds; Peterson et al., 2005: 5-7 and 8-9-year-olds).

Although almost all recent studies investigating multiple open-ended interviews included some measure of accuracy, no consensus emerged regarding the relationship between multiple interviews and accuracy. In the experiments of Hubbard et al. (2016) and Knutsson et al. (2016), the accuracy of children's reports did not decrease across interviews. Studying children's recall of real-life stressful events, Baugerud et al. (2014) and Peterson (2015) also found that accuracy remained constant across multiple interviews. The results provide further support for the view that multiple interviews do not inherently compromise the accuracy of children's testimony.

Differences in research methods partially account for why some studies found a decrease in accuracy across multiple interviews while others reported the opposite. Whilst children's memories of real-life events remained accurate in some studies (Baugerud et al., 2014; Peterson, 2015), others found a decrease in accuracy (Peterson, 2011). Peterson (2015) contrasted children's recall of a traumatic injury and the following hospital stay and found that children's reports about the injury remained highly accurate, whilst the accuracy of their statements regarding the hospital stay decreased. Consistent with the theory that children's recall of stressful events is superior to their memories of neutral events Peterson (2015) suggests that children's memories of the injury remained accurate because they experienced the accident as more stressful than the hospital stay.

However, the accuracy of children's recall also remained constant across multiple interviews in staged experiments, suggesting that differences in the salience and stressfulness of the events experienced cannot fully account for contrasting results regarding the impact of multiple interviews on accuracy. Hubbard et al. (2016) found no difference in the accuracy of children's recall of repeated play sessions across three interviews conducted over consecutive days. Similarly, Knutsson et al. (2011) reported that the proportion of correct details did not decrease in a second interview conducted 4 days after the first one. However, the accuracy of children's recall decreased after a 5-minute break in Salmon et al.'s study (2012), although errors in the second interview resulted mainly from follow-up prompts rather than open-ended invitations. Experimental studies also evaluated the effectiveness of methods designed to improve the accuracy of children's recall across multiple interviews. Neither mental context reinstatement (Dietze et al., 2011), nor free drawings (Salmon et al., 2012) influenced the accuracy of children's testimony.

Consistency

Previous reviews concluded that reminiscence in multiple interviews is the rule, rather than the exception (Goodman & Quas, 2008; La Rooy et al., 2009). By definition, new details recalled in multiple interviews decrease the consistency of children's testimony, therefore, multiple interviews have a negative impact on consistency.

Consistent with previous findings, recent field studies analysing the consistency of children's forensic testimonies provided as part of investigations in Israel (Katz & Hershkowitz, 2013) and the United Kingdom (Waterhouse et al., 2016) found significant reminiscence effects in later interviews, and therefore poor consistency across interviews. Both studies reported that the decrease in consistency in multiple interviews was mainly driven by reminiscence and the omission of details mentioned in previous interviews. Katz &

Hershkowitz (2013) analysed interviews conducted with strict adherence to the NICHD protocol, separated by a break of only 7 minutes and found no contradictory information reported within or across interviews. In contrast, Waterhouse et al.'s sample included interviews separated by much longer delays that consisted mainly of closed-ended questions with few invitations and this study found a low proportion of contradictory information reported in the interviews (Waterhouse et al., 2016). From the 14 interviews involving contradictions, 6 had low forensic relevance, 5 included delayed disclosures, and 2 demonstrated a changing understanding of the concept of touch.

Baugerud et al. (2014) investigated the accuracy of consistent, omitted and reminisced details over the course of several interviews about children's removal from their family by child protective services. Consistent with previous research (Gilbert & Fisher, 2006; Krix et al., 2015; La Rooy et al., 2005), results demonstrated that low consistency does not necessarily indicate low accuracy, as children's recall remained highly accurate in the second interview despite a high proportion of new details recalled. In children's free recall provided in the second interview, 94-99% of reminisced information was accurate, although this proportion decreased to 79-87% in response to focused questions. Similarly, although Price et al. (2016) found that the recall of children with repeated experience appeared less consistent and less accurate than the reports of children who took part in a single play session, the difference in accuracy disappeared when 'internal intrusions', details that children experienced but attributed to the wrong occurrence of the event, were taken as correct. These results challenge laypeople's and legal professionals' assumption that the consistency of witnesses' statements is a reliable measure of the accuracy of their memories (Krix et al., 2015).

Children's age

Trace strength theory predicts that the memory of younger children would be more adversely affected by long delays than the recall of older children, suggesting that multiple interviews may affect young children's testimony more adversely (Brainerd & Reyna, 1998). On the other hand, young children usually provide less complete testimonies in a single interview and have a shorter attention span, which may mean that multiple interviews are often necessary with very young children (Faller, Cordico-Steele, & Nelson-Gardell, 2010; Langballe & Davik, 2017). Despite the abundance of studies investigating the role of age in mediating the effects of multiple interviewing, no consensus emerged in the literature.

The majority of studies analysed in previous reviews (Goodman & Quas, 2008; La Rooy et al., 2009) found that younger children forgot more than older children both in terms of the number of details reported and the accuracy of details across interviews (Baker-Ward, Gordon, Ornstein, Larus, & Clubb, 1993; Peterson et al., 2005; Pipe et al., 1999; Salmon & Pipe, 1997). Two studies found that children's age was associated with the effectiveness of providing an additional interview between two recall occasions separated by a long delay to preserve their memories but their results contradicted each other; in Tizzard-Drover & Peterson's study (2004), this inoculation effect occurred for 3-4 year-olds only whilst Gee & Pipe (1995) found that 9 year-olds, but not younger children, benefitted from an additional interview.

The role of children's age in mediating the effects of multiple interviews on the accuracy and completeness of their recall was widely studied in the past decade. Despite previous research demonstrating young children's ability to recall events accurately in open-ended interviews (Quas, Goodman, Ghetti, & Redlich, 2000), most experiments conducted since 2009 reported that the recall of younger children was less accurate than the memories of older children, regardless of the number of interviews (e.g. Baugerud et al., 2014; Hubbard et al., 2016; Peterson, 2011). This discrepancy might result from methodological limitations, as

many of the studies reviewed used non-participatory events (e.g. videos), which might be less salient for younger children (Quas et al., 2000). However, developmental differences were not always observed in the number of details children reported. Analysing real-life forensic interviews, Leander (2010) found that age did not influence the number and content of details reported by 5-17-year-old children. These contradictory findings might be partially accounted for by the relatively old mean age of children in Leander's sample (M = 10.63, SD = 3.18) compared with the other studies reviewed.

Results regarding whether multiple interviews affect the recall of younger and older children differentially were inconsistent. Studying children's recall of sexual abuse (Katz & Hershkowitz, 2013) and removal from their families by CPS (Baugerud et al., 2014), field studies found that the increase in the number of details reported in multiple interviews versus a single interview was more substantial for younger children. However, Peterson (2015) reported that the younger the children were, the more they forgot, both in terms of the number of details reported and their accuracy.

Delay

Unfortunately, in forensic contexts, children often have to testify about events they experienced years prior to the investigation. The impact of this delay is further exacerbated by the passage of time between multiple interviews conducted by the police and court hearings for witnesses called to testify in trials (Plotnikoff & Woolfson, 2009). As longer delays lead to more forgetting (Salmon & Pipe, 2000), determining whether multiple interviews may help preserve or recover information from months or even years ago is crucial. Studies analysed for previous reviews suggested that the extent of reminiscence (Powell & Thompson, 1997) and the accuracy of newly recalled details (Pipe et al., 1999) decreases when the delay between interviews is longer.

Leander (2010) and Katz & Hershkowitz (2013) reported that the time elapsing between the alleged abuse and the first forensic interview ranged between one day and a year or more. In contrast, participants in all experiments published in the past decade were interviewed within a month of the event they were asked to recall. However, delays between interviews were more varied, from only minutes separating the first and second interview (Salmon et al., 2012), to follow-up interviews conducted years after the initial recall attempt (Peterson 2011, 2015). Contrasting results regarding the impact of multiple interviews on the accuracy and completeness of children's reports are not easily explained in terms of the length of delay between the event and interviews or between the first and second interview. Despite a similar length of delay between the event children were questioned about and the final interview, Dietze et al. (2011) and Salmon et al. (2012) found that the accuracy of children's recall decreased in the second interview, whilst Hubbard et al. (2016) and Knutsson et al. (2011) found that the proportion of correct details remained constant across interviews. Furthermore, whereas Katz & Hershkowitz (2013) suggested that interviews separated by a few minutes can effectively increase the number of details reported by children without leading to an increase in the number of contradictions, Salmon et al. (2012) reported that after a 5-minute break, children reported few new details and the accuracy of the reminisced details was low.

Overall, it is difficult to draw any conclusions about the role of delay in mediating the impact of multiple interviews on the completeness and accuracy of children's recall, as no recent studies using open-ended methods manipulated the timing of interviews within experiments, and studies using different delays also differed in the methodologies used and the age of children studied, preventing direct comparisons.

Table 1 *Results of Studies Published After 2008 Relying on Open-ended Questioning methods*

Researcher(s)	Number of interviews	Last interview	Control group	Age group	Type of event	Correct	Errors	Accuracy
Baugerud et al. (2014)	2	3 months	no	3-12	Real-life stressful	0	n/a	0
Peterson (2011)	2-13	2 years	no	2-13	Real-life stressful	0	n/a	-
Peterson (2015) Injury	3-6	10 years	yes	13-15	Real-life stressful	0	n/a	0
Peterson (2015) Hospital	3-6	10 years	yes	13-15	Real-life stressful	-	n/a	-
Hubbard et al. (2016)	3	1 week	no	4-10	Staged	n/a	n/a	0
Price et al. (2016) Exp. 1	2	4 months	no	7-8	Staged	-	n/a	n/a
Price et al. (2016) Exp. 2	2	1 year	no	4-5	Staged	-	n/a	n/a
				6-7		-	n/a	n/a
Price et al. (2016) Exp. 3	2	1 year	no	4-5	Staged	-	n/a	n/a
Salmon et al. (2012) Exp. 1	2	10 days	no	5-7	Staged	-	+	-
Dietze et al. (2011)	2	3 weeks	no	6,9,11	Film	+	+	n/a
Knutsson et al. (2011)	2	8 days	yes*	9-11	Film	+	n/a	0
				Adult		+	n/a	+

(continued)

(continued)

Azad et al. (2014)	1-6	n/a	no	3-17	Field	n/a	n/a	n/a
Katz & Hershkowitz (2013)	2	1 year	no	5-14	Field	n/a	n/a	n/a
Leander (2010)	1-3	10 years	no	5-17	Field	n/a	n/a	n/a
Waterhouse et al. (2016)	2-5	1 year	no	3-14	Field	n/a	n/a	n/a

Note: In the columns labelled Correct, Errors, and Accuracy “0” denotes no statistically significant change in recall across multiple interviews, “-“ denotes a significant decrease across repeated interviews, “+” denotes a significant increase across repeated interviews, and “n/a” identifies dependent variables that were not measured. Where developmental effects were explored the results are listed separately for each age group. In the column labelled Delayed Control, "*" denotes a significant effect compared with control.

Suggestive Questioning Paradigms

The question of whether multiple forensic interviews increase children’s suggestibility attracted the interest of researchers and legal professionals following highly publicized cases of alleged child sexual abuse in which children’s testimonies were deemed inadmissible due to improper questioning techniques, such as the use of suggestive props, misleading questions, fantasy play and relentless repeated questioning about the same details, even if children denied the allegations (Bruck & Ceci, 1999). Although current guidelines on conducting forensic interviews with children recommend the avoidance of these methods (Lamb et al., 2009; Ministry of Justice, 2011; Scottish Government, 2011), research on the effects of suggestion is unfortunately still relevant, as many forensic interviews are not conducted in accordance with best practice guidelines (Lamb, 2016). A summary of the results of all studies published since 2008 that investigated children’s suggestibility in multiple interviews is presented in Table 2.

The studies analysed in previous reviews (Goodman & Quas, 2008; La Rooy et al., 2009) used a variety of suggestive methods, including peer pressure, social pressure, requests that children imagine false events, misleading physical props, praise and encouragement for correct answers, interviewer-provided misinformation, appeals to help the interviewer and negative reinforcement when children did not comply. More than half of these experiments concluded that multiple interviews increased children's suggestibility (e.g. Ornstein, Gordon, & Larus, 1992, Principe & Ceci, 2002; Quas & Schaaf, 2002). La Rooy et al. (2009) highlighted that studies involving multiple methods of suggestion, false events or long delays between interviews found increased suggestibility in later interviews while studies that used only misleading yes/no questions often did not find an increase in the amount of erroneous information reported by children (e.g. Flin et al., 1992; Gordon & Follmer, 1994).

However, the combination of multiple methods of suggestion and multiple interviews did not always lead to more inaccuracies in children's testimony than using the same methods in a single interview. For instance, in Quas et al.'s (2007) experiment, biased interviewers attempted to convince children that they played in the company of a man when in fact they were on their own. Children interviewed on a single occasion made more errors and acquiesced more frequently to the suggestion that the man was in the room than children interviewed multiple times using the same questioning technique, suggesting that multiple interviews can sometime decrease the negative impact of suggestion on the accuracy of children's testimony (Quas et al., 2007).

Studies published in the last decade examining the effect of multiple interviews on children's suggestibility involved props, post-event leading information, false events and misleading questions. Unlike earlier studies, most experiments relied on a single method of suggestion and many research designs included a non-suggestive control condition, aiming to model situations that can occur in forensic contexts. Melinder et al.'s (2010) study had

especially high forensic relevance, as it involved a comparison of the accuracy and completeness of children's recall of a medical examination in multiple interviews conducted by trained police interviewers using verbal interviewing methods and play-oriented, prop-assisted interviews conducted by clinicians. Unlike police interviews, which were found to elicit highly accurate information in the second interview, clinician-led interviews involving less free recall questions and more prop-assisted play led to a decrease in accuracy (Melinder et al., 2010). Remarkably, children's recall in a third, open-ended interview was equally accurate across the two conditions, suggesting that non-suggestive multiple interviews may ameliorate the effect of previous suggestive recall occasions (Melinder et al., 2010)

Equally relevant to forensic investigations were the results of studies evaluating courtroom procedures by conducting a direct examination under experimental conditions, followed by a cross-examination. The studies involved a control group who experienced a second direct examination rather than a cross-examination to establish whether cross-examination style questioning is detrimental over and above the potential negative effects of multiple interviews (Fogliati & Bussey, 2014; Jack & Zajac, 2013; Righarts et al., 2014; O'Neill & Zajac, 2013). Results demonstrated a consensus in that cross-examination style second interviews, including suggestive and challenging questions, led to a decrease in the accuracy of children's testimony (Fogliati & Bussey, 2014; Jack & Zajac, 2013; Righarts et al., 2014; O'Neill & Zajac, 2013). However, in Jack & Zajac (2013) and O'Neill & Zajac (2013) accuracy decreased in the second interview of the control group as well, although to a smaller extent than in the cross-examination group. In contrast, Fogliati & Bussey (2014) found that the children's recall in the control condition was equally accurate in the first and second interview. Righarts et al.'s study (2014) also involved a third interview, taking place after the direct-examination and the cross-examination. Similarly to Melinder et al. (2013), results showed that although children's recall was less accurate in the cross-examination than

the direct examination, the accuracy of their recall in the third, non-suggestive interview was comparable to the accuracy of their initial recall (Righarts et al., 2014).

The remaining studies investigated whether multiple interviews exacerbate or ameliorate the impact of false suggestions on the accuracy of children's recall. Studying the reports of children with mild and moderate intellectual disabilities, Brown et al. (2015) found that multiple interviews provided an 'inoculation effect' against forgetting and suggestibility (Brown, Lewis, & Lamb, 2015). Children interviewed 1 week and 6 months after they participated in a first aid training provided more accurate and detailed description of the event and were less likely to acquiesce to false suggestions than children in the control group interviewed for the first time 6 months after the training (Brown et al., 2015). These findings suggest that there may be individual differences in the impact of multiple interviews on the quality of children's testimony, and children with intellectual disabilities may benefit markedly from multiple interviews.

London et al. (2009) investigated the impact of post-event misinformation on the accuracy of children's recall in first interview 1 or 2 weeks after the event and a second interview 10 months or 1 year later (London, Bruck, & Melnyk, 2009, Experiments 1 & 2). Results showed that misleading suggestions had a lasting effect on children's memory when they were questioned using recognition probes, but the frequency of suggested details in children's free recall decreased from the first to the second interview. Price & Connolly (2013) used a similar methodology to investigate the persistence of misleading information after a yearlong delay in children with a repeated or a single experience of a play session. The study found that children with repeated experience were more likely to acquiesce to suggestions that a non-experienced detail occurred than children with a single experience, indicating that the combination of suggestive questions and multiple interviews may

especially damage the accuracy of children's testimony when they are asked to describe events they experienced repeatedly (Price & Connolly, 2013).

In contrast to London et al.'s (2009) finding that the effect of false information on the accuracy of children's testimony decreased across multiple interviews, Otgaar et al. (2012) found no difference in children's likelihood to acquiesce to suggestions that they participated in a hot air balloon ride between the first and second interviews (Otgaar, Versuchere, Meijer & Van Oorsouw, 2012). However, this study involved a false memory paradigm including the interviewer presenting children with an entirely false narrative in each interview using context reinstatement and guided imagery techniques, therefore, it is remarkable that the second interview did not lead to an increased number of children acquiescing to the suggestions.

Overall, evidence from studies published in the past decade supports the view that suggestive questioning methods decrease the accuracy of children's testimony, but also suggests that multiple interviews do not necessarily exacerbate the impact of suggestion. Providing children with an opportunity to recall the events in a non-suggestive interview following earlier suggestions led to an increase in accuracy compared with earlier interviews (London et al., 2009; Melinder et al., 2013; Righarts et al., 2014).

Children's age

Consistent with Trace Strength Theory (Brainerd & Reyna, 1998), most studies analysed in previous reviews (Goodman & Quas, 2008; La Rooy et al., 2009) found that younger children were more suggestible than older children and that multiple interviews used in combination with other suggestive techniques compromised the accuracy of their testimonies to a greater extent.

Few studies published in the last decade compared the suggestibility of different age groups across multiple interviews. Consistent with previous findings, Fogliati & Bussey (2014) reported that younger children were less accurate than older children in the second interview when both interviews were non-suggestive. However, when a first non-suggestive interview was followed by a second interview including suggestive and credibility-challenging questions, the performance of 6-year-olds and 8-year-olds deteriorated equally (Fogliati & Bussey, 2014). In contrast, O'Neill & Zajac (2013) found that 5- and 6-year-olds changed more of their original responses in response to suggestive questions than 9- to 11-year-olds and the changes they made were also less likely to be correct. Similarly, Jack & Zajac (2013) found that 9-11-year-old children were more likely to change their answers in response to cross-examination style questioning than adolescents and adults, although the recall of 14-16-year-olds was also significantly affected by suggestive and credibility challenging questions.

Delay

Studies analysed in previous reviews (Goodman & Quas, 2008; La Rooy et al., 2009) that involved a delayed control condition demonstrated that long delays contributed to increased suggestibility independently from the impact of multiple interviews (e.g. Baker-Ward et al., 1993; Quas et al., 2007).

In contrast to the predictions of Trace Strength Theory (Brainerd & Reyna 1998), the results of studies published in the past decade did not unanimously support the view that longer delays between interviews lead to increased suggestibility.

Using a misinformation paradigm, London et al. (2009) found that the impact of false reminders on the accuracy of children's recall was lower 10 months and 16 months after the event in comparison to the initial interview 1 or 2 weeks after. Similarly, Righarts et al.

(2014) found that children were no more suggestible in response to misleading questions following an 8-month delay than a few days after the event. Interestingly, O'Neill & Zajac (2013) reported a significant interaction between children's age, the delay between interviews and the accuracy of their responses to misleading questions. One week after the event to be remembered, younger (5-6) and older (9-11) children were equally likely to change their initial responses but after 6 months the number of contradictions in younger children's recall increased to a greater extent.

The small number of recent studies that manipulated the delay between interviews experimentally or compared the recall of different age groups does not allow for strong conclusions regarding the role of these variables in mediating the accuracy of children's recall across multiple suggestive interviews, but their results appear to complicate the previous consensus regarding the relationship between age, delay and accuracy.

Table 2 *Results of Studies Published After 2008 Relying on Suggestive Questioning Methods*

Researchers	Type of suggestion	Number of interviews	Last interview	Control group	Age	Type of event	Increase in suggestibility
Melinder et al. (2010)	1. Misleading questions 2. Misleading props	2 or 3	6 weeks	Yes	4	Medical Exam.	yes
Brown et al. (2015)	1. Misleading questions	1 or 2	6 months	yes	4-12	Staged	No
Fogliati & Bussey (2014)	1. Misleading questions 2. Cross-examination	2	minutes	no	6,8	Staged	n/a
Jack & Zajac (2013)	1. Misleading questions 2. Cross-examination	2	8 months	no	9-11, 14-16, Adult	Staged	n/a
London et al. (2009) Exp. 1	1. Misleading information 2. Misleading questions	4	16 months	no	4-6	Staged	No
London et al. (2009) Exp. 2	1. Misleading information 2. Misleading questions	3	10 months	no	4-9	Staged	No
O'Neill & Zajac (2013)	1. Misleading questions 2. Cross-examination	2	6 months	no	5-6 9-11	Staged	n/a
Otgaar et al. (2012)	1. Event creation	2	n/a	no	8-10	Staged	No
Price & Connolly (2013)	1. Misleading questions.	2	1 year	no	4-5 6-7	Staged	n/a
Righarts et al. (2014)	1. Misleading questions 2. Cross-examination	3	8 months	no	5-6	Staged	n/a

Conclusions

In the last decade, the methodologies used to study repeated interviewing have become more diverse and more focused on forensic applicability. Field studies are starting to bridge the gap between laboratory findings and real police investigations, whilst attempts to replicate forensic contexts in experimental studies have increased the external validity of laboratory research.

Although hypermnesia occurred only infrequently in experimental research, most studies reported that multiple interviews increased children's cumulative recall through reminiscence (e.g. Baugerud, Magnussen, & Melinder, 2014; Peterson, 2011; 2015; Salmon et al., 2012). The proportion of reminisced details was higher in forensic transcripts than in experimental research (Katz & Hershkowitz, 2013; Leander, 2010; Waterhouse et al., 2016). Importantly, Leander et al. (2010) found that newly recalled details in the second and third interview had high relevance for the investigation.

Consistent with the conclusions of previous reviews (Goodman & Quas, 2008; La Rooy et al., 2009), studies using a diverse range of methodologies to measure the recall of children of all ages reported that children's recall can remain highly accurate in multiple interviews, indicating that multiple interviews do not inherently decrease the accuracy of children's testimony (Baugerud et al. 2014; Brown et al., 2015; Fogliati & Bussey, 2014; Knutsson et al., 2014; London et al., 2009; Melinder et al., 2010; Peterson, 2015; Righarts et al., 2014). However, in experiments where the accuracy of children's recall did decrease over the course of multiple interviews, the impact of delay or suggestive interviewing techniques did not always account for this tendency (Jack & Zajac, 2013; O'Neill & Zajac, 2013; Salmon et al., 2012). Further research is needed to clarify the processes underlying of the negative impact of multiple interviews on children's accuracy in a subset of studies.

The results of studies analysed in the present review and previously (Goodman & Quas, 2008; La Rooy et al., 2009) unanimously indicate that multiple interviews decrease the consistency of children's testimony, as children omit details mentioned in the first interview and add new details in later interviews (Baugerud et al., 2014; Dietze et al., 2013; Katz & Hershkowitz, 2013; Knutsson et al, 2014; Leander, 2010; Waterhouse et al., 2016). However, Baugerud et al. (2014) found that children's recall remained highly accurate across multiple interviews, despite the low proportion of consistent details between interview 1 and interview 2, suggesting that consistency should not be understood as a correlate of accuracy.

Consistent with previous reviews, recent studies confirmed that all kinds of suggestive questioning methods, including cross-examination techniques, false memory paradigms, misinformation and misleading questions decrease the accuracy of children's testimony (Fogliati & Bussey, 2014; Jack & Zajac, 2013; London et al., 2009; Melinder et al., 2010; Righarts et al., 2014; O'Neill & Zajac, 2013; Otgaar et al., 2012). However, the results of Melinder et al. (2010) and Righarts et al. (2014) suggest that providing children with an opportunity to recall the events in response to open-ended questions can ameliorate the effects of previous suggestive interviews. These findings demonstrate that suggestive questioning does not permanently alter the memory traces children form about an event, which may have crucial implications for re-direct examinations by prosecutors after suggestive cross-examinations in court.

The studies analysed in the current review provide further evidence that children are able to provide complete and accurate testimonies even following very long delays, especially when questioned about stressful or traumatic events (Peterson 2011; 2015). Although both previous research and the present review shows that younger children often provide less complete descriptions of events they experienced, recent studies raise the possibility that multiple non-suggestive interview sessions can help younger children report

more forensically relevant details (Baugerud et al., 2014; Leander, 2010). Further research is needed to investigate the potential interaction between children's age and the delay between interviews found by O'Neill & Zajac (2013).

The present review highlights that while recent studies have attempted to move away from highly controlled laboratory-based cognitive research towards field research and laboratory conditions which resemble the field context more closely, the forensic applicability of this body of evidence is limited by the inconsistent and often contradictory findings reported, partially explained by the wide range of methodologies and measures used by researchers. On the one hand, it is crucial to conduct studies involving a range of ages, delays and questioning methods, as the conditions of real-life forensic interviews are also highly variable. However, results would be easier to interpret if integrated measures of the quality of children's testimonies were developed, encompassing both cognitive concepts such as accuracy and completeness and markers of forensic and legal relevance, such as content and consistency. The development of integrated measures would assist professionals in determining whether multiple interviews contributed positively to children's recall in previous investigations and whether current and future investigations may benefit from the use of multiple interviews. The following section will examine the question of whether the concept of narrative coherence may be of use to researchers and professionals seeking integrated measures of the quality of children's testimony.

1.3 Developing an Integrated Measure of the Quality of Children's Testimony

A comprehensive review of the literature revealed that in contrast to the first wave of studies which focused almost exclusively on the effects of multiple interviews on the accuracy and completeness of children's accounts of staged events, recent research began to develop a more nuanced view of the influence of multiple interviews on children's ability to provide reliable testimonies. Whilst very few field studies have been conducted on multiple

interviews before 2008, recent studies started to further our understanding of the dynamics of real-life multiple interviews, which differ from laboratory contexts in many vital aspects.

Crucially, research began to construct a more integrated conceptualisation of the real-life effects of multiple interviewing by developing multidimensional measures of the completeness of children's testimony and investigating individual differences that may determine whether certain witnesses benefit from multiple interviewing. Early studies measured completeness either by comparing children's accounts against a list of details characterising the event-to-be-remembered or by counting the details mentioned by children, which roughly correspond to the number of clauses in each sentence. However, the usefulness of these measures for field studies is limited; when the researcher does not control the event to be remembered, there is no list of expected details to compare children's accounts to and the number of details mentioned says very little about the contribution of each interview to the investigation in question. This problem is aggravated by researchers' ignorance to "ground truth", which prevents field studies from measuring the accuracy of children's testimony. Recognising the scarcity of measures describing children's real-life testimonies, recent studies increasingly began to categorise details on the basis of their influence on the investigation. Almost all field studies of multiple interviews distinguish between new versus repeated and consistent versus contradictory details, however, newer studies also consider the content of these details and their relevance to the investigation. These new measures help interpret the quality of children's testimonies not only from a psychological, but also from a legal perspective, calling for the development of integrated approaches to studying children's testimony which are both science-based and legally meaningful.

Studying narrative coherence provides researchers with the opportunity to integrate two, often isolated, strands of research on the quality of forensic testimonies; legal research

investigating the credibility of witnesses and cognitive psychological research focused on witness memory. In both threads of research, there has been a growing recognition that narrative coherence is a critical component of children's testimony. From a legal standpoint, the coherence of a narrative is a strong predictor of its credibility (e.g. Pennington & Hastie, 1991) whilst recent studies in cognitive psychology suggest that more coherently organised memories may be recalled in a more detailed and accurate manner (e.g. Peterson, Morris, Baker-Ward, & Flynn, 2014; Wang, Bui & Song, 2015). Westcott & Kynan (2004) pointed out that children's reports in forensic interviews are often lacking in temporal elements of narrative coherence, which might have a negative impact on their credibility in the courtroom.

Studying the life narratives of children, adolescents and adults, Habermas & Silveira (2008) found that young children have difficulties with constructing coherent narratives of autobiographical events. However, the testimonies children are required to provide in forensic contexts differ strongly from life narratives, therefore, it is important to ascertain whether the limitations on the narrative coherence of young children's recall also applies to investigative interviews. Habermas & Silveira (2008) reported that providing participants with a second opportunity to recall their life narratives did not affect the narrative coherence of their accounts, but research on linguistic markers of narrative ability found an interaction between the frequency of linguistic markers and the number of interviews (Knutsson et al., 2011; Salykaly, Talwar, Lindsay, Bala, & Lee, 2013; Thomas, Chen, Gordon, & Thenbrink, 2015) pointing to the potential value of exploring the relationship between narrative coherence and multiple interviews further.

The Legal Concept of Narrative Coherence

The story-telling model of legal decision-making proposed by Pennington & Hastie (1992) suggests that jurors' verdict decisions are based on a three-stage process involving the evaluation of evidence through story construction, the representation of the decision alternatives by learning verdict category attributes and reaching a decision through the classification of the story into the best fitting verdict category. Pennington & Hastie (1992) define a story as a "causal chain of events in which events are connected by causal relationships of necessity and sufficiency". Jurors find a story persuasive if it fulfils the certainty principles of coverage, uniqueness, goodness of fit and coherence (Pennington & Hastie, 1992). As a court case progresses, jurors independently assess their belief in the complainant's story and their belief in the defendant's story before comparing the two and choosing the version of the events they accept as true (Pennington & Hastie, 1992).

The story-telling model of legal decision-making received considerable empirical support (Pennington & Hastie, 1988; 1992; Voss, Wiley, & Sandak, 1999; Wilmott, Boduszek, Debowska, & Woodfield, 2018) and is widely used in the legal and psycho-legal literature (Wilmott et al., 2018; Yale, 2013). Evaluating the applicability of the story construction model of legal decision-making, Pennington & Hastie (1992) found that coherence predicted mock jurors' decisions. Mock jurors were more likely to convict the defendant when the evidence was manipulated to increase temporal coherence and tended to explain their decisions based on actual or inferred causal links in the evidence (Pennington & Hastie, 1988; 1992). Further support for jurors' reliance on story construction was provided by Voss et al. (1999) who showed that jurors guilt ratings were lower when the prosecutor's closing statement was manipulated to either disrupt the temporal order of the narrative or lack explicit causal connections between events, although this effect was only significant when the evidence was ambiguous (Voss & van Dyke, 2001). To overcome the limitations of previous research on the story-telling model that focused almost exclusively on individual juror-level

decision-making rather than collective group decisions, Wilmott et al. (2018) assessed mock jurors' use of certainty principles when deciding the verdict both pre- and post-deliberation. Mock jurors' answers on a scale measuring their use of certainty principles affected their verdict decisions; scores on the complainant believability subscale correlated positively with guilty decisions, whereas scores on the defendant believability subscale showed a negative correlation with guilty verdicts, both before and after the deliberation phase (Wilmott et al., 2018).

From the four certainty principles proposed by Pennington & Hastie (1992), coverage, uniqueness and goodness of fit each concern the story's ability to provide a better explanation of the events in question than other possible stories, whilst coherence is an internal characteristic of each story, regardless of the quality of other evidence. In Pennington & Hastie's model, a story is coherent if it corresponds to jurors' pre-existing knowledge of the state of the world (plausibility), provides a full picture with all expected components (completeness) and is free of logical contradictions (consistency) (Pennington & Hastie, 1992). Alternative explanations of coherence have placed less weight on the component of plausibility and proposed that causal and temporal connectedness is a separate element in determining the persuasiveness of arguments (MacCormick, 2005; Rideout, 2008). For instance, Rideout refers to coherence as a combination of consistency and completeness, with the two contributing equal weight to persuasion: "A story may be internally consistent, and yet remain unconvincing if it is incomplete" (Rideout, 2008, p. 65). In contrast, MacCormick (2005) emphasises the need for coherent narratives to fulfil the principles of "universal causation" and "rational motivation" through causal and temporal linking of details.

To summarise, the story-telling model of legal decision-making suggests that when arguments in court are coherent, meaning that they fulfil the criteria of consistency, plausibility, completeness and causal and temporal connectedness, jurors perceive them as

more credible, and this affects their verdict decisions. Although the model was conceived to explain specifically the decision-making process of individual jurors presented with the full range of evidence associated with a court case, certainty principles may also apply to the perceived credibility of individual witness statements.

Narrative Coherence in Cognitive Psychology

The definition of narrative coherence in cognitive psychology is more constrained than the legal understanding of the concept. While legal scholars often use it as an umbrella term to describe the components of consistency, plausibility, completeness and causal-temporal connectedness, in cognitive psychology, the phrase initially referred specifically to narrative structure or the extent to which details are causally connected in a text (Black & Bern, 1981). However, later research began to distinguish between the concepts of cohesion or the extent to which "event details are presented in a connected form through temporal and causal relations" (Kulkofsky, Wang, & Ceci, 2007) and coherence referring to a higher order structure characterising the entire narrative rather than individual sentences (Peterson et al., 2013). Cohesion is sometimes described as a prerequisite to coherence, however, a linguistically cohesive story can be incoherent if it lacks a global structure and cannot account for why events occurred not only in terms of causal relations but also regarding the motivation of human action through emotive, cognitive and perceptive processes (Reese et al., 2011; Nelson & Fivush, 2004; Peterson et al., 2014, O'Kearney, Speyer, & Kenardy, 2006). McAdams (2006) claims that narrative coherence essentially refers to the characteristics of a story which serve towards overcoming "the problem of being understood in a social context." Whilst "legal coherence" focuses on persuasiveness and credibility, ultimately the two definitions both describe coherence as the characteristic of a story which determines whether communication between the speaker and the listener will be successful.

Models of narrative coherence

Given that narrative coherence occupies a fundamental but not well circumscribed position in storytelling and communication more generally, operationalising the concept for research in cognitive psychology proved a considerable challenge. Early research on narrative coherence consisted of two main strands; story schema models and linguistic approaches (Reese et al., 2011). Peterson and McCabe's (1983) story grammar structure approach defined coherence in terms of sequences and episodes contributing to the completeness of a story. From lowest to highest, complexity sequences included descriptive, action and reactive sequences whilst episodes were categorised as abbreviated, complete, complex or interactive. The story grammar model can be considered as the origin of approaches that describe coherence as a sum of local characteristics or details in the text. In contrast, high point analysis (Peterson & McCabe, 1983) focuses on the global structure of texts, analysing how the story moved from its start to the end point. Both local and global approaches are used in current research on narrative coherence however, the unitary construct of coherence implicit in these early models has been succeeded my multidimensional approaches (e.g. Kulkofsky et al., 2007; Reese et al., 2011).

Multidimensional approaches to narrative coherence

Multidimensional approaches to narrative coherence discern separate dimensions of coherence, all of which contribute to listeners' impression of the global coherence of a story. Habermas and Bluck (2000) distinguished between temporal, thematic and causal dimensions. Temporal coherence refers to the extent to which the chronology of a story is defined, whilst causal coherence signifies cohesion or the extent to which details in a story are connected to form an integrated whole (Habermas & Bluck, 2000). Finally, a story is thematically coherent if there is a clear focus to it with few arbitrary changes of topic

(Habermas & Bluck, 2000). In addition, Habermas and Bluck (2000) included a fourth, autobiographical dimension which describes the extent to which the story includes all elements that listeners expect a life story to include. The Narrative Coherence Coding Scheme, developed by Reese et al. (2011) based on the multidimensional approach of Habermas & Bluck (2000), includes the additional category of contextual coherence. In contrast to chronological coherence, which refers to the extent to which events within the story are organised temporally, contextual coherence refers to the extent to which the story is defined in time and place (Reese et al., 2011). A further dimension of personal meaning, including subjective evaluations, emotion terms or mental state language is frequently used in combination with the Narrative Coherence Coding Scheme (Peterson et al., 2013, O'Kearney et al., 2006; Nelson & Fivush, 2004, Reese et al., 2011).

This multidimensional model of narrative coherence is supported by Reese et al.'s (2011) and Habermas & Silveira's (2004) research on the development of coherence across the lifespan. Reese et al. (2011) found that rather than appearing at one developmental stage, as suggested by a unitary theory of narrative coherence, different aspects of coherence emerge gradually at different points in the lifespan. Whilst the ability to narrate a story in chronological order develops between early and middle childhood, establishing contextual coherence in autobiographical memories remains challenging until adolescence, as it requires advanced perspective-taking skills (Reese et al., 2011). Similarly, studying the life stories of children, adolescents and adults, Habermas and Silveira (2008) reported that temporal coherence improved most between the ages of 8 and 12 and causal coherence between the ages of 12 and 16 but thematic coherence or the ability to maintain a clear focus and convey personal meaning was still developing between the ages of 16 and 20. As children are less able to recount past experiences in a coherent manner than adults (Habermas & Silveira, 2008; Reese et al., 2011), it is essential to explore whether the narrative coherence of

children's recall may be associated with other memory and communication skills required in investigative interviews.

Narrative coherence and memory

Early research on narrative coherence has shown that linguistic cohesion affects the memorability of texts. In Black & Bern's (1981) experiment, participants remembered causally related sentences better than unrelated sentence pairs, both in free recall and cued recognition tests. More recent research indicates that the extent to which participants form causal connections between more complex stimuli also affects the accuracy and completeness of their recall (Chae, Kulkofsky, Debaran, Wang, & Hart, 2016; Kulkofsky et al., 2007; Wang et al., 2015). Studying children's recall of unrelated pictures, Wang et al. (2015) found that children who narrated a coherent story when presented with the pictures recalled them in more detail and with higher accuracy 6 months later than children who made no narrative connections between the pictures. Similarly, Chae et al. (2016) found that the coherence of children's narratives about an event was correlated with the completeness and accuracy of their recall. Using a story grammar approach to narrative coherence, Brown et al. reported that the presence or absence of story grammar elements predicted the accuracy of children's accounts and their resistance to suggestive questions (Brown, Brown, Lewis, & Lamb, 2018).

In contrast to research indicating that narrative coherence predicts the accuracy of recall (Brown et al., 2018; Chae et al., 2016; Wang et al., 2015), Kulkofsky et al. (2007) found that whilst children who provided more coherent narratives also recalled more details of a staged event they participated in, the proportion of accurate information in their narratives decreased as narrative coherence increased. The researchers suggested that this was the result of a trade-off between the length and accuracy of children's recall (Kulkofsky et al., 2007). When the length of narratives was controlled, children who provided more coherent

narratives made a lower proportion of errors than children whose stories were less coherent (Kulkofsky et al., 2007).

Whilst studies investigating the association between the narrative coherence of children's memory and their accuracy and completeness demonstrated that these variables are correlated (Chae et al., 2016; Wang et al., 2015; Kulkofsky et al., 2007), they were not able to ascertain the direction of this relationship – does the coherent organisation of memories lead to better recall or are more complete and accurate memories easier to organise coherently? This problem was addressed by Peterson et al.'s (2014) longitudinal study investigating the survival of childhood memories over a 2-year delay. Using a multidimensional framework of narrative coherence consisting of thematic, chronological and contextual elements, Peterson et al. (2014) found that each dimension contributed to the persistence of memories over time. In addition, events that were described with emotion terms were found to be 2.5 times more likely to be remembered 2 years later than those described without emotion (Peterson et al., 2014). These findings suggest that narrative coherence is a separate dimension of autobiographical memory which contributes to the accuracy and completeness of recall independently.

Emotional experiences and narrative coherence

Child witnesses may struggle to provide coherent narratives of their experiences in forensic interviews not only because the elements of narrative coherence are still emerging at a young age, but also because of the nature of the events they are asked to recall. Early research on the impact of trauma on memory suggested that children who were highly distressed in the emergency room following an injury were less able to produce coherent narratives of their experiences than children who were less upset (Peterson & Biggs, 1998). However, recent studies of children's recall of traumatic experiences indicated that narratives

of emotionally negative events were characterised by a higher frequency of causal and chronological markers of coherence than positive experiences (O'Kearney et al., 2006; Peterson et al., 2014). Similarly, Fivush et al. (2003) found that children growing up in communities exposed to a high extent of violence reported an equal number of details about emotionally positive and negative experiences, but narrated negative experienced more coherently (Fivush, Hazzard, McDermott Sales, Sarfati, & Brown, 2003).

O'Kearney et al. (2006) suggest that the increase in coherence associated with negative emotional experiences might result from children's efforts to find meaning in traumatic events but emphasise that this effect might not apply to memories of trauma that is secret or private in nature, such as intra-familial physical and sexual abuse. It is also important to note that while most children recognise that sexual abuse is wrong while it is happening, victims commonly report that following the incident they initially felt more confused than upset and only interpreted the experience as traumatic retrospectively (Clancy, 2009). Altogether, the current body of literature does not support the idea that forming coherent narratives of traumatic events is more difficult than recalling neutral and positive events coherently (Fivush et al., 2003; O'Kearney et al., 2006; Peterson et al., 2014), however, there are no conclusive findings relating specifically to the case of childhood sexual abuse.

Narrative coherence and multiple interviews

Field studies reviewed in this chapter have shown that multiple interviews increase the completeness of children's testimony, as central and sensitive details emerge in later interviews (Katz & Hershkowitz, 2013; Leander, 2010; Waterhouse et al., 2016). At the same time, multiple interviews tend to decrease the consistency of children's testimony through the omission of previously mentioned details, the addition of new details and potential

contradictions between interviews (Baugerud et al., 2014; Price et al., 2016). However, the impact of multiple interviews on the frequency of causal-temporal connections in children's accounts has not been widely studied, despite the influence of these connections on the credibility (Pennington & Hastie, 1988; 1992; Voss et al., 1999) and accuracy (Chae et al., 2016; Kulkofsky et al., 2008; Wang et al., 2015) of children's recall. Habermas and Silveira (2008) asked participants between the ages of 8 and 20 to narrate their entire life stories with the help of a previously constructed timeline on two separate occasions. The second interview took place two weeks after the first one and was conducted by a different interviewer. Half of the group also participated in a training session intended to improve autobiographical narration while the other group participated in a control task. The authors found that neither repetition, nor the training session affected the relative frequency of local markers of coherence or listeners' ratings of the global coherence of participants' life stories (Habermas & Silveira, 2008).

Habermas & Silveira's (2008) findings are at odds with Thomas et al.'s (2015) research demonstrating that multiple interviews lead to a 'practice effect' in adults' recall that resulted in a decrease in participants' use of verbal markers of uncertainty, such as pauses and filler words. However, the research design of Habermas & Silveira's (2008) study differed from the setup of forensic interviews on a range of crucial issues, which makes it difficult to apply their findings to forensic contexts. Firstly, the aim of the study was to investigate the coherence of life stories rather than memories of a specific event or series of events in participants' lives. In forensic interviews, victims are expected to recall only instances of abuse and other experiences related to the suspect, location and timeline of the abuse. However, they are expected to recall these events in much more detail and emphasis is placed on reporting every detail accurately. This style of painstaking analysis of one's memories might prevent the formation of coherent narratives as victims constantly need to monitor the

accuracy of the memories reported. Due to the cognitive effort associated with the initial recall attempt, subsequent recall occasions might allow victims to focus less on identifying the details they need to recollect and more on piecing them together. This help may be particularly important for children as they are both less aware of the type of communication expected during forensic interviews and less able to provide coherent narratives of events that occurred (Peterson et al., 2013; Reese et al., 2011).

Secondly, in Habermas & Silveira's study, interviewers were instructed to avoid asking questions, therefore narratives were presented entirely in a monologic format. In contrast, forensic interviews are always dialogic, although the extent of interviewers' contribution differs greatly depending on the interview protocol, the skill and experience of the interviewer and children's willingness or unwillingness to communicate. Studies of parent-child reminiscing of the past have shown that the communicative style of conversational partners influences the coherence, as well as the accuracy and completeness of children's recall of past events (e.g. Fivush, Haden, & Reese, 2006; Langley, Coffman & Ornstein, 2017; Larkina & Bauer, 2010). Fivush et al. (2006) found that the children of mothers who use a high-elaborative style during conversation (for instance by asking follow-up questions and encouraging their children to share more) tended to provide more detailed and coherent narratives of events both when talking to their mothers and when producing free narratives in conversation with others. Whilst most forensic interviewing guidelines aim to encourage children to provide as much unaided narrative as they can, the length of these narratives is often very limited, especially for young children. Even when they limit their input to open-ended questions, interviewers still influence the coherence of children's recall by assisting them with chronology ("Tell me what happened next"), causality ("Why did you hide?"), the inclusion of central details ("Where did this happen?"), and references to subjective experiences ("How did you feel?"). In contrast, if interviewers' questions jump

around in time or are endlessly repeated without a clear reason why, constructing coherent narrative may be exceedingly difficult for children. Children who were assisted by coherent questioning in their first interview may begin the second interview with the advantage of having previously built a coherent narrative with the help of the interviewer and manage to recall the memories in a more coherent manner. However, this positive effect would not apply to children questioned in an incoherent manner whose subsequent narratives might be even more disorganised than their initial recall.

Unfortunately, Habermas and Silveira (2008) provided no analysis of how similar the life stories recalled on the two different recall occasions were. Field studies show that up to 58.5% of details mentioned in second and third interviews are new, indicating that children might recall other aspects of events or event entirely different events in these interviews (Katz & Hershkowitz, 2013). Whilst this means that a 'practice effect' resulting from repetition might not be present in multiple forensic interview, children's increased familiarity with the aims and style of a forensic interview and the scaffolding provided by the interviewer may still lead to an increased frequency of causal-temporal connections across interviews. Whilst Habermas and Silveira (2008) found that neither multiple recall occasions, nor explicit attempts at improving participants' ability to construct coherent and connected narratives were effective, the specific circumstances of forensic interviews may lead to a different relationship between repetition and the coherence of children's narratives.

1.4 An Overview of the Present Studies

The aim of the present studies was to explore the impact of multiple interviews on the narrative coherence of children's testimony and to investigate the relationship between elements of narrative coherence and the accuracy and credibility of children's recall. For the purpose of the present research, the 'legal' definition of narrative coherence was adopted, consisting of the completeness, consistency and connectedness of children's testimony.

Completeness referred to the amount and type of details in children's recall, consistency signified the lack of explicit contradictions in testimonies and connectedness referred to the extent to which the details mentioned fit together to form a whole. Plausibility, the additional element included in Pennington & Hastie's (1992) model of coherence, was not analysed, because rather than constituting an internal characteristic of a testimony, it describes the relationship between the testimony and the external world. Completeness, consistency and connectedness were further broken down in each study into multiple variables characterising the quality and forensic relevance of children's testimony. The resulting variables can be mapped onto the dimensions of narrative coherence proposed by previous research, including context, chronology, theme and emotion (Peterson et al., 2014; Reese et al., 2011).

Three studies were conducted, with each consecutive study addressing the methodological limitations of previous research. The results from Study 1 informed the research questions and methods chosen in Study 2, and the findings from both Study 1 and Study 2 were considered when Study 3 was designed. Through using a wide methodological framework, this thesis aimed to evaluate whether narrative coherence can provide an integrated framework for laboratory-based and field research aiming to assist the development of multiple interviewing guidelines.

Study 1. Narrative Coherence in Multiple Forensic Interviews with Child Witnesses Alleging Physical and Sexual Abuse

Study 1 investigated the narrative coherence of children's testimonies in a sample of 58 interviews with 29 children alleging sexual or physical abuse in the United Kingdom and the Republic of Ireland. The influence of multiple interviews on the completeness, consistency and connectedness of children's testimonies was assessed alongside analyses of the relationship between children's age and the quality of their testimony. Previous studies

reported that multiple interviews affect the components of narrative coherence in different directions; the completeness of children's accounts tends to increase as a result of multiple interviews whilst the consistency of their testimony decreases. Habermas & Silveira (2008) found that the coherence of children's life narratives was not affected by repetition, but no prior forensic field studies have investigated the effect of multiple interviews on the connectedness of children's accounts. Therefore, it was expected that multiple interviews will increase the completeness and decrease the consistency of testimonies, whilst analyses regarding linguistic connectedness were exploratory. As research in cognitive psychology has indicated that the level of connectedness of children's memories predicts the completeness of their recall after a delay (Peterson et al., 2013), it is important to establish whether multiple forensic interviews can help children organise and structure their memories more coherently and thereby recall more forensically relevant details about the events. The results of this study could assist investigators in deciding when multiple interviews with child witnesses are appropriate and suggest ways to effectively scaffold the coherence of children's memories.

Study 2. The Impact of Multiple Interviews on the Accuracy and Narrative Coherence of Children's Descriptions of a Staged event

Overcoming the limitations of field research regarding researchers' ignorance of 'ground truth', Study 2 involved an experimental design aiming to establish the relationship between components of narrative coherence and the accuracy of children's recall in multiple interviews. Eighty-three children aged between 8 and 11 years took part in a first aid training session and were interviewed using the NICHD protocol. To discern between the effects of delay and those of multiple interviewing, the study involved a control group of children interviewed 5 weeks after the first aid training session alongside the experimental group interviewed 1 week and 5 weeks after the event. Previous research on the relationship between the narrative coherence and accuracy of children's recall has been inconsistent, with

some studies reporting that coherent memories are recalled more accurately (Chae et al., 2016) while others found that the proportion of accurate details decreases as narrative coherence increases (Kulkofsky et al., 2007). It was expected that the completeness of children's recall would increase when information provided in both the 1-week interview and the 5-week interview of the experimental group would be considered, whilst their accuracy would decrease. Taking the view that the decrease in accuracy is the result of delay rather than multiple interviewing per se, it was expected that the accuracy of 5-week interviews would be the same regardless of whether children were in the experimental group interviewed for the second time or in the control group interviewed for the first time. As forensic interviews aim to reveal the truth, the benefits of multiple forensic interviews in increasing the narrative coherence of children's testimony only confer advantages in real-life investigations if they have no detrimental impact on the accuracy of recall.

Study 3. The Impact of Narrative Coherence and Multiple Interviews on Mock Jurors' Ratings of the Credibility of Child Witnesses

Testing the story-telling model's implication that narrative coherence is a key predictor of witness credibility (Pennington & Hastie, 1992) Study 3 investigated the impact of multiple interviews and narrative coherence on 106 mock jurors' impressions of the reliability, competence and credibility of the witness in a fictional child sexual abuse case and their verdict decisions using a between-subjects design. In addition to contrasting jurors' impressions of a testimony elicited in a single interview with an account provided over the course of two interviews, the study also assessed whether the perceived age of the child witness and the consistency of the testimony may influence the impact of multiple interviews on credibility. It was expected that mock jurors would rate the witness as more reliable, competent and credible when presented with two interviews, rather than a single interview due to the increase in the completeness of children's testimony, and this difference would

also be reflected in their verdict decision. Additionally, consistent interviews and younger witness age were also expected to lead to higher ratings and verdict decisions favouring the alleged victim. The results of Study 3 have crucial implications with regards to whether the benefits of multiple interviews found in Study 1 and Study 2 translate to positive courtroom outcomes for alleged victims of child sexual abuse.

Chapter 2. Narrative Coherence in Multiple Forensic Interviews with Child Witnesses Alleging Physical and Sexual Abuse

2.1 Introduction

The story-telling model of legal decision-making (Pennington & Hastie, 1992) suggests that the narrative coherence of witnesses' accounts plays a crucial role in forensic investigations, as coherent narratives, defined in legal terms as stories which are complete, consistent, and causally and temporally connected, are more credible than stories lacking in coherence (McAdams, 2006; Pennington & Hastie, 1988; Rideout, 2008; Voss et al., 1999). Although the story-telling model was initially conceived to describe the decision-making processes of jurors during a trial, the model has implications for credibility judgments during police investigations too, as cases judged non-credible at this stage never reach a courtroom. Bennett & Feldman (1981) suggest that the narrative coherence of witnesses' stories is especially critical in situations where credibility judgements need to be made in the absence of corroborating evidence, which is often the case for child abuse investigations.

Children's ability to provide coherent narratives in forensic interviews may be limited by several factors. Firstly, Reese et al.'s (2011) research on the narrative coherence of children's autobiographical accounts suggests that although young children can give truthful and detailed accounts of events that happened to them, the ability to provide chronologically, contextually and thematically coherent narratives does not fully develop until late childhood or adolescence. Constructing a consistent timeline of events may be particularly difficult for young children due to their difficulty with understanding and using temporal concepts (Graffam Walker, Kenniston, Inada, & Caldwell, 2013), especially in cases where they need to give accounts of repeated experiences (Brubacher, Powell, & Roberts, 2014). Despite the

central role of storytelling in forensic investigations (Pennington & Hastie, 1992) and children's difficulty with recounting their past experiences in a coherent manner (Reese et al., 2011), few studies applied a story-telling framework to evaluate the narrative quality of children's accounts elicited in forensic interviews.

Models of Narrative Coherence in Developmental Psychology

In contrast to the story-telling model's definition of narrative coherence as an umbrella term comprising of the completeness, consistency and connectedness of arguments (McAdams, 2005; Pennington & Hastie, 1992; Rideout, 2008), psychology studies generally use a narrower concept of narrative coherence, broadly defined as a framework for organising event details (Feltis, Powell & Roberts, 2011). In this form, narrative coherence is described as independent from other measures of the quality of children's stories, including accuracy, consistency and descriptive detail (Brown, et al., 2018; Reese et al., 2011).

Recent laboratory and field studies analysing the narrative coherence of children's accounts of real-life experiences fall into three broad categories. Story-grammar approaches categorise the information provided by the speaker into 6 grammar elements, including the setting of the story, the initiating event, the character's internal responses, the attempt or action, the consequence of the action and the character's reaction. (Feltis, Powell, Snow, & Hughes-Scholes, 2010; Feltis et al., 2011; Westcott & Kynan, 2004). In contrast, multidimensional models rely on the broader categories of context, chronology, theme and evaluation to measure the narrative coherence of autobiographical memories (Habermas & Bluck, 2000; Habermas & Silveira, 2008; Peterson et al., 2014; Reese et al., 2011; Morris, Baker-Ward, & Bauer, 2010; Nelson & Fivush, 2004). Finally, the narrative cohesion approach measures the 'degree to which event details are presented in a connected form through temporal and causal relations' (Kulkofsky et al., 2008, p. 23).

Whilst multidimensional models, story schema approaches and the narrative cohesion approach are distinct from one another, elements of their conceptualisation of narrative coherence overlap (Brown et al., 2018). For instance, Brown et al. (2018) mapped the multidimensional model's elements of chronology, context and evaluation onto the temporal element of setting, the physical and social elements of setting and the reaction and internal response of characters, respectively, in the story schema approach. Similarly, temporal markers in the narrative cohesion approach measure the same aspect of recall as the element of temporal setting in the story schema model and the dimension of chronology in multidimensional models.

Narrative Coherence in Forensic Interviews with Children

Only the story grammar approach has been applied to measure the narrative coherence of forensic interviews with child witnesses (Feltis et al., 2010; Westcott & Kynan, 2004). Analysing *Memorandum of Good Practice* interviews with children alleging sexual abuse in the United Kingdom, Westcott and Kynan (2004) found that although children's accounts did conform to a basic story structure, the scarcity of details and the lack of a causal-temporal structure in young children's testimonies limited the coherence of their narratives. A third of the cases examined was rated as 'disordered' in terms of causal and temporal relations and the extent of disorder was especially high for pre-schoolers (Westcott & Kynan, 2004). Another aspect of narrative coherence which children of all ages struggled with was describing 'subjective details', their personal reactions to the abuse. Only 10% of children spontaneously described their physical perceptions and 20% provided descriptions of their emotional state. However, this proportion was increased to 33% and 46.6%, respectively, when interviewers asked questions focused on subjective details. Noting the low proportion of open-ended questions in their sample, Westcott and Kynan (2004) suggested that the use

of overly specific questions compromised the narrative coherence, especially the temporal and causal organisation of children's accounts.

Westcott & Kynan's (2004) results regarding the impact of overly specific questioning on the narrative coherence of children's testimony are consistent with previous research demonstrating the superiority of open-ended questions over specific questions in eliciting details of children's allegations in forensic interviews (Lamb et al., 2003). Feltis et al.'s (2010) study of police interviews with 5- to 15-year-old children conducted along the lines of the NICHD protocol in Australia provided further support for the effect of question type on the coherence of children's testimony. Open-ended questions were more effective than closed-ended questions in eliciting story grammar elements in children's testimony, across all age groups and interview phases (Feltis et al., 2010). However, not all open-ended questions were equally beneficial. Broad open-ended ('Tell me everything that happened') yielded the most story grammar elements, followed by chronological questions ('What happened next?'), whilst open-ended questions asking the child to elaborate or clarify a previously mentioned detail were less effective.

In addition to the impact of question type on the frequency of story grammar elements in children's testimony (Feltis et al., 2010), laboratory research suggests that children's age and the characteristics of the event they are questioned about also affects the narrative coherence of recall (Feltis et al., 2011; Reese et al., 2011). Relying on the multidimensional model of narrative coherence, Reese et al. (2011) found that even pre-schoolers were able to maintain a topic of conversation, but they found it difficult to place events in time and place and organise them according to chronology. Similarly, Habermas & Silveira (2008) found that temporal coherence increased dramatically between the ages of 8 and 12, causal coherence increased most between 12-year-olds and 16-year-olds and thematic coherence was still in development between 16 and 20 years. Using a story grammar approach, Feltis et

al. (2011) reported that the frequency of schema elements in children's accounts correlated positively with age. In addition, the children in Feltis et al.'s (2011) sample also recalled more story grammar elements when the delay between the interview and the events children were asked to recall was shorter, and when children experienced repeated occurrences of the same event rather than a single event.

Overall, research on children's ability to provide coherent narratives in forensic interviews suggested that children struggle with providing a causal-temporal framework and describing their subjective reactions to the events they are questioned about (Westcott & Kynan, 2004). Whilst children's age and the type of abuse they experienced may limit their ability to form coherent narratives (Feltis et al., 2011; Reese et al., 2011), interviewers can support witnesses in providing coherent accounts through using open-ended questions (Feltis et al., 2010).

Multiple Forensic Interviews with Child Witnesses

Many children interviewed for legal purposes in the United Kingdom are questioned more than once (Plotnikoff & Woolfson, 2001; Waterhouse, Ridley, Bull, La Rooy, & Wilcock, 2016), but the narrative coherence of children's testimonies provided over the course of multiple interviews has not yet been investigated. Whilst the impact of multiple interviews on the completeness and consistency of children's accounts has been the subject of both field research (Hershkowitz & Terner, 2007; Katz & Hershkowitz, 2013; Leander, 2010; Waterhouse et al., 2016) and laboratory studies (e.g. Dietze et al., 2013; Knutsson et al., 2011; Peterson, 2011; 2015; Price et al., 2016), the third component of the 'story-telling' model of coherence, the extent to which details are connected through causal and temporal links, is yet to be explored.

Multiple interviews generally increase the completeness of children's testimony as repeated recall occasions allow witnesses to recall new information (La Rooy et al., 2009). In laboratory studies, the completeness of children's recall in a second interview has sometimes exceeded the number of details reported in the first interview (e.g. Knutsson et al., 2011; La Rooy et al., 2005). However, even when the number of details reported in each successive interview declines, the overall amount of information reported in the interviews may increase through reminiscence. Consistent with laboratory findings indicating that reminiscence leads to an increase in completeness over multiple interviews, several field studies found that a significant amount of forensically relevant information was recalled in subsequent interviews but absent from the first interview (Hershkowitz & Terner, 2007; Katz & Hershkowitz, 2013; Leander, 2010; Waterhouse, 2016).

As multiple interviews lead to the recall of new information, they inherently decrease the consistency of children's accounts through the addition of new details and the omission of previously mentioned details in later interviews. Examining cross-examination techniques in Scottish courts, Szojka et al. found that defence lawyers often challenged the credibility of children's statements on the basis of inconsistencies in their accounts (Szojka, Andrews, Lamb, Stolzenberg, & Lyon, 2017), despite laboratory research showing a lack of correlation between the consistency and overall accuracy of witness statements elicited through multiple interviews (Baugerud et al., 2014; Gilbert & Fisher 2006). Defence lawyers most commonly referred to contradictions between details mentioned by the witness, rather than additions of new details or omissions of previously mentioned details (Szojka et al., 2017). Previous research on the frequency of contradictions in children's accounts elicited over multiple interviews have been inconclusive; while Katz & Hershkowitz (2013) found no contradictory details within or across interviews, Waterhouse et al. (2016) reported a low number of forensically relevant contradictions. Contrasting results may be partially accounted for by the

lower frequency of open-ended questions in Waterhouse et al.'s (2016) sample compared with the NICHD interviews analysed by Katz & Hershkowitz (2013).

Although the connectedness of children's accounts elicited over the course of multiple forensic interviews has not yet been addressed, Habermas & Silveira's (2008) study analysed the impact of repeated recall on the extent to which events in children's and adults' life narratives were causally and temporally connected. Comparing the thematic, causal and temporal coherence of 8-year-olds, 12-year-olds, 16-year-olds and 20-year-olds in two 15-minute interviews conducted two weeks apart, Habermas & Silveira (2008) found that multiple interviews had no effect on temporal and causal elements of narrative coherence.

However, the extent to which Habermas & Silveira's results can be interpreted in a forensic context are limited. First, forensic interviews are based on a question-answer format, and lengthy uninterrupted narratives are not frequent in interviews with children. The question-answer format might modify the effect of multiple interviews on narrative coherence – it can either serve as a scaffold, as suggested by Bennett and Feldman (1981), or have a disjointing effect on children's accounts, when the questions asked are overly specific (Westcott & Kynan, 2004). Secondly, children are asked to recall events in much more detail during forensic interviews than when questioned about their life stories, which may allow a larger role for reminiscence in later interviews. Third, the youngest children in Habermas & Silveira's (2008) sample were 8-years old, but the testimonies of younger children may be particularly lacking in causal and temporal connections due to their limited ability to organise their recall according to context and chronology (Reese et al., 2011)

The Present Study

The purpose of this study was to investigate the extent to which children construct coherent narratives in multiple forensic interviews and explore the impact of multiple

interviews on each of the three components of narrative coherence, as defined by the story-telling model (McAdams, 2005; Pennington & Hastie, 1992; Rideout, 2008); completeness, consistency and connectedness. 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, the overall amount of information available to the interviewers after the two interviews was compared with the amount of information available after the first interview.

The completeness of children's testimony was measured through the amount of details children reported, the forensic relevance and content of these details, and the frequency of subjective details, including emotions, cognitions and perceptions. To determine the consistency of children's accounts, new and repeated information was discerned both within the same interview and between the two interviews and the frequency of contradictions was calculated in each interview. Finally, connectedness was measured using a modified version of the Narrative Cohesion Coding Scheme developed by Kulkofsky et al. (2007). Following research showing that open-ended questions increased the narrative coherence of children's testimony (Feltis et al., 2010), interviewers' questions were categorised according to their narrative coherence.

Although a new coding scheme was developed to conform more closely to the legal definition of narrative coherence and to allow for the analysis of very long narratives elicited over the course of multiple interviews, the components of narrative coherence in this study can be mapped onto multidimensional models of coherence (Reese et al., 2011). Specifically, the content types of time and location indicate context, the proportion of sensitive information conveys a measure of theme, subjective details measure evaluative content and chronology is indexed by the frequency of temporal markers.

Hypotheses

Previous field research strongly suggests that multiple interviews increase the completeness of children's testimonies (Hershkowitz & Terner, 2007; Katz & Hershkowitz, 2013; Leander, 2010; Waterhouse et al., 2016). Therefore, it was expected that 1) multiple interviews will increase the overall number of details, 2) multiple interviews will increase the number of time and location details, 3) multiple interviews will increase the number of subjective details and 4) multiple interviews will increase the number of sensitive details. Based on the results of previous studies investigating the impact of repeated recall on the consistency of witnesses' accounts (Baugerud et al., 2014; Gilbert & Fisher, 2006, Krix et al., 2015), it was expected that children's testimonies will become less consistent as a result of multiple interviews.

Due to the lack of previous research on the connectedness of children's testimonies elicited over multiple forensic interviews, analyses regarding this aspect of narrative coherence were exploratory and no specific hypotheses were established. Scaffolding provided by interviewers and the 'practice effect' associated with repeated recall may facilitate children's recall, leading to a higher frequency of markers of causal and temporal connectedness in multiple interviews. However, it is also possible that children have difficulties with integrating details mentioned over different recall occasions on the same causal-temporal scale. Alternatively, Habermas & Silveira's (2008) finding that the frequency of causal-temporal connections in children's accounts is not affected by multiple recall occasions may be replicated in forensic interviews.

The relationship between the types of questions asked by interviewers and the connectedness children's responses was also explored. Previous research has shown that open-ended questions result in longer narratives, however, no studies have measured the

extent to which interviewers follow a coherent questioning strategy which allows children to build causal and temporal connections within their narratives. It was expected that more coherent questioning strategies will result in more connected responses.

2.2 Method

Sample

Transcriptions of 56 forensic interviews with 28 alleged victims of child physical and sexual abuse aged between 3 and 14 years (*M* = 7.83, *SD* = 3) obtained from an existing dataset, were analysed. Interviews were conducted and transcribed for the purpose of police investigations. Permission to conduct the study was granted by the Royal Holloway, University of London ethics committee ahead of the start of data collection. Children were interviewed between 2 and 11 times, however, only the first and second interviews with each child were examined. The delay between interviews ranged from less than an hour to over a year and a half (*M* = 111.52 days, *SD* = 159.49). The interviews took place in Scotland (*N* = 17), England (*N* = 6) and the Republic of Ireland (*N* = 5) between 2001 and 2014. The majority of children were interviewed twice by the same interviewer (*N* = 16) whilst some were interviewed by two different interviewers (*N* = 12). The witnesses interviewed included females (*N* = 15) and males (*N* = 13) alleging physical abuse (*N* = 5), sexual abuse (*N* = 17) or both physical and sexual abuse (*N* = 6). Most witnesses described multiple instances of abuse (*N* = 23). Cases involved a single victim (*N* = 12), two co-victims (*N* = 10) or three co-victims (*N* = 6).

Coding

Only the substantive part of the interviews was analysed, defined as questions asked after the interviewer initially transitioned to probes related to the context of the allegations. Interviewers' utterances were coded according to their coherence. Codes for children's

responses were divided into three main categories: completeness, consistency and connectedness codes. Completeness codes included the number of details in each utterance, their content, their subjectivity and their relevance to the investigations. To measure consistency, each detail was coded according to their novelty and new details were coded as either consistent or inconsistent with previously reported details. Connectedness was coded by identifying local markers of linguistic coherence.

Question coherence

In order to explore the relationship between the narrative coherence of interviewers' questions and the responses elicited, a novel coding scheme was developed for question coherence.

1. *Logically coherent*. Questions that are connected logically to the child's previous utterance. These questions often direct the child's attention to aspects of an event that were not previously mentioned but are related to the details currently discussed.
2. *Chronologically coherent*. Questions that are chronologically related to the child's previous utterance, enquiring about what happened before or after an event mentioned by the child.
3. *Neutral*. Questions that have no explicit connection the previous statement made by the child or the interviewer but do not actively disrupt a coherent questioning strategy. This category includes questions that prompt children to disclose more details about topics that were discussed previously but without explicitly transitioning from the current topic to an earlier topic.

4. *Incoherent*. Questions that are contrary to a coherent questioning strategy, for example because they ignore children's previous utterances, jump around time or excessively repeat the content of previous questions.

5. *Child-led.* Questions that follow children's lead without attempting to refocus their attention. These questions often prompt children to elaborate on their previous utterance.

Number of details

The number of details, approximately corresponding to the number of clauses, was counted in each of children's utterances. Both independent clauses and subordinate clauses were coded as separate details.

Content

Children's responses were coded according to the type of information they referred to. Content codes were not exclusive, however, multiple mentions of the same detail type in the same utterance were only coded once.

1. *Victim*. Mentions of the child himself or herself.
2. *Suspect*. Mentions of the alleged perpetrator of the abuse.
3. *Co-victim.* Mentions of another alleged victim of the same perpetrator.
4. *Witness.* Mentions of other persons.
5. *Location.* Mentions of places or of the positioning of objects or people.
6. *Time.* Mentions of specific dates or times.
7. *Action.* Mentions of actions, including verbal statements.
8. *Object.* Mentions of objects and pets.

Table 3 *Examples of the Content Codes Assigned to Children's Responses*

Example	Content codes
"On Saturday, my dad bought me a new bicycle."	Time, suspect, victim, action, object
"My mum shouted at Mr P."	Witness, action, suspect
"The living room had a large sofa since last Christmas."	Location, object, time

Subjectivity

Details referring to subjective, personal descriptions of the event from the point of view of the victim or other persons were identified and categorised into one of three categories.

1. *Perceptions.* Mentions of visual, auditory or tactile experiences, including pain.
2. *Emotions.* Mentions of affective processes, such as embarrassment, happiness or anger. Expressions of preferences ("I liked / did not like it.") were included in this category.
3. *Cognitions.* Mentions of thoughts, aims and beliefs.

Relevance to the investigation

Children's responses were coded according to whether they were sensitive, central or peripheral from the point of view of the investigators. *Sensitive* content was defined as details describing the time and place of the abuse and details related to the immediate lead up to the abuse, the content of the abuse, or the immediate aftermath of the abuse. *Central* content was

defined as details directly related to the location of the abuse, the time scale of the abuse, the alleged perpetrator of the abuse, co-victims and witnesses of the abuse and essential family relationships without referring to specific instances of abuse. *Peripheral* content was defined as focused on details that were only indirectly related to the events that the interview focused on, such as non-plot related descriptions of time, place, people or events.

Novelty

Each detail mentioned by children was coded according to whether they were *new* or *repeated.* Repeated details were also categorised according to whether they were *repeated within the same interview* or *repeated across interviews.* Details could be coded as repeated both within the same interview and across interviews.

Consistency

Each repeated detail was coded either as *consistent* or *inconsistent* with previous mentions of the same detail. Inconsistent details were also coded based on whether they contradicted details mentioned *within the same interview* or *across interviews.* Details could be coded as contradictory both with details within the same interview or across interviews. If a detail was once mentioned inconsistently, all future mentions were coded as inconsistent, unless an explanation was provided to resolve the contradiction.

Connectedness

Local markers of linguistic coherence were identified using a modified version of the coding scheme developed by Kulkofsky et al. (2007). Marker types were not exclusive, however, multiple mentions of the same marker type in the same utterance were only coded once.

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.

Table 4 *Examples of Linguistic Markers of Connectedness*

Type of marker	Examples
Simple temporal markers	First, next, then, before, after, etc.
Complex temporal markers	When, until, while, etc.
Markers of causal relations	Because, so, in order to, as, etc.
Markers of optional states	Sometimes, usually, always, probably, etc.

Inter-rater reliability

A random selection of 20% of the transcripts (*N* = 12) were independently re-coded. Cohen's κ was used to assess agreement between coders. Reliability was very good for question type, κ = .90, *SE* = .01, 95% *CI* [.88, .92], question coherence, κ = .85, *SE* = .01, 95% *CI* [.83, .87], relevance κ = .89, *SE* = .01, 95% *CI* [.87, .91], novelty κ = .92, *SE* = .02, 95% *CI* [.88, .96], consistency, κ = .91, *SE* = .02, 95% *CI* [.87, .94] and markers of connectedness, κ = .95, *SE* = .01, 95% *CI* [.93, .97]. Reliability was substantial for content, κ = .77, *SE* = .04, 95% *CI* [.69, .85] and subjective details κ = .77, *SE* = .04, 95% *CI* [.69, .85].

2.3 Results

Preliminary analyses and excluded variables

Initial analyses using mixed ANOVAS showed that measures of the completeness, consistency and coherence of children's testimony (the number of details reported, the consistency of details, the relevance of details for the investigation, the content of details and the frequency of markers of connectedness) were not associated with the gender of the witness the identity of the interviewer, the delay between interviews, the frequency of the abuse and the type of abuse reported. Therefore, these variables were excluded from further analyses.

Children's age was normally distributed (Kolmogorov-Smirnov: $D(28) = .13$, $p = .20$) and included in the analyses as a continuous variable. Children's age was added as a covariate to analyses where previous research suggested a relationship between age and the dependent variable, and the scatterplot also suggested a linear relationship. Although ANCOVAs are often used by researchers to statistically control for the effects a confounding variable on the dependent variable (Schneider, Reich, & Mozuraitis, 2015), in the current study, the aim was to assess potential interactions between the effects of the covariate and the dependent variable. Therefore, an interaction term between age and the dependent variables was included in each analysis. In within-subjects designs, the covariate needs to be centred by subtracting the mean of the covariate from each covariate value to avoid an increase in Type 1 error rates or a loss of power (Schneider et al., 2015). Therefore, children's age was centred when used as a covariate in within-subjects analyses. As the probability of Type 1 errors is elevated in within-subject designs involving an interaction between the covariate and the dependent variable (Schneider et al., 2015), alternative statistical tests were conducted following each ANCOVA, without the covariate. These confirmed the main effect in each analysis (Appendix A).

Completeness

On average, witnesses reported 121.75 (*SD* = 147.49) new details in the first interview and 99.29 (*SD*= 123.14) new details in the second. Table 5 contains the number of details children reported in each content category and the proportion of those details from all new details (the same details could belong to multiple content categories).

Overall, number of details correlated positively with children's age in both interview 1 and interview 2. As previous research has found strong associations between children's age and the completeness of their accounts and the scatterplot also suggested a linear relationship between age and the number of details children reported, age was added as a covariate to the analysis (Figure 1).

Table 5 *The Number of Details Reported in Each Content Category in Interview 1 and Interview 2*

	Interview 1		Interview 2	
	M	*SD*	*M*	*SD*
New details	**121.75**	**147.49**	**99.29**	**123.14**
Victim	49.04	57.86	46.32	53.76
Suspect	45.39	52.11	37.11	40.52
Witness	26	42.3	23.47	45.9
Co-victim	12.11	18.88	6.68	11.26
Action	62.07	80.13	54.86	74.62
Object	13.14	19.31	11.43	18.57
Location	24.14	41.02	15.57	19.23
Time	9.14	16.68	7.57	10.19
Subjectivity	**16.71**	**20.07**	**16.71**	**20.07**
Cognition	2.75	3.79	2.71	4.08
Perception	3.82	5.14	4.57	6.44
Emotion	10.14	11.14	9.43	13.62

Note: The number of details in each content category does not add up to the total number of new details, as a single detail could belong to multiple content categories

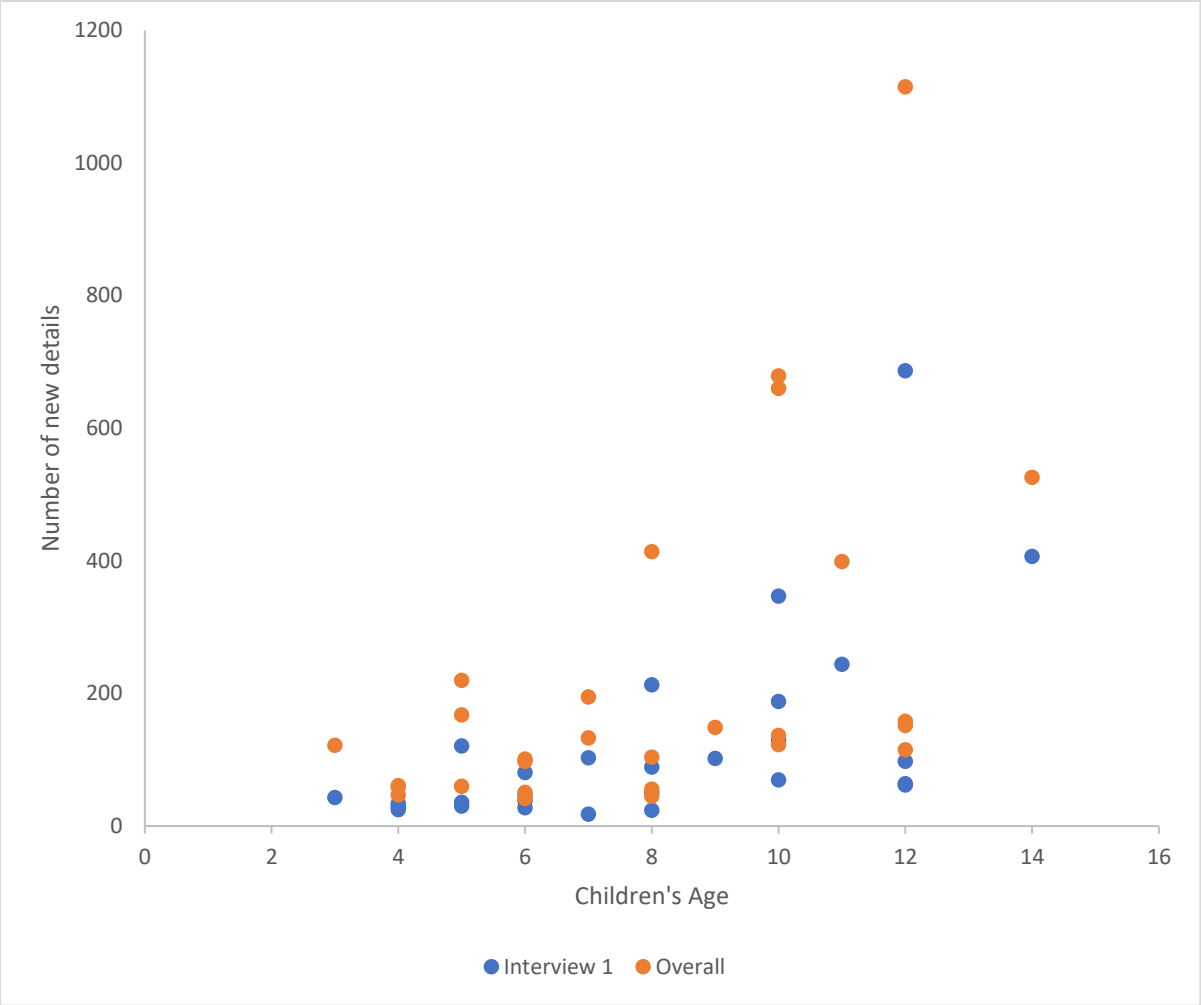


Figure 1 *Children's Age and the Number of New Details in Interview 1 and Overall*

Number of details

To assess whether investigators were aware of significantly more details after the second interview than after the first one, a RM-ANCOVA was conducted assessing the potential effect of interview number (first interview, overall) on the number of details witnesses reported. Children's age added to the analysis as a covariate. Significantly more details were reported overall than in the first interview only, $F(1,26) = 20.44$, $p < .001$, $\eta_p^2 =$

.44. Older children reported significantly more details than younger children, $F(1, 26) = 11.74$, $p = .002$, $\eta_p^2 = .31$. There was a significant interaction between interview number and children's age, $F(1, 26) = 4.32$, $p = .048$, $\eta_p^2 = .14$. Inspection of Figure 1 suggests that the difference between the number of details reported in the first interview and the number of details increased as the age of witnesses increased.

Content

Most of the new details mentioned focused on the victim, the suspect, actions, witnesses and locations. Mentions of objects, co-victims and time were less frequent in both interviews, as were subjective details, including emotions, perceptions and cognitions. Children reported less new details in each content category in the second interview, but the proportional frequency of each type of detail remained similar.

To assess whether interviewers were aware of significantly more details in each content category after the second interview than the first interview, a two-way RM ANOVA was conducted investigating the potential effect of interview number (first interview, overall) on the number of details reported in each content category (suspect, victim, witness, co-victim, action, object, location, time). Mauchley's test indicated that the assumption of sphericity was violated for content, $\chi^2(27) = 378.95$, $p <. 001$, and the interaction between content and interview number, $\chi^2(27) = 323.33$, $p < .001$. Consequently, Greenhouse-Geisser corrections were applied ($\varepsilon = .20$, $\varepsilon = .23$, respectively). Children reported significantly more details overall than in the first interview only, $F(1, 27) = 18.17$, $p < .001$, $\eta_p^2 = .40$. There was also a significant main effect of content, $F(1.39, 37.39) = 23.45$, $p < .001$, $\eta_p^2 = .40$, and a significant interaction between interview number and content, $F(1.62, 43.68) = 12.77$, $p < .001$, $\eta_p^2 = .32$, which were further explored using post-hoc tests.

Pairwise comparisons using the Bonferroni adjustment (adjusted alpha levels $p < .001$) for multiple comparisons revealed that significantly more details were reported related

to actions than related to objects, witnesses, co-victims, locations and time. Significantly more details were reported related to the victim than related to objects, witnesses, co-victims, locations and time. The suspect was mentioned on significantly more occasions than objects, witnesses, locations, co-victims or time. Significantly more details were reported related to location than temporal details. All pairwise comparisons are presented in Table 6.

The two-way interaction between interview number and content was followed up with 8 paired samples t-tests comparing the number of details reported in each category in interview 1 and overall (adjusted alpha level $p < .006$). Significantly more details were reported overall than in the first interview about the suspect, the victim, co-victims, actions objects, locations and time but not about witnesses. All pairwise comparisons are presented in Table 6.

Table 6 *Summary of Analyses of the Content of Children's Responses in Interview 1 and Overall*

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Main effect	Interview	1,26	F =18.17	<.001*	η_p^2 = .40	
Main effect	Content	1.38,37.39	F =18.16	<.001*	η_p^2 = .40	
Interaction	Interview X Content	1.62,43.68	F =12.77	<.001*	η_p^2 = .32	
Follow up	Action vs. Object	27	t = 3.98	=.012	d = .89	[44.68,140.03]
Follow up	Action vs. Victim	27	t = 2.09	=.05	d = .17	[.44,42.71]
Follow up	Action vs. Suspect	27	t = 2.60	=.02	d = .30	[7.23,61.63]
Follow up	Action vs. Witness	27	t = 5.45	<.001**	d = .62	[42.06,92.87]
Follow up	Action vs. Co-victim	27	t = 4.23	<.001**	d = .95	[50.52,145.76]
Follow up	Action vs. Location	27	t = 4.10	<.001**	d = .70	[38.11,114.31]
Follow up	Action vs. Time	27	t = 4.35	<.001**	d = .97	[52.95,147.82]
Follow up	Victim vs. Object	27	t = 4.91	<.001**	d = .94	[41.18,100.39]
Follow up	Suspect vs. Object	27	t = 5.40	<.001**	d = .94	[35.91,79.95]
Follow up	Witness vs. Object	27	t = 1.80	=.08	d = .38	[-3.47,53.26]
Follow up	Co-victim vs. Object	27	t = -.88	=.39	d = .19	[-19.28,7.71]

(continued)

(continued)

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Follow up	Object vs. Location	27	t = -1.98	=.06	d = .35	[-32.84,.56]
Follow up	Object vs. Time	27	t = 1.68	=.11	d = .26	[-1.74,17.46]
Follow up	Suspect vs. Victim	27	t = -2.22	=.04	d = .10	[-24.73,-.98]
Follow up	Victim vs. Witness	27	t = 5.70	<.001**	d = .49	[29.38,62.40]
Follow up	Victim vs. Co-victim	27	t = 4.84	<.001**	d = 1.04	[44.13,109.01]
Follow up	Victim vs. Location	27	t = 5.24	<.001**	d = .67	[33.26,76,03]
Follow up	Victim vs. Time	27	t = 5.27	<.001**	d = 1.07	[48.05,109.24]
Follow up	Suspect vs. Witness	27	t = 4.00	<.001**	d = .39	[16.07,50.00]
Follow up	Suspect vs. Co-victim	27	t = 5.30	<.001**	d = 1.06	[39.07,88.36]
Follow up	Suspect vs. Location	27	t = 4.90	<.001**	d = .60	[24.30,59.27]
Follow up	Suspect vs. Time	27	t = 5.88	<.001**	d = 1.10	[42.84,88.73]
Follow up	Witness vs. Co-victim	27	t = 2.49	=.02	d = .48	[5.47,55.89]
Follow up	Witness vs. Location	27	t = 1.13	=.27	d = .12	[-7.09,24.59]
Follow up	Witness vs. Time	27	t = 2.65	=.01	d = .52	[7.38,58.12]
Follow up	Co-victim vs. Location	27	t = -3.12	=.004	d = .50	[-36.35,-7.51]

(continued)

(continued)

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Follow up	Co-victim vs. Time	27	t = .65	=.52	d = .08	[-4.46,8.60]
Follow up	Location vs. Time	27	t = 3.78	<.001**	d = .55	[10.97,37.03]
Follow up	Action 1 vs. Action All	27	t = 3.89	=.001***	d = .47	[25.92,83.79]
Follow up	Object 1 vs. Object All	27	t = 3.26	=.003***	d = .45	[4.23,18.63]
Follow up	Victim 1 vs. Victim All	27	t = 4.56	<.001***	d = .56	[25.47,67.17]
Follow up	Suspect 1 vs. Suspect All	27	t = 4.85	<.001***	d = .55	[21.40,52.82]
Follow up	Witness 1 vs. Witness All	27	t = 2.71	=.01	d = .35	[5.66,41.26]
Follow up	Co-victim 1 vs. Co-victim All	27	t = 3.14	=.004***	d = .29	[2.31,11.04]
Follow up	Location 1 vs. Location All	27	t = 4.56	<.001***	d = .34	[9.11,24.03]
Follow up	Time 1 vs. Time All	27	t = 3.93	=.001***	d = .35	[3.62,11.52]

Note: * denotes significance at $p < .05$, ** denotes significance at adjusted alpha level $p < .001$, *** denotes significance at adjusted alpha level $p < .006$

Subjectivity

On average, witnesses reported 16.71 new subjective details in both interview 1 and interview 2 (Table 1). In both interviews, emotions were the most frequently mentioned subjective details, followed by perceptions and cognitions.

To assess whether interviewers gathered significantly more subjective details altogether in the two interviews than from the first interview only, a two-way RM ANOVA was conducted investigating the potential effect of interview number (first interview, overall) on the number of details reported in each subjective content category (emotion, cognition, perception). Mauchley's test indicated that the assumption of sphericity was violated for subjectivity, $\chi^2(2) = 31.57$, $p <. 001$, and the interaction between subjectivity and interview number, $\chi^2(2) = 20.6$, $p < .001$. Consequently, Greenhouse-Geisser corrections were applied ($\varepsilon = .59$, $\varepsilon = .65$, respectively). Children reported significantly more subjective details overall than in the first interview only, $F(1, 27) = 15.84$, $p < .001$, $\eta_p^2 = .37$. There was also a significant main effect of subjectivity, $F(1.17, 31.71) = 12.70$, $p = .001$, $\eta_p^2 = .32$, and a significant interaction between interview number and subjectivity, $F(1.29, 34.9) = 8.51$, $p = .003$, $\eta_p^2 = .24$, which were further explored using post-hoc tests.

Pairwise comparisons using the Bonferroni adjustment (adjusted alpha levels $p < .017$) for multiple comparisons revealed that significantly more subjective details focused on emotions than on perceptions or cognitions. All pairwise comparisons are presented in Table 7.

The two-way interaction between interview number and subjectivity was followed up with 3 paired samples t-tests comparing the number of details reported in each subjective content category in interview 1 and overall (adjusted alpha level $p <.017$). Significantly more

details were reported overall than in the first interview about emotions, perceptions and cognitions. All pairwise comparisons are presented in Table 7.

Table 7 *Summary of Analyses of the Subjective Content of Children's Responses in Interview 1 and Overall*

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Main effect	Interview	1,27	F =15.84	<.001*	η_p^2 = .37	
Main effect	Subjectivity	1.17,31.71	F =12.70	=.001*	η_p^2 = .32	
Interaction	Interview X Subjectivity	1.29,34.9	F =8.51	=.003*	η_p^2 = .24	
Follow up	Cognition vs. Perception	27	t = -2.27	=.03	d = .35	[-5.57,-.28]
Follow up	Cognition vs. Emotion	27	t = -3.77	=.001**	d = .83	[-21.77,-6.44]
Follow up	Perception vs. Emotion	27	t = -4.44	=.002**	d = .63	[-17.91,-4.45]
Follow up	Emotion 1 vs. Emotion All	27	t = -3.66	=.001**	d = .52	[-14.71,-4.15]
Follow up	Cognition 1 vs. Cognition All	27	t = -3.52	=.002**	d = .49	[-4.3,-1.13]
Follow up	Perception 1 vs. Perception All	27	t = -2.07	=.001**	d = .60	[-7.07,-2.07]

Note: * denotes significance at $p < .05$, ** denotes significance at adjusted alpha level $p < .017$

Sensitive Details

Child witnesses reported a larger number of new sensitive details in the second interview (*M* = 41.57, *SD* = 53.58) than in the first interview (*M* = 32.07, *SD* = 48.69). The

percentage of sensitive details from all details witnesses reported increased from 30.6% (*SD* = 25%) in the first interview to 48.1% (*SD* = 29%) in the second interview. A paired t-test showed that interviewers were aware of significantly more new sensitive details after the second interview than after the first interview, $t(27) = 4.09$, $p < .001$, $d = .63$, 95% *CI* = [20.69, 62.45]. A second paired t-test showed that the increase in the percentage of sensitive details from the first to the second interview was significant, $t(27) = 2.49$, $p = .02$, $d = .62$, 95% *CI* = [.03, .32]

Consistency

In the first interview, 23% (*SD* = 12%) of all information reported was repeated, whilst in the second interview this percentage increased to 38.5% (*SD* = 15%). Overall, 32% (*SD* = .12%) of details reported in the two interviews altogether were repeated. In the second interview, information was more frequently repeated within the interview (*M* = .28, *SD* = .12) than across the two interviews (*M* = .17, *SD* = .16). Some information was repeated both within the same interview and across the two interviews.

The percentage of consistent repeated details decreased from 87.1% (*SD* = 15%) in interview 1 to 84.0% (*SD* = 16%) in interview 2. Considering details reported over the course of the two interviews, 84.1% (*SD* = 16%) of repeated details were consistent with previously reported information. Accordingly, inconsistencies within the same interview (*M* = .08, *SD* = .12) were also more common than inconsistencies across interviews (*M* = .04, *SD* = .06). Some repeated information was inconsistent with previously reported information in the same interview as well as in the previous interview (*M* = .04, *SD* = .06).

Inspection of the scatterplot suggested that older children reported a higher percentage of consistent information in the first interview than younger children, both in the first interview and overall (Figure 2). Therefore, children's age was included as a covariate in the analysis.

To assess whether the percentage of consistent responses was associated with interview number (first interview, overall), a two-way RM ANCOVA was conducted investigating the potential effect of interview number (first interview, overall) on the percentage of consistent responses. One witness provided no repeated information in either interview, therefore this witness was excluded from the analysis. The effect of interview number on consistency was non-significant, $F(1,26) = .47$, $p = .50$, $\eta_p^2 = .02$. There was no significant effect of age, $F(1,26) = 2.50$, $p = .13$, $\eta_p^2 = .09$. There were no significant interactions.

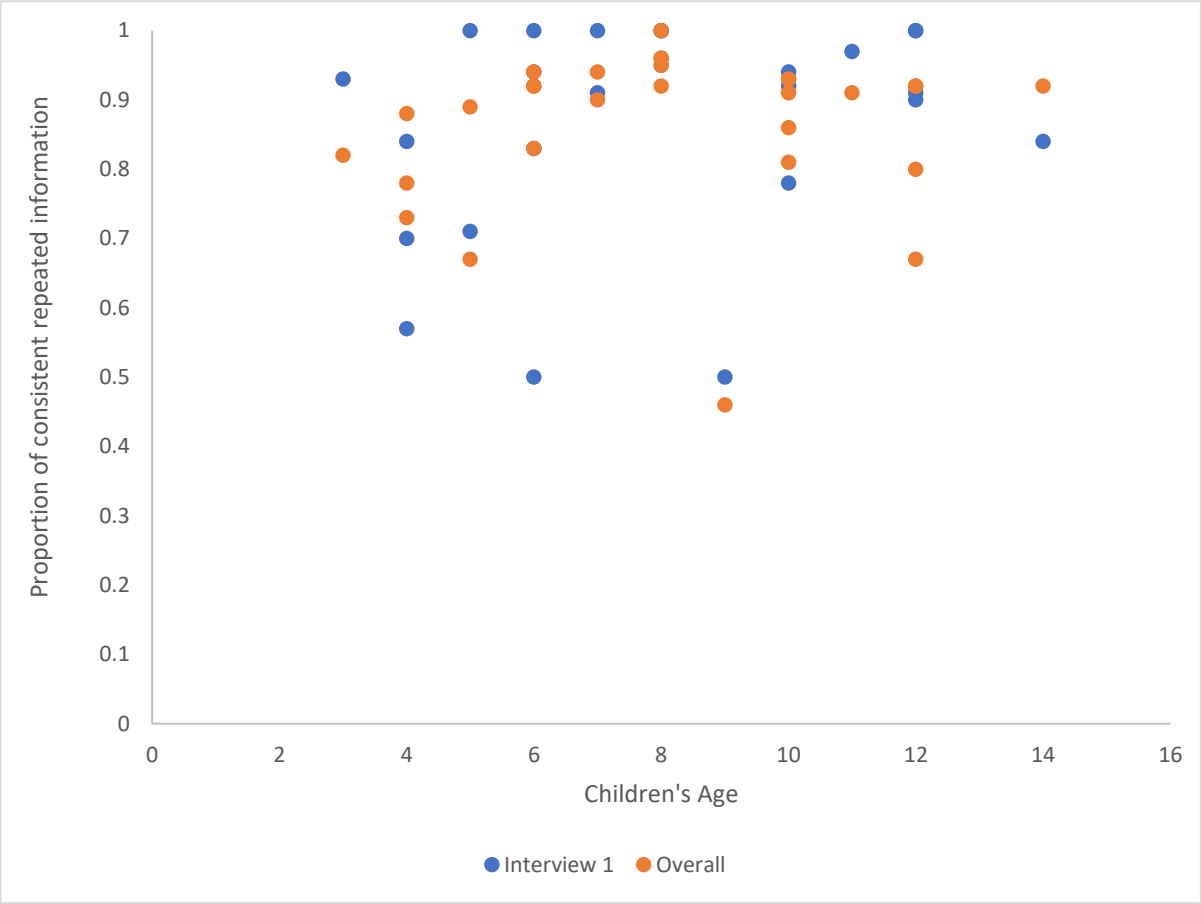


Figure 2 *Children's Age and the Percentage of Consistent Information in Interview 1 and Overall*

Markers of Connectedness

In the first interview, 15.6% (*SD* = 9%) of details mentioned by children contained markers of connectedness whilst this number increased to 17.9% (*SD* = 9%) in the second interview. The frequency of simple temporal markers, complex temporal markers and causal

connections increased from the first to the second interview whilst the frequency of markers of optional states decreased (Table 8).

Table 8 *Percentage of Responses Including Linguistic Markers of Connectedness in Interview 1 and Interview 2*

	Interview 1		Interview 2	
	M	*SD*	*M*	*SD*
Markers of connectedness	**15.6%**	**9%**	**17.9%**	**9%**
Simple temporal markers	3.79%	5%	4.51%	6%
Complex temporal markers	3.33%	3%	4.92%	4%
Causal relations	5.3%	4%	6.22%	6%
Optional states	3.2%	3%	2.26%	3%

Inspection of the scatterplot of the relationship between age and the frequency of markers of connectedness in the first and second interviews suggests that markers of connectedness were more frequently used by older witnesses than younger witnesses (Figure 3). As the graph indicated the existence of a linear relationship between age and the frequency of connectedness markers in both interviews, age was added as a covariate to the analyses regarding markers of connectedness.

To assess whether children's accounts were significantly more connected in the second interview than in the first one, a RM-ANCOVA was conducted assessing the potential effect of interview number (first interview, overall) on the frequency of the four types of connectedness markers (simple temporal, complex temporal, causal, optional) witnesses reported. Children's age was added to the analysis as a covariate. Older children used

markers of connectedness more frequently than younger children, $F(1, 26) = 25.02$, $p < .001$, $\eta_p^2 = .49$. There was also a significant main effect of connectedness, $F(3, 78) = 4.26$, $p = .008$, $\eta_p^2 = .14$, which was further explored using post-hoc tests. No main effect was found for interview number, $F(1, 26) = 1.41$, $p = .25$, $\eta_p^2 = .05$. No significant interactions were found.

To follow up the main effect of connectedness, 6 pairwise comparisons were conducted on the types of connectedness markers using the Bonferroni adjustment (adjusted alpha levels $p < .008$) for multiple comparisons. Markers of causal connections were significantly more frequently used than markers of optional states. All pairwise comparisons are presented in Table 9.

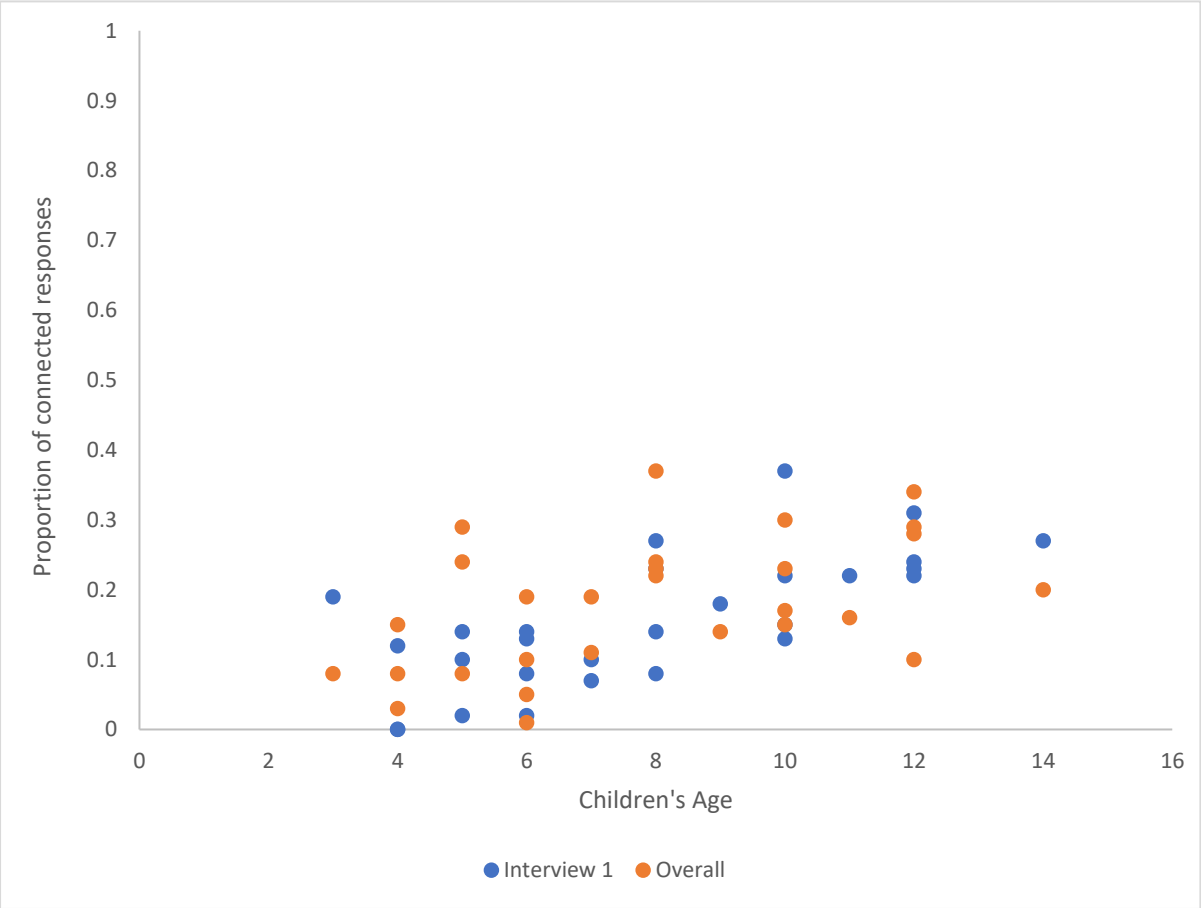


Figure 3 Children's Age and the Percentage of Connected Responses in Interview 1 and Overall

Table 9 *Summary of Analyses of the Percentage of Linguistic Markers in Children's Responses in Interview 1 and Overall*

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Main effect	Connectedness	3,78	F = 4.26	=.25	η_p^2 =.05	
Main effect	Age	1,26	F = 25.02	<.001*	η_p^2 =.49	
Main effect	Interview	1,26	F = 1.41	=.008*	η_p^2 =.14	
Interaction	Connectedness X Age	3,78	F = 2.45	=.07	η_p^2 =.09	
Interaction	Connectedness X Interview	3,78	F = 1.07	=.37	η_p^2 =.04	
Interaction	Age X Interview	1,26	F = 1.30	=.27	η_p^2 =.06	
Interaction	Age X Interview X Connectedness	3,78	F = .62	=.60	η_p^2 =.02	
Follow up	Simple tm vs. Complex tm	26	t = -.29	=.78	d = .08	[-.03,.02]
Follow up	Simple tm vs. Causal	26	t = -1.08	=.29	d = .28	[-.05,.02]
Follow up	Simple tm vs. Optional	26	t = 1.86	=.08	d = .49	[-.002,.05]
Follow up	Complex tm vs. Causal	26	t = -1.08	=.29	d = .25	[-.04,.01]
Follow up	Complex tm vs. Optional	26	t = 2.74	=.01	d = .75	[.007,.05]
Follow up	Causal vs. Optional	26	t = 2.98	=.006**	d = .83	[.01,.07]

Note: * denotes significance at $p < .05$, ** denotes significance at adjusted alpha level $p <$ .008

Question Coherence and Markers of Connectedness

The percentage of linguistically connected details yielded by question type was calculated by dividing the number of linguistically connected details by the number of overall details elicited by questions belonging to each coherence category. The highest percentage of linguistically connected details was elicited by questions which were logically or chronologically connected to children's previous utterances or which followed the direction of conversation set out by the child. A far lower percentage of linguistically connected responses were elicited by questions which had either no apparent coherent structure or which were confusing and incoherent (Table 10).

Due to the low number of interviews in which any chronological (Interview 1: $N = 9$, Interview 2: $N = 9$) or incoherent (Interview 1: $N = 13$, Interview 2: $N = 15$) questions were asked, question coherence was regrouped into three categories for analysis. The new categories consisted of child-driven questions, connected questions (including logical and chronological questions) and unconnected questions (including questions with no apparent coherence or a negative effect on coherence). Following the regrouping, 22 transcripts included were analysed.

Table 10 *The Percentage of Linguistically Connected Responses Elicited by Each Question Type in Interview 1 and Interview 2*

	Interview 1		Interview 2	
	M	*SD*	*M*	*SD*
Child-driven	17.7%	16%	27.9%	27%
Chronological	18.7%	20%	33.3%	31%
Logical	16.5%	10%	16.9%	10%
Neutral	9.4%	10%	9.2%	14%
Incoherent	8.6%	19%	9.2%	18%

To assess whether question coherence and the number of interviews was associated with the proportion of markers of connectedness in witnesses' accounts, a RM-ANOVA was conducted assessing the potential effect if interview number (first interview, second interview) and the regrouped question coherence categories (child-driven, connected, unconnected) on the frequency of markers of connectedness in children's responses. Mauchley's test indicated that the assumption of sphericity was violated for question coherence, $\chi^2(2) = 8.77$, $p = .01$, and the interaction between question coherence and interview number, $\chi^2(2) = 17.60$, $p < .001$. Consequently, Greenhouse-Geisser corrections were applied ($\varepsilon = .74$, $\varepsilon = .63$, respectively). A significant main effect was found for question coherence, $F(1.48,30.99) = 14.52$, $p < .001$, $\eta_p^2 = .41$, which was further explored using post-hoc tests. No main effect was found for interview number, $F(1, 21) = .74$, $p = .40$, $\eta_p^2 = .03$. There was no significant interaction between interview number and question coherence, $F(1.26,26.49) = 1.74$, $p = .20$, $\eta_p^2 = .08$.

To follow up the main effect of question coherence on the frequency of markers of connectedness, pairwise comparisons were conducted on question coherence using the Bonferroni adjustment (adjusted alpha levels $p < .017$) for multiple comparisons. A significantly higher proportion of linguistically connected responses were elicited using child-driven and connected questions than using unconnected questions. All pairwise comparisons are presented in Table 11.

Table 11 *Summary of Analyses of the Proportion of Linguistic Markers in Elicited by Question Type in Interview 1 and Interview 2*

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Main effect	Question Coherence	1.48,30.99	F = 14.52	<.001*	$\eta_p^2 = .41$	
Main effect	Interview	1,21	F = .74	=.40	$\eta_p^2 = .03$	
Interaction	Question Coherence X Interview	1.26,26.49	F = 1.74	=.19	$\eta_p^2 = .08$	
Follow up	Child-driven vs. Connected	21	t = 2.32	=.03	d = .59	[.01,.23]
Follow up	Child-driven vs. Unconnected	21	t = 3.19	=.004**	d = .89	[.07,.31]
Follow up	Connected vs. Unconnected	21	t = 2.59	=.016**	d = .70	[.01,.13]

Note: * denotes significance at $p < .05$, ** denotes significance at adjusted alpha level $p < .017$

2.4 Discussion

Consistent with the hypotheses, results of the present study indicated that the completeness of children's testimonies increased significantly when information was collected over two interviews rather than a single interview. Not only did children report more new details over the course of two interviews, they also mentioned a higher number of crucial time and location details as well as subjective details related to emotions, cognitions and perceptions, and sensitive details which were directly related to the alleged abuse. In fact, children reported a higher proportion of forensically relevant details in the second interview than in the first interview. Contrary to expectations, when both contradictions within the same interview and contradictions across interviews were considered, there was no significant difference in the proportion of contradictory repeated details between the first and second interviews. Exploratory analyses of linguistic connectedness revealed that the proportion of markers of connectedness remained stable across interviews but was significantly associated with the coherence of interviewers' questions. Children's age was associated with the completeness and connectedness of their testimonies, and the impact of multiple interviews on the completeness of narratives.

Completeness

Child witnesses in the current study provided more new information over the course of two interviews than in a single interview in every content category, except for information related to other witnesses. Consistent with the results of Patterson & Pipe (2009), most details reported in the second interview were new rather than repeated. Crucially, interviewing children a second time also led to a significantly higher number of new time and location details available to investigators. Time and location details play a fundamental role in narrative coherence; in multidimensional models, they provide the context of the narrative (Reese et al., 2011), whilst in story grammar approaches, they describe the temporal and

physical setting for the story (Brown et al., 2018). Westcott & Kynan (2004) reported that in 26% of single forensic interviews analysed, children did not provide sufficient information about the physical setting of the events, and this number rose to 50% for temporal setting. Time and location details are vital for the particularisation of the alleged offenses, which is a prerequisite of the successful prosecution of child abuse cases (Powell & Thompson, 1997). Reflecting frequencies found in previous research (Connolly & Read, 2006), most child witnesses in this sample alleged that they had been victims of multiple instances of abuse. Particularisation is especially difficult when adults and children testify about multiple offences (e.g. Brubacher et al., 2014; Connolly & Gordon, 2014; Connolly & Price, 2006; Roberts & Powell, 2006), as memories about repeated occurrences of an event are often organised into "scripts" of general features and selecting a specific occurrence from several repetitions poses a source monitoring challenge (Johnson et al., 1993). Therefore, the increase in the number of time and location details resulting from a second interview may have important implications for developing techniques to aid the process of particularisation when interviewing child witnesses alleging multiple abuse.

In addition to time and location details, subjective details relating to children's emotions, cognitive states and physical perceptions also increased in number over the course of two interviews. In the story-telling framework of legal decision-making, first-person descriptions are essential components of credible testimonies (Pennington & Hastie, 1992). Subjective descriptions are also required for coherent narratives in the multidimensional model (Peterson et al., 2014; Reese et al., 2011) and in story grammar approaches (Brown et al., 2018). In the multidimensional model, subjective first-person descriptions are aspects of the 'evaluative' or 'emotional' dimension (Peterson et al., 2014; Reese et al., 2011), while in story grammar approaches, they are divided into the categories of 'internal response', referring to the emotions, cognitions and goals of the characters before the event and

‘reaction’, describing the emotions, cognitions and goals of the characters in response to the event (Brown et al., 2018). Previous field studies have shown that children only infrequently describe subjective reactions when describing sexual or physical abuse (Lyon, Scurich, Choi, Handmaker, & Blank, 2012, Westcott & Kynan, 2004), despite their ability to verbalise emotions in laboratory settings (Ahern & Lyon, 2013). The results of the present study indicate that in some contexts, multiple interviews can increase the number of subjective reactions children describe, including references to the emotional states, cognitions, perceptions and emotional states of themselves and others.

Not only did child witnesses report a higher number of sensitive details overall than in a single interview, they also referred to a higher proportion of details directly related to the alleged offence in the second interview than in the first one. This tendency could potentially relate to children’s increased comfort with the interview situation and increased trust in the interviewer during the second interview compared with the first one. This explanation is consistent with recommendations of the Extended Forensic Interview guide that interviewers only approach the topic of the abuse in the second or third interview when talking to children alleging sexual abuse to allow sufficient rapport building (Carnes et al., 1999). Alternatively, the increased proportion of sensitive details might reflect children’s increased understanding of the format of forensic interviews and the type of information interviewers are interested in. The way witnesses are expected to recall their memories is highly unusual for children, who are rarely asked to describe past events to this level of detail, especially to listeners who themselves are not knowledgeable about the event. Due to the question-answer format of forensic interviews, skilled interviewers can guide the conversation towards crucial topics even when using exclusively open-ended questions and carefully chosen follow-up questions (Lamb, Orbach, Hershkowitz, Esplin, & Horowitz, 2007). In addition, previous research suggests that interviewers may ask more sensitive questions in further interviews, shifting

from contextual to abuse-related details gradually over the course of multiple interviews (Patterson & Pipe, 2009). Whether due to better rapport between the child and the interviewer or to children's increased understanding of forensic interviews, the increased proportion of sensitive details in the second interview suggests that interviewers may gain a large amount of abuse-related information in multiple interviews which would not come to surface in single interviews.

Consistency

Based on the low consistency found in multiple interviews with children in previous research (Baugerud et al., 2014; Price et al., 2016), contradictions in children's recall were expected to increase in frequency across the two interview occasions. In contrast, when both contradictions within the same interview and across the two interviews were taken into account, results showed no significant difference in the proportion of consistent repeated details between the first and second interview. In line with previous findings (Katz & Hershkowitz, 2013; Waterhouse et al., 2016), there were few contradictions between interviews. The frequency of within-interview contradictions was similar in the two interviews and higher than the frequency of contradictions between interviews. Thus, although the overall number of contradictions increased in multiple interviews, the proportion of contradictions remained constant.

Research in forensic psychology cautions against the use of inconsistencies as correlates for accuracy (Gilbert & Fisher, 2006) but pointing out inconsistencies in witnesses' accounts is a common credibility challenging strategy during cross-examination (Szojka et al., 2017). The present findings suggest that multiple interviews do not necessarily damage the credibility of children's accounts by increasing the proportion of contradictions. However, inconsistencies are also introduced into children's accounts via omissions of

previously mentioned details and additions of new details, and the proportion of these types of inconsistencies was very high in the current sample. Children's credibility may also be challenged on the basis of details that were "left out" from earlier interviews and emerged in later interviews only (Szojka et al., 2017), but this is an inherent result of multiple interviewing – additional interviews are conducted to reveal new details, but new details always mean more inconsistencies.

Connectedness

As no previous research investigated the extent to which forensic interviews with children elicit linguistically connected accounts, hypotheses regarding this component of narrative coherence were exploratory. In the present study, the proportion of linguistic connections in children's accounts remained stable across interviews. This result is consistent with those reported by Habermas & Silveira (2008) who found that repeated recall opportunities did not affect linguistic connectedness in children's autobiographical accounts.

While multiple interviews by themselves do not appear to assist children in constructing linguistic connections, results suggest that interviewers can support child witnesses in building coherent accounts through using questioning strategies that encourage causal-temporal connections between details. Questions in which interviewers provided connections to previously mentioned details elicited a higher proportion of linguistically connected responses than questions which were unconnected to previous topics of conversation. This tendency supports the view that interviewers can help children in constructing temporally and causally connected responses by providing scaffolding through a coherent story framework (Graffam Walker & Warren, 1995). This might be particularly helpful for young children with a limited understanding of temporal relations and story-telling rules (Reese et al., 2011).

However, the highest proportion of linguistically connected responses were elicited by questions that followed the child's lead rather than enforcing a direction of questioning decided by the interviewer, indicating that under some circumstances, children are capable of providing a story framework on their own. This finding is consistent with Feltis et al.'s (2010) results showing that broad open-ended questions yielded more story grammar elements than specific follow-ups. Thus, providing opportunities for children to connect their stories based on their own framework appears to be the most successful strategy for interviewers to follow, but when the task of scaffolding children's narratives does fall on interviewers, their questions need to be logically or chronologically connected in an explicit manner. When interviewers are unresponsive to the direction of conversation suggested by children's responses and ask seemingly unconnected questions that jump in time or between topics, the frequency of markers of linguistic connections in children's accounts decreases.

Age and Narrative Coherence

The present study provided an insight into developmental differences in children's ability to provide coherent testimonies in response to multiple forensic interviews. Consistent with studies, older children provided more complete responses in both interviews than younger children. Additionally, the linguistic connectedness of testimonies also increased as the age of witnesses increased. The age differences in the narrative coherence of children's testimony found in the present study are consistent with findings from research using a multidimensional model of narrative coherence indicating that thematic, chronological, contextual and evaluative elements of coherence develop at different rates across the lifespan (Habermas & Silveira, 2008; Reese et al., 2011).

Although children of all ages provided more complete accounts over the course of the two interviews than in the first interview only, the increase in the number of details increased

with children's age. These results are contrary to the increased benefit of multiple interviews for the youngest children reported by Baugerud et al. (2014) and Katz & Hershkowitz (2014). However, even though the increase in the number of details was proportionally higher for older children, a small increase can be valuable in forensic investigations where younger children often provide only brief descriptions in a single interview (Faller et al., 2010; Langballe & Davik, 2017).

Limitations and Future Research

Whilst the current research provided a crucial insight into the impact of multiple interviews on the narrative coherence of children's forensic testimonies, the design of the study left several questions open for future research. First, the field research design of the current study prevented analyses involving accuracy, as the "ground truth" regarding the events children described was not known to the researchers. The impact of multiple interviews on the accuracy of children's accounts is a much-debated topic and research in the laboratory needs to explore the relationship between accuracy and narrative coherence further before any recommendations can be made about using multiple interviews as a method to scaffold the narrative coherence of child witnesses' accounts. Gaining complete, coherent and consistent accounts is essential for investigations, but inaccurate details can lead to miscarriages of justice (Ceci & Bruck, 1993).

Second, due to the challenges associated with accessing data from real-life police investigations, the sample size was small, but comparable to other field studies involving multiple interviews (Leander, 2010: $N = 27$; Waterhouse et al., 2016: $N = 21$). The sampling method of this study did not allow for the control of age and other witness characteristics. Due to the small number of witnesses and the large age differences in the current sample, it was not possible to investigate the relationship between age and more complex measures of

narrative coherence, such as the content of children's responses or the effectiveness of questioning strategies in eliciting linguistically connected responses. A larger sample size would allow researchers to categorise witnesses into meaningful age groups and offer more nuanced conclusions about the impact of multiple interviews in different age groups.

Third, research by Waterhouse et al. (2016) suggests that law enforcement officers may choose to conduct multiple interviews for a range of reasons. Whilst the most common justifications for conducting a second interview in Waterhouse et al.'s (2016) study was non-disclosure in the first interview and the disclosure of new details after the first interview, other reasons included the child requesting to stop the interview and additional evidence surfacing that appears to contradict the child's statements. The focus of the second interview might differ according to the reason why multiple interviews were deemed necessary, and this might influence how multiple interviews impact the completeness, consistency and connectedness of children's accounts. Future research with larger sample sizes could code the reason why multiple interviews are conducted and evaluate whether interviews with different purposes affect the narrative coherence of children's testimonies in different ways.

Fourth, the type and quantity of complex real-life data provided in multiple forensic interviews prevented the use of coding schemes developed for multidimensional models of coherence (Reese et al., 2011), and local, quantitative measures of narrative coherence were used instead. Further research is needed to investigate the relationship between different models and coding approaches of narrative coherence and to determine whether legal and psychological models of coherence can be integrated.

In addition to exploring the relationship between narrative coherence and accuracy in multiple interviews, future research should clarify the impact of the three components of narrative coherence on the credibility of children's testimony. The story-telling model of

legal decision-making suggests that completeness, consistency and connectedness contribute independently to the credibility of witnesses' accounts (McAdams, 2006; Pennington & Hastie, 1988; Rideout, 2008; Voss et al., 1999). Although previous studies have reported that adults judge narratively coherent accounts as 'better stories' (Schneider & Winship, 2006), research has not yet investigated the effect of narrative coherence on mock jurors' credibility judgements of children's reports elicited in multiple interviews.

Conclusion

The present study was the first to investigate the narrative coherence of children's accounts elicited in multiple forensic interviews. Although further research is needed to establish the relationship between narrative coherence and the accuracy and credibility of children's testimonies, results suggest that multiple interviews increase children's ability to provide coherent testimonies. Multiple interviews were associated with different changes in the components of narrative coherence; while the completeness of children's accounts increased, their consistency and connectedness remained stable. With regards to connectedness, the least well-understood component of narrative coherence, findings suggest that children provide the most linguistically connected accounts when interviewers either provide scaffolding through asking logically and temporally connected questions or allow witnesses to direct the conversation according to their own causal-temporal story framework.

Chapter 3. The Impact of Multiple Interviews on the Accuracy and Narrative Coherence of Children's Descriptions of a Staged Event

3.1 Introduction

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., 2007) or at the 'global' level characterising the entire narrative (Peterson et al., 2013). 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. As the story-telling model suggests that narrative coherence is a central element in legal decision-making (Pennington & Hastie, 1992), it is essential to consider narrative coherence integrated in a legal framework, alongside the accuracy of children's testimony.

The results of Study 1 demonstrated that in forensic interviews with children alleging physical and sexual abuse, multiple interviews are associated with a positive effect on the narrative coherence of testimonies, as they increase the completeness of witnesses' testimonies, without compromising their consistency and causal-temporal connectedness. These findings are consistent with previous results from both field studies and laboratory studies that indicated that multiple interviews increase the number of details reported by children (e.g. Hershkowitz & Terner, 2007; La Rooy et al., 2005; Leander, 2010), introduce only a small proportion of contradictory information into children's accounts (Katz & Hershkowitz, 2013; Waterhouse et al., 2016), and do not affect their causal-temporal connectedness (Haberman & Silveira, 2008).

Whilst the increase in narrative coherence suggests that multiple interviews can provide a useful tool for interviewers aiming to gather more details about the allegations, their use cannot be recommended without exploring the relationship between narrative coherence and accuracy in multiple interviews. Accuracy cannot be measured in field studies, as researchers are usually ignorant of 'ground truth' when analysing real-life forensic interviews. Although material evidence, such as pictures or recordings may be used to evaluate the accuracy of testimonies, these are rarely available to researchers. Although previous laboratory studies have examined the relationship between accuracy and narrative coherence as well as the effect of multiple interviews on accuracy, the three-way relationship between multiple interviews, narrative coherence and accuracy has yet to be explored.

The Accuracy of Children's Testimony

In the past decades, research has shown that under optimal conditions, schoolchildren and even pre-schoolers are able to provide accurate descriptions of events they experienced (for a review see Lamb et al., 2008). However, the extent to which children's recall is accurate is influenced by a range of factors, including the style of questioning used by the interviewer and children's developmental level (Lamb et al., 2003). Children's responses are most accurate in response to open-ended questions, whilst suggestive and overly specific questions lead to a decrease in accuracy (e.g. Lamb et al., 2008; Leichtman & Ceci, 1995). This tendency is especially strong 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 even very young children can monitor the accuracy of their recall.

Even truthful and overall accurate statements elicited through open-ended questioning can include a substantial number of inaccuracies (Brubacher, Peterson, La Rooy, Dickinson, & Poole, 2019). Among other mechanisms, inaccuracies appear in children's statements through intrusions, resulting from a failure to inhibit event-irrelevant thoughts, and conjunctions, details of other events falsely reported as relating to the current event through 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. The impact of inaccuracies in children's testimony in forensic contexts can range from inconsequential mistakes regarding peripheral details to severe consequences associated with errors related to types of touch or the identity of the alleged perpetrator.

Multiple interviews and accuracy

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 findings suggesting that multiple interviews decrease the accuracy of children's accounts have resulted from 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). In Leichtman & Ceci's (1995) 'Sam Stone' study, the accuracy of children's accounts often remained high across multiple interviews in the control condition where no suggestive methods were used.

Studies using non-suggestive questioning cast doubt on the view of multiple interviewing as an inherently suggestive method (e.g. Baker-Ward et al., 1993; Hubbard et al., 2016; Knutsson et al., 2011). Peterson et al. (2013) found that children's accounts of an

injury requiring a hospital stay remained highly accurate in multiple interviews, even when the delay between interviews was several years long. Furthermore, under some circumstances, multiple interviews may exert a positive influence on 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. In field studies investigating the effect of multiple interviews on the accuracy of children's testimonies, it is not possible to 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 more forgetting due to the length of time elapsed, rather than from instructions to recall the events multiple times. However, experimental studies can overcome this limitation by involving a delayed control condition; a group of participants interviewed a single time 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; Knutson et al.,

2016), whilst other studies found a decrease in accuracy after comparable delays (Dietze et al., 2911; Salmon et al., 2012).

Narrative coherence and accuracy

Although laboratory studies have established that the narrative coherence of children's accounts correlates positively with the amount of details they report (e.g. Chae et al., 2016; Kulkofsky et al., 2008; Peterson et al., 2013), there is less consensus regarding the relationship between narrative coherence and accuracy (Kulkofsky et al., 2008). 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).

However, 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). In Chae et al.'s (2016) study, 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 effort into 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).

While the relationship between completeness, connectedness and accuracy has mainly been explored by studies analysing the narrative quality on children's accounts (Chae et al., 2016; Kulkofsky et al., 2008; Wang et al., 2015), the relationship between consistency and accuracy has been described in studies investigating the impact of multiple interviews on 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, the relationship between consistency and accuracy in multiple interviews may be 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 their results, Baugerud et al. (2014) suggest that low consistency across multiple interviews does not necessarily implicate low accuracy.

Kulkofsky et al. (2008) also studied the influence of the delay between the event to be remembered and the interview on the relationship between narrative coherence and accuracy. In this study, children were interviewed a single time either 1 week or 1 month after a play event they participated in. 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 (Kulkofksy 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. The consistency of children's accounts was not analysed in this study (Kulkofsky et al., 2008).

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 does not hold for differences in the type of event children are asked to remember and the delay between the event and the interviews.

The Present Study

In the present study, primary school children were interviewed about their memories of a first aid training either on two occasions, 1 week and 5 weeks after the event, or on a single occasion 5 weeks after the event. To distinguish between the effects of delay and

multiple recall occasions, between-subjects analyses were conducted comparing the second recall of the experimental group and the single interview of the control group. In the experimental group, within-subjects analyses were also conducted to compare the completeness, connectedness and accuracy of information available to the interviewers after the first interview with the information gained overall in the two interviews.

As the two groups were matched for length of delay, and the interviews were conducted in a non-suggestive, open-ended manner, it was expected that the recall of the two groups should be comparable on all measures, whilst the overall information reported across the two interviews of the experimental condition would differ from information reported in the first interview in some respects.

Specifically, it was expected that 1) information provided over the course of two interviews would be more complete than information provided in the first interview only, however, 2) the information provided overall would also be less accurate due to the influence of forgetting over the 5-week delay. Due to the delay between the two interviews, it was expected that, 3) consistent details would be more accurate than reminisced details in the second interview of the experimental group. Based on the results of Study 1, a change in the connectedness of details across interviews was not anticipated.

3.2 Method

Participants

Ninety-seven children 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 in

Surrey (*N* = 21) and a primary school in Essex (*N* = 55). Children were divided into two age groups: 8-9-year-olds (*N* = 37, *M* = 104.05 months, *SD* = 1.02) and 10-11-year-olds (*N* = 39, *M* = 123.41 months, *SD* = .74).

Design and Procedure

Permission to conduct the study was granted by the Royal Holloway, University of London ethics committee ahead of the start of data collection. The experiment was based on a 2 (age: 8-9 years, 10-11 years) x 2 (single interview, multiple interview) design. Children in the experimental group (*N* = 41) were interviewed 1 week and 5 weeks after the event to-be-remembered whilst children in the delayed control group (*N* = 35) were only interviewed 5 weeks after. Comparisons were made both between subjects, comparing the performance of the two age groups and the multiple and single interview groups, and within-subjects, comparing the recall of children in the first interview with their overall recall provided over the course of the two interviews.

Initial contact with schools and after school clubs was made via an information pack (Appendix B) 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 (Appendix C). 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 a research assistant. 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 usefulness 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 (Appendix D). 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, a different researcher (the author) interviewed children assigned to the experimental group individually about their experiences of the first aid training. The interview protocol (Appendix E) 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 (e.g. Lamb et al, 2007, 2009) 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 = 121) were coded by the first author according to the completeness, consistency, connectedness and accuracy of children’s responses. The connectedness of children’s responses to the question instructing them to “Think back and try to tell me everything that happened from the beginning to the end” was coded separately from their answers to different questions.

Completeness

The first aid training event included 35 target details. The presence of target details was identified in transcripts and the completeness of children's recall was then assessed by calculating the proportion of target elements present in their accounts from the overall number of possible target details. 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 12.

Table 12 *List of Target Details by Training Phase*

Phase of first aid training	Target details
Introduction	Partner
	Teacher
Inspecting the first aid kit	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)
What is an emergency?	Emergency or not
	Cards
	Sorting (the cards)
	Example of emergency
	Example of not emergency
How to help	Video about choking
	Hit back
	Five times
	Check object
	Repeat
	Video about fainting
	Turn them on their side
	Tilt their head
	Listen for breathing
	Call for help

(continued)

(continued)

Phase of first aid training	Target details
Calling for help	999
	Instruction sheets
	Ambulance
	Type of accident
	Phone number
	Location
	Directions
	Certificate

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.

Accuracy

Incorrect details were defined as demonstrably untrue details related to the first aid training mentioned by the children. Off-topic details were ignored, as were details whose veracity 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 13.

Table 13 *Incorrect Details by Training Phase*

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

Causal-temporal connectedness

Connectedness codes fell into two categories: local markers of connectedness and global coherence.

Local markers of connectedness

Local markers of connectedness were coded using the linguistic cohesion coding scheme developed by Kulkofsky et al. (2007) in the same manner as in Study 1.

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.

Global coherence

Global coherence was coded according to a modified version of the Narrative Coherence Coding Scheme developed by Reese et al. (2011). Interviews received a score of 1 to 4 based on the number of main events (Familiarity with the first aid box, Emergency or not, How to help, Calling for help) about which chronological information was given. Both unmarked chronological ordering (e.g. *We did the first aid box. We played with the cards.)* and marked violations of temporal ordering (e.g. *We put a plaster on. That was before the videos.)* were considered correct.

Inter-rater Reliability

A random selection of 20% of the transcripts (N = 24) were coded for inter-rater reliability. Cohen's κ was used to assess agreement between the main coder and an independent second coder. There was perfect agreement 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 local markers of connectedness, $\kappa = .95$, $SE = .01$, 95% CI [.93, .97] and substantial for global markers $\kappa = .77$, $SE = .01$, 95% CI [.55, .98].

3.3 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]. The distribution of participants in experimental conditions were well-balanced, with no significant differences in age group, $\chi^2(1) = .23$, $p = .63$, and gender, $\chi^2(1) = .39$, $p = .53$ across conditions.

Between-subjects Analyses

To distinguish between the effects of delay and multiple interviews, between-subjects ANOVAs comparing the recall of children in the delayed control condition with the second recall of children in the experimental condition were conducted.

Completeness

On average, children in the delayed control condition recalled 13.11 (*SD* = 5.16) correct details, whilst the average number of correct details in the second interview of the experimental condition was 14.49 (*SD* = 5.46) (Table 14). Girls recalled more correct details than boys both overall in the two interviews of the experimental condition (Girls: *M* = 15.00, *SD* = 1.12, Boys: *M* = 11.18, *SD* = 1.19), and in the control condition (Girls: *M* = 20.00, *SD* = 1.13, Boys: *M* = 16.47, *SD* = 1.05).

Table 14 *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	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
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

To assess whether the condition children were assigned to affected the completeness of their accounts, a three-way ANOVA was conducted assessing the effects of condition (single, multiple), children's age (8-9-years old, 10-11-years-old) and gender (boy, girl) on the number of correct details recalled in the second interview. Girls recalled significantly more details than boys, $F(1, 68) = 15.99$, $p < .001$, $\eta_p^2 = .19$. No main effect was found for condition, $F(1, 68) = .33$, $p = .57$, $\eta_p^2 = .01$, or age group $F(1, 68) = .25$, $p = .62$, $\eta_p^2 = .004$. There was, however, a significant interaction between condition and age group, $F(1, 68) = 5.06$, $p = .03$, $\eta_p^2 = .07$ (Figure 4), which was further explored using post-hoc tests. There were no other significant interactions.

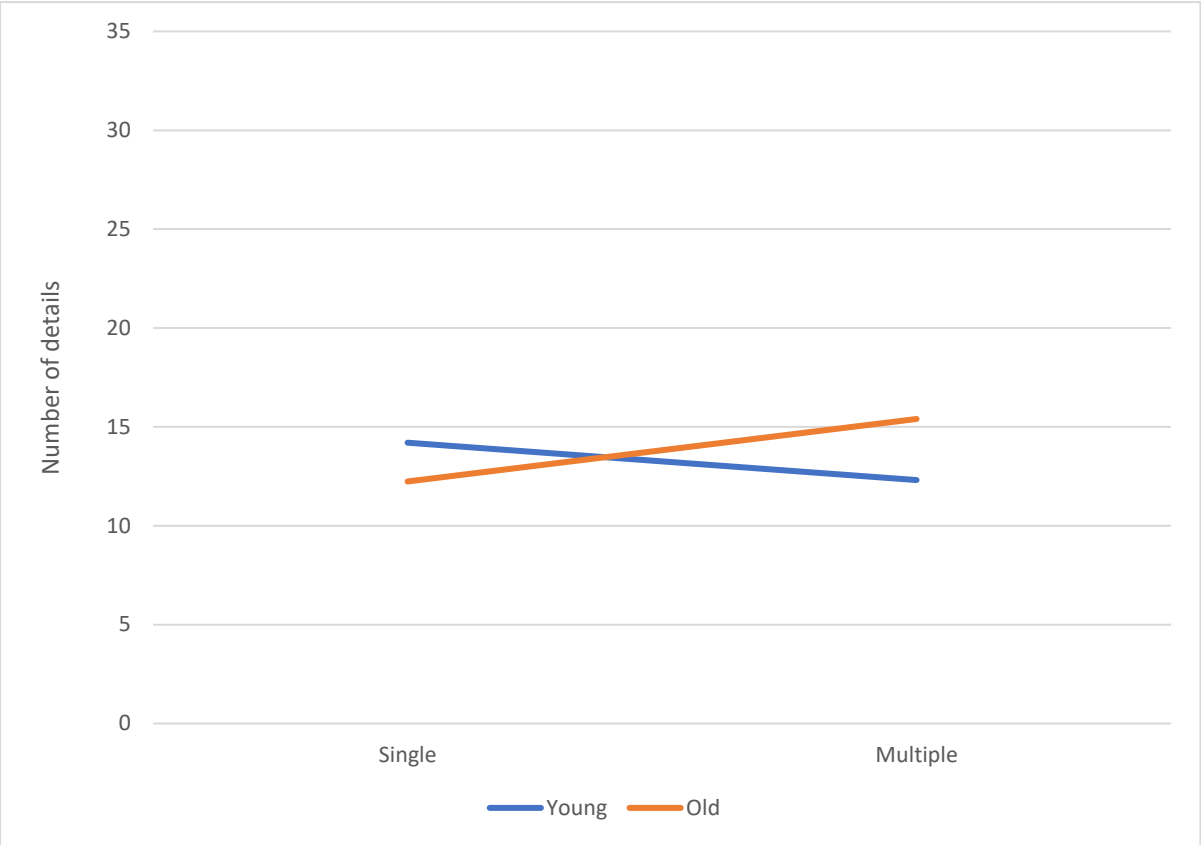


Figure 4 *The Interaction Between Condition and Age Category on the Number of Details*

Two paired t-tests were conducted to explore the interaction between condition and age group (adjusted α = .025). There was no significant difference between the control group and the experimental group for 8-9-year-olds. The difference between control group and the experimental condition for 10-11-year-olds was also non-significant at the adjusted alpha levels. All pairwise comparisons are presented in Table 17.

Table 15 *Analyses of the Completeness of Children's Recall by Condition, Age and Gender*

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Main effect	Condition	1,68	$F = .33$	=.57	$\eta_p^2 = .01$	
Main effect	Age	1,68	$F = .25$	=.62	$\eta_p^2 = .004$	
Main effect	Gender	1,68	$F = 15.99$	<.001*	$\eta_p^2 = .19$	
Interaction	Condition X Age	1,68	$F = 5.06$	=.03*	$\eta_p^2 = .07$	
Interaction	Condition X Gender	1,68	$F = .67$	=.42	$\eta_p^2 = .01$	
Interaction	Age X Gender	1,68	$F = .09$	=.77	$\eta_p^2 = .01$	
Interaction	Condition X Age X Gender	1,68	$F = 3.40$	=.07	$\eta_p^2 = .05$	
Follow up	Single vs. Multiple (8-9)	35	$t = .37$	=.71	$d = .12$	[-3.12,4.51]
Follow up	Single vs. Multiple (10-11)	38	$t = -2.08$	=.04	$d = .66$	[-6.50,-.09]

Note: * denotes significance at $p < .05$

Accuracy

On average, children recalled .73 (*SD* = .74) incorrect details in the second interview of the experimental condition, whilst the average number of incorrect details in the delayed control condition was 1.14 (*SD* = 1.12). (Table 14). Details reported in the second interview were 94% (*SD* = 6%) correct on average, compared with 92% (*SD* = 8%) in the experimental condition.

To assess whether the condition children were assigned to affected the accuracy of their accounts, a three-way ANOVA was conducted assessing the effects of condition (single, multiple), children's age (8-9-years old, 10-11-years-old) and gender (boy, girl) on the proportion of correct details from all details recalled in the second interview. There was no significant main effect of condition, $F(1, 68) = 2.31$, $p = .13$, $\eta_p^2 = .03$, age group, $F(1, 68) <.005$, $p = .99$, $\eta_p^2 < .001$, or gender $F(1, 68) = 3.54$, $p = .06$, $\eta_p^2 = .05$. There were no significant interactions.

Connectedness

A proportional score for linguistic markers was calculated by dividing the number of linguistic markers in children's accounts by the overall number of details children reported, in order to ensure the independence of connectedness and completeness scores. The average proportion of linguistic markers of connectedness was 1.11 (*SD* = .68) in the second interview of the experimental condition, and 1.25 (*SD* = 1.13) in the delayed control condition (Table 15).

Table 16 *The Proportion of Linguistic Markers of Connectedness in Children's Responses in Each Interview*

	Single 5 weeks		Repeated 1 1 week		Repeated 2 5 weeks		Repeated Overall	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
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

To assess whether the condition children were assigned to affected the frequency and type of linguistic markers of connectedness in their accounts, a four-way RM ANOVA was conducted assessing the effects of condition (single, multiple), marker type (simple temporal, complex temporal, causal, optional), children's age (8-9-years old, 10-11-years-old) and gender (boy, girl) on the proportion of linguistic markers of connectedness in the 5-week interview. Mauchley's test indicated that the assumption of sphericity was violated for marker type, $\chi^2(5) = 42.32$, $p < .001$. Consequently, a Greenhouse-Geisser correction was applied ($\varepsilon = .69$).

There was a significant main effect of marker type $F(2.08, 141.43) = 33.40$, $p < .001$, $\eta_p^2 = .33$, which was further explored using post-hoc tests. There was no significant main effect of condition, $F(1, 68) = .31$, $p = .58$, $\eta_p^2 = .01$, age group, $F(1, 68) = .02$, $p = .89$, $\eta_p^2 = 0$, or gender $F(1, 68) = .91$, $p = .34$, $\eta_p^2 = .01$. There were no significant interactions.

The main effect of marker type was followed up by 6 paired t-tests (adjusted α = .008). Simple temporal markers were used significantly more frequently than complex temporal markers, markers of optional states, and markers of causal connections. Markers of optional states were more frequent than complex temporal markers and markers of causality. All pairwise comparisons are presented in Table 18.

Table 17 *Analyses of the Connectedness of Children's Responses by Condition, Age and Gender*

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Main effect	Condition	1,68	$F = .31$	=.58	$\eta_p^2 = .01$	
Main effect	Age	1,68	$F = .02$	=.88	$\eta_p^2 <.001$	
Main effect	Gender	1,68	$F = .91$	=.34	$\eta_p^2 = .01$	
Main effect	Marker type	2.08, 141.43	$F = 33.40$	<.001*	$\eta_p^2 = .33$	
Interaction	Condition X Age	2.08, 141.43	$F = 1.28$	=.26	$\eta_p^2 = .02$	
Interaction	Condition X Gender	2.08, 141.43	$F = 1.20$	=.28	$\eta_p^2 = .02$	
Interaction	Age X Gender	2.08, 141.43	$F = 1.55$	=.22	$\eta_p^2 = .02$	
Interaction	Condition X Age X Gender	2.08, 141.43	$F = 1.31$	=.26	$\eta_p^2 = .02$	
Interaction	Marker type X Condition	2.08, 141.43	$F = .99$	=.38	$\eta_p^2 = .01$	
Interaction	Marker type X Age	2.08, 141.43	$F = .25$	=.79	$\eta_p^2 =.004$	
Interaction	Marker type X Gender	2.08, 141.43	$F = .72$	=.49	$\eta_p^2 = .01$	
Interaction	Marker type X Age X Gender	2.08, 141.43	$F = .83$	=.48	$\eta_p^2 = .01$	

(continued)

(continued)

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Interaction	Marker type X Age X Condition	2.08, 141.43	*F* = .49	=.62	η_p^2 = .01	
Interaction	Marker type X Gender X Condition	2.08, 141.43	*F* = .10	=.92	η_p^2 =.001	
Interaction	Marker type X Age X Gender X Condition	2.08, 141.43	*F* = .33	=.73	η_p^2 = .01	
Follow up	Simple tm vs. Complex tm	75	*t* = 7.08	<.001**	*d* =1.03	[.29,.52]
Follow up	Simple tm vs. Optional	75	*t* = 2.76	=.007**	*d* = .40	[.05,.31]
Follow up	Simple tm vs. Causal	75	*t* = 9.57	<.001**	*d* =1.45	[.36,.55]
Follow up	Complex tm vs. Optional	75	*t* = -6.76	<.001**	*d* = .59	[-.30,-.16]
Follow up	Complex tm vs. Causal	75	*t* = 1.06	=.29	*d* = .19	[-.04,.13]
Follow up	Optional vs. Causal	75	*t* = 6.06	<.001**	*d* = .90	[.18,.37]

Note: * denotes significance at *p* < .05, ** denotes significance at *p* < .008

Global connections

On average, children established correct temporal order for less than 2 out of the 4 key activities included in the first aid training (Table 16). To assess whether the condition children were assigned to affected the number of global connections established between key events in their accounts, a three-way ANOVA was conducted assessing the effects of condition (single, multiple), children's age (8-9-years old, 10-11-years-old) and gender (boy, girl) on the number of global connections present in the 5-week interview.

There was no significant main effect of condition, $F(1, 68) = 1.09$, $p = .30$, $\eta_p^2 = .02$, age group, $F(1, 68) = .41$, $p = .53$, $\eta_p^2 = .01$, or gender $F(1, 68) = .46$, $p = .50$, $\eta_p^2 = .01$. There were no significant interactions.

Table 18 *Number of Key Events Placed in Correct Temporal Order in Each Interview*

	Single 5 weeks		Repeated 1 1 week		Repeated 2 5 weeks	
	M	*SD*	*M*	*SD*	*M*	*SD*
Global connections	1.71	1.24	1.83	1.73	1.49	1.39

Within-subject analyses

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 remembered more details in their second interview than during their first recall. Explicit contradictions were rare, with 7 children mentioning contradictory details in the first interview, 11 in the second interview and 13 across the two interviews. Children whose accounts included contradictory information generally mentioned a single contradiction (Interview 1: *M* = 1, *SD* = 0; Interview 2: *M* = 1.18, *SD* = .40, Interview 3: *M* = 1.19, *SD* = .40).

To assess whether the consistency of information affected the accuracy of children's accounts, a three-way RM-ANOVA was conducted assessing the effects of consistency (consistent, omitted, reminisced), children's age (8-9-years old, 10-11-years-old) and gender (boy, girl) on the accuracy of information recalled in the second interview. Due to the limited definition and small number of contradictions, contradictory information was excluded from

the analysis. Four children provided no repeated details, no omitted details or no reminisced details, therefore, they were excluded from the analysis. Mauchley's test indicated that the assumption of sphericity was violated for consistency, $\chi^2(2) = 10.34$. Consequently, a Greenhouse-Geisser correction was applied ($\varepsilon = .78$). A significant main effect was found for consistency, $F(1.56,49.86) = 6.12$, $p = .008$, $\eta_p^2 = .16$, which was further explored using post-hoc tests. No main effect was found for age group, $F(1, 32) = .31$, $p = .58$, $\eta_p^2 = .01$, or gender, $F(1, 32) = .32$, $p = .08$, $\eta_p^2 = .09$. There were no significant interactions.

The main effect of consistency was followed up by 3 paired t-tests (adjusted $\alpha = .017$). Consistent information ($M = .97$, $SD = .06$) was more accurate than reminisced information ($M = .82$, $SD = .26$). The accuracy of details mentioned in the first interview but omitted from the second ($M = .90$, $SD = .19$) did not differ significantly from consistent or reminisced information. All pairwise comparisons are presented in Table 19.

Table 19 *Analyses of The Effect of Consistency on the Accuracy of Children's Recall by Age and Gender*

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Main effect	Consistency	1.56, 49.86	$F = 6.12$	=.008*	$\eta_p^2 = .16$	
Main effect	Age	1,32	$F = .32$	=.58	$\eta_p^2 = .01$	
Main effect	Gender	1,32	$F = 3.32$	=.08	$\eta_p^2 = .09$	
Interaction	Consistency X Age	1.56, 49.86	$F = 1.75$	=.18	$\eta_p^2 = .05$	
Interaction	Consistency X Gender	1.56, 49.86	$F = 2.03$	=.15	$\eta_p^2 = .06$	
Interaction	Consistency X Age X Gender	1.56, 49.86	$F = .02$	=.97	$\eta_p^2 <.001$	
Interaction	Age X Gender	1,32	$F = .08$	=.78	$\eta_p^2 =.002$	
Follow up	Consistent vs. Omitted	38	$t = 2.22$	=.03	$d = .50$	[.01,.13]
Follow up	Consistent vs. Reminisced	37	$t = 3.35$	=.002**	$d = .79$	[.06,.24]
Follow up	Omitted vs. Reminisced	35	$t = 1.55$	=.13	$d = .34$	[-.03,.21]

Note: * denotes significance at $p < .05$, ** denotes significance at $p < .017$

Completeness

On average, children reported 15.63 (*SD* = 5.00) in the first interview of the experimental condition and 18.54 (*SD* = 5.30) details over the course of the two interviews (Table 14). To assess whether the number of interviews affected the completeness of children's accounts, a RM-ANOVA was conducted assessing the effects of interview number (first interview, overall), children's age (8-9-years old, 10-11-years-old) and gender (boy,

girl) on the number of correct details recalled. Significantly more correct details were reported overall than in the first interview only, $F(1,37) = 59.18$, $p < .001$, $\eta_p^2 = .61$. No main effect was found for age group, $F(1, 37) = .53$, $p = .47$, $\eta_p^2 = .01$, or gender, $F(1, 37) = 3.74$, $p = .07$, $\eta_p^2 = .09$. A significant interaction was found between interview number and gender, $F(1, 37) = 4.38$, $p = .04$, $\eta_p^2 = .11$ (Figure 5), which was further explored using post-hoc tests. There were no other significant interactions.

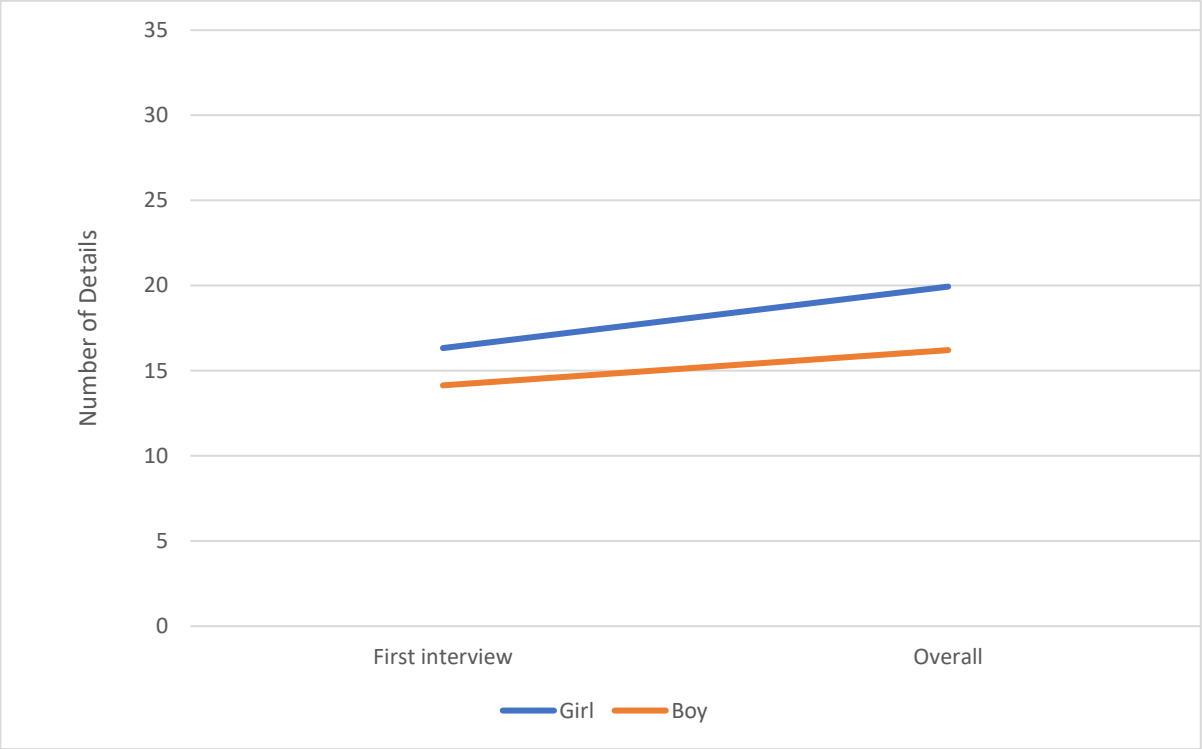


Figure 5 *The Interaction Between Gender and Interview Number on the Number of Details*

Two paired t-tests were conducted to explore the interaction between gender and interview number (adjusted $\alpha = .025$). There was no significant difference between the number of correct details reported by boys and girls in the first interview. The difference between the number of correct details reported over the course of the two interviews by boys and girls was also non-significant at the adjusted alpha levels. All pairwise comparisons are presented in Table 20.

Table 20 *Analyses of the Completeness of Children's Testimony by Age and Gender in the First Interview and Overall*

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Main effect	Interview	1,37	F = 59.18	<.001*	η_p^2 = .62	
Main effect	Gender	1,37	F = 3.57	=.07	η_p^2 = .09	
Main effect	Age	1,37	F = .53	=.47	η_p^2 = .01	
Interaction	Interview X Gender	1,37	F = 4.38	=.04*	η_p^2 = .11	
Interaction	Interview X Age	1,37	F =.002	=.97	η_p^2 <.001	
Interaction	Gender X Age	1,37	F = 1.99	=.17	η_p^2 = .05	
Interaction	Interview X Gender X Age	1,37	F = 1.06	=.31	η_p^2 = .03	
Follow up	Boy vs. Girl (Interview 1)	39	t = -1.26	=.21	d = .41	[-5.17,1.20]
Follow up	Boy vs. Girl (Overall)	39	t = -2.20	=.03	d = .71	[-6.78,-.28]

Note: * denotes significance at $p < .05$

Accuracy

On average, children recalled .61 (*SD* = .80) incorrect details in first second interview of the experimental condition, and .73 (*SD* = .74) incorrect details overall (Table 14). Details reported in the first interview were 96% (*SD* = 6%) accurate on average, compared with 95% (*SD* = 6%) in the experimental condition.

To assess whether the number of interviews affected the accuracy of children's accounts, a RM-ANOVA was conducted assessing the effects of interview number (first interview, overall), children's age (8-9-years old, 10-11-years-old) and gender (boy, girl) on

the proportion of correct details from all details mentioned. Accuracy in the first interview was significantly higher than over the course of the two interviews, $F(1,37) = 9.06$, $p = .005$, $\eta_p^2 = .20$. No main effect was found for age group, $F(1, 37) = 1.74$, $p = .20$, $\eta_p^2 = .05$, or gender, $F(1, 37) = 2.14$, $p = .15$, $\eta_p^2 = .06$. There were no significant interactions.

Connectedness

The average proportion of linguistically connected details was 1.02 (*SD* = .56) in the first interview of the experimental condition and 1.06 (*SD* = .55) over the course of the two interviews (Table 15). To assess whether the number of interviews affected the frequency and type of linguistic markers of connectedness in their accounts, a RM-ANOVA was conducted assessing the effects of interview number (first interview, overall), marker type (simple temporal, complex temporal, causal, optional), children's age (8-9-years old, 10-11-years-old) and gender (boy, girl) on the percentage of linguistic markers of connectedness in children's accounts. Mauchley's test indicated that the assumption of sphericity was violated for marker type, $\chi^2(5) = 34.21$, $p < .001$, and the interaction between marker type and interview number, $\chi^2(5) = 51.41$, $p < .001$. Consequently, Greenhouse-Geisser corrections were applied ($\varepsilon = .69$ and $\varepsilon = .59$, respectively).

There was a significant main effect of marker type $F(2.08, 77.05) = 47.43$, $p < .001$, $\eta_p^2 = .56$, which was further explored using post-hoc tests. There was no significant main effect of interview number, $F(1, 37) = 2.08$, $p = .16$, $\eta_p^2 = .05$, age group, $F(1, 37) = .66$, $p = .42$, $\eta_p^2 = .02$, or gender $F(1, 68) = .10$, $p = .76$, $\eta_p^2 = .003$. There were no significant interactions.

The main effect of marker type was followed up by 6 post-hoc tests (adjusted $\alpha = .008$). Simple temporal markers were used significantly more frequently than complex temporal markers and markers of causal connections. Markers of optional states were more

frequent than complex temporal markers and markers of causality. There were no other significant differences.

Table 21 *Analyses of the Connectedness of Children's testimony by Age and Gender in the First Interview and Overall*

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Main effect	Interview	1,37	$F = 2.08$	=.16	$\eta_p^2 = .05$	
Main effect	Age		$F = .02$	=.88	$\eta_p^2 <.001$	
Main effect	Gender		$F = .91$	=.34	$\eta_p^2 = .01$	
Main effect	Marker type	2.08, 77.05	$F = 47.43$	<.001*	$\eta_p^2 = .56$	
Interaction	Interview X Age	1,37	$F = .18$	=.67	$\eta_p^2 = .01$	
Interaction	Interview X Gender	1,37	$F = .09$	=.76	$\eta_p^2 =.002$	
Interaction	Age X Gender		$F = 1.55$	=.22	$\eta_p^2 = .02$	
Interaction	Interview X Age X Gender	1,37	$F = 3.10$	=.09	$\eta_p^2 = .08$	
Interaction	Marker type X Interview	1.77, 65.57	$F = 1.23$	=.30	$\eta_p^2 = .03$	
Interaction	Marker type X Age	2.08, 77.05	$F = 1.99$	=.14	$\eta_p^2 =.05$	
Interaction	Marker type X Gender	2.08, 77.05	$F = .96$	=.39	$\eta_p^2 = .03$	
Interaction	Marker type X Age X Gender	2.08, 77.05	$F = .39$	=.69	$\eta_p^2 = .01$	

(continued)

(continued)

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Interaction	Marker type X Age X Interview	1.77, 65.57	$F = 1.30$	=.28	$\eta_p^2 = .03$	
Interaction	Marker type X Gender X Interview	1.77, 65.57	$F = .37$	=.67	$\eta_p^2 = .01$	
Interaction	Marker type X Age X Gender X Interview	1.77, 65.57	$F = 3.09$	=.06	$\eta_p^2 = .08$	
Follow up	Simple tm vs. Complex tm	40	$t = 10.05$	<.001**	$d = 2.05$	[.34,.52]
Follow up	Simple tm vs. Optional	40	$t = 3.27$	=.002**	$d = .70$	[.07,.31]
Follow up	Simple tm vs. Causal	40	$t = 9.10$	<.001**	$d = 1.87$	[.33,.50]
Follow up	Complex tm vs. Optional	40	$t = -5.48$	<.001**	$d = 1.15$	[-.33,-.15]
Follow up	Complex tm vs. Causal	40	$t = -1.02$	=.31	$d = .18$	[-.07,.03]
Follow up	Optional vs. Causal	40	$t = 6.55$	<.001**	$d = 1.01$	[.15,.29]

Note: * denotes significance at $p < .05$ ** denotes significance at $p < .008$

Global connections

To assess whether the condition children were assigned to affected the number of global connections established between key events in their accounts, a RM-ANOVA was conducted assessing the effects of interview number (first interview, second interview), children's age (8-9-years old, 10-11-years-old) and gender (boy, girl) on the number of global connections present.

There was no significant main effect of interview number, $F(1, 37) = 1.54$, $p = .22$, $\eta_p^2 = .04$, age group, $F(1, 37) = .34$, $p = .53$, $\eta_p^2 = .02$, or gender $F(1, 37) = .14$, $p = .71$, $\eta_p^2 = .004$. There were no significant interactions.

Correlations

The completeness and connectedness of children's accounts were positively correlated, both when information gained over the course of the two interviews was considered ($r = .59$, $p < .001$), and in each interview (Interview 1: $r = .67$, $p < .001$; Interview 2: $r =.55$, $p < .001$). However, consistency in the second interview was only significantly correlated with the completeness of information in the first interview ($r = .49$, $p = .001$), not in the second interview or overall. There was no significant correlation between consistency and connectedness overall, or in either interview.

The number of global links established between key events in the first interview was positively correlated with connectedness in the first interview ($r = .37$, $p = .02$). The number of global links established in the second interview was positively correlated with connectedness in the second interview ($r = .36$, $p = .02$) and overall connectedness in the two interviews ($r = .32$, $p = .04$). There were no significant correlations between global links and other measures of the quality of children's testimony.

The overall accuracy of children's accounts provided over the course of the two interviews was positively correlated with the completeness of overall information ($r = .37$, $p = .02$), as well as the completeness of information provided in the first interview ($r = .33$, $p = .04$). The consistency of children's accounts in the second interview correlated positively with their accuracy in the second interview (r = .32, p = .04), but not overall. There were no significant correlations between accuracy and connectedness.

Table 22 *Correlations between the accuracy, completeness, connectedness and consistency of children's testimony*

		Overall accuracy	Accuracy 1	Accuracy 2	Completeness Overall	Completeness 1	Completeness 2	connectedness Overall	Connectedness 1	Connectedness 2	Global links 1	Global links 2	Consistency
Accuracy 1	*r*	.80*	-										
	p	<.001	-										
Accuracy 2	*r*	.80*	.45*	-									
	p	<.001	.003	-									
Overall Completeness	*r*	.37*	.24	.34*	-								
	p	.017	.13	.033	-								

(continued)

(continued)

		Overall accuracy	Accuracy	Accuracy 2	Completeness	Overall Completeness 1	Completeness 2	Connectedness	Overall connectedness 1	Connectedness 2	Global links 1	Global links 2	Consistency
Completeness 1	*r*	.33*	.24	.27	.90*	-							
	p	.035	.13	.09	<.001	-							
Completeness 2	*r*	.28	.19	.38	.85*	.66*	-						
	p	.074	.24	.014	<.001	<.001	-						
Overall Connectedness	*p*	.08	.21	.19	.58*	.64*	.52	-					
	r	.60	.19	-.03	<.001	<.001	.001	-					
Connectedness 1	*r*	.05	.14	-.06	.59	.67*	.41*	.93	-				
	p	.76	.39	.70	<.001	<.001	.008	<.001	-				

(continued)

(continued)

		Overall accuracy	Accuracy	Accuracy 2	Completeness	Overall 1	Completeness 2	Completeness connectedness	Overall 1	Connectedness 2	Connectedness	Global links 1	Global links 2	Consistency
Connectedness 2	*r*	.11	.25	-.03	.49*	.51*	.55*	.93	.93	-				
	p	.51	.12	.87	.001	.001	<.001	<.001	<.001	-				
Global links 1	*r*	-.06	-.10	-.06	.21	.30	.08	.28	.37*	.14	-			
	p	.72	.54	.71	.20	.05	.59	.08	.02	.37	-			
Global links 2	*r*	-.001	-.004	-.02	.23	.14	.28	.32*	.23	.36*	.16	-		
	p	.99	.98	.23	.15	.38	.08	.04	.16	.02	.32	-		
Consistency	*R*	.28	.29	.32*	.14	.49*	.08	.26	.24	.27	.11	-.08	-	
	p	.07	.06	.04	.39	.001	.64	.10	.13	.09	.40	.62	-	

3.4 Discussion

Results supported the expectations that a second interview would lead to an increase in the overall number of details reported by children, but that there would be no difference between the second interview of the experimental group and the recall of the delayed control group. Interviews conducted 5 weeks after the first aid training were almost identical, with no difference between the completeness, accuracy and connectedness of interviews in the single and multiple conditions. However, within-subjects analyses showed that the cumulative information received by interviewers over the course of the two interviews was both significantly more complete and significantly less accurate than information revealed in the first interview. In the second interview, reminisced details were also less accurate than consistent details. As expected, the connectedness of children's accounts remained stable across interviews.

Completeness

Although hypermnesia was only present in approximately one quarter of the sample, reminiscence between interviews was the norm, leading to a significantly larger number of new 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, or the emergence of new allegations or new evidence after the first interview (Waterhouse et al., 2016). In all these cases, 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 the results of Study 1, suggest that this assumption is generally correct.

Interestingly, the difference between the number of details reported in the first interview and overall was proportionally smaller in this experiment than in Study 1. 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 in a staged study may be less salient than autobiographical memories about stressful or traumatic events (Baker-Ward et al., 1993; Baugerud et al., 2014; Peterson, 2011, 2015), which could play a role in the difference between the proportional amount of reminiscence found in the field study and the experimental study. Additionally, the present study consisted of a single event which children were asked to recall. Most witnesses in Study 1 testified about repeated events, therefore, the large amount of new information surfacing in later interviews might have been the result of a shift in the topic of conversation between different occurrences (Patterson & Pipe, 2009). Finally, analyses of completeness in the current study included only correct details, while it was not possible to distinguish between correct and incorrect details in the field study. Therefore, some newly recalled details in later interviews in Study 1 may have been incorrect.

Connectedness

In the present study, the connectedness component of narrative coherence was measured using two different methods. Firstly, the proportion of local linguistic connections between details was coded in the same way as in Study 1 (Kulkofsky et al., 2008). Secondly, the number of global links established between the four key activities included in the first aid training was measured to provide a score on a modified version of the temporal element of the Narrative Coherence Coding Scheme (Reese et al., 2011). Using two different approaches allowed the researchers to establish whether there is a correlation between widely used methods of measuring temporal coherence in children's testimony (Kulkofsky et al., 2008; Reese et al., 2011). Consistent with expectations, children's local and global connectedness scores were positively correlated. On average, children provided correct global connections between less than 2 out of the 4 key activities included in the first aid training. The proportion

of globally connected key events from all key events was similar to the proportion of details connected using local simple temporal markers from all details mentioned.

Children's ability to place only half of the events experiences correctly on a temporal scale in the present study is consistent with previous research showing that child witnesses often struggle with providing temporal information in their testimony (Westcott & Kynan, 2004). However, when all types of local markers of connectedness were considered, including complex temporal markers, markers of optional states and markers of causal connections, results are more encouraging. Altogether, children used more than one marker of connectedness for each detail mentioned, indicating that they did attempt to provide a causal-temporal framework of the events spontaneously in response to open-ended questioning.

Consistent with the results of Study 1, the proportion of local and global markers of connectedness remained stable across multiple interviews, suggesting that multiple interviews by themselves do not help children place details on a causal-temporal framework.

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), results indicate that multiple interviews, by themselves, do not result in a decrease in the accuracy of children's accounts, but the delay between interviews does compromise accuracy. There was no difference in the accuracy of interviews conducted 5 weeks after the first aid training, regardless of whether they were preceded by an interview 1 week after the event or not, suggesting 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 5-week interviews in both the single and multiple conditions was on average less accurate than the 1-week interview in the multiple condition, demonstrating the damaging effect of delay on the accuracy of children's accounts. This suggests that even though multiple interviews *per se* do not have a negative effect on children's accuracy, 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. A second factor contributing to the decrease in accuracy from the first interview to the overall information repeated over the course of the two interviews is the presence of new incorrect details in Interview 2. Not only did children recall a higher number of incorrect details in the second interview than the first interview, the errors they made were also different from the ones they mentioned in the first interview, leading to a small increase in the overall number of incorrect details. 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 1 additional incorrect detail reported on average, leading to a 1% decrease in accuracy.

Results suggest that some, but not all, components of narrative coherence are related to the accuracy of recall. In contrast to the notion of a completeness-accuracy trade-off suggested by Kulkofsky et al. (2008), in the present study, the overall accuracy of children's accounts correlated positively with the completeness of their recall in interview 1 and overall. Although the completeness and connectedness of children's accounts were strongly correlated, results do not provide support for the theory that causal-temporal connectedness allows children to report more complete and accurate details, as there were no significant correlations between connectedness and accuracy in either interview. Consistent with

Baugerud et al.'s (2014) findings, the consistency of children's testimony was not correlated with the overall accuracy of their recall. However, the consistency of details reported in the first interview showed a positive correlation with the accuracy of recall in the second interview.

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. 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, Lamb, La Rooy, & Pipe, 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 6 interviews, the witness 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 Study 1, in the present study, most details recalled by children were consistent across the two interviews, with only 3 details reminisced in the second interview, and 4 details mentioned in the initial interview but omitted from the second one, on average. Consistent with Study 1, the rate of explicit contradictions was low both within and across interviews.

The accuracy of details in the two interviews was associated with their consistency. Details that were consistent across the two interviews were 97% accurate, whilst details reminisced in the second interview were only 82% correct. The large 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 the recall of adults (Gilbert & Fisher, 2006) and children (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 multiple interviews did not correlate with their overall accuracy in the present study or 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 Study 1 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 the field study, 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.

Children's Age and Gender

In Study 1, the completeness and connectedness of children's recall was positively correlated with age. Developmental differences in the narrative coherence of children's testimony were not replicated in the present study, presumably due to the narrower age range of participants in the experiment (8-11 years), compared with the field study (3-14 years). The only significant difference between 8-9-year-olds and 10-11-year-olds in the present study was an interaction between the age group children belonged to and the condition they were assigned to that was associated with the completeness of their testimony in the 5-week interview. Plotting the interaction showed that older children reported more details in the multiple interview condition, whilst younger children provided more complete accounts in the single interview condition. Although follow-up tests did not confirm the significance of this difference, the tendency of younger children to report more information if they have been interviewed previously does not provide support for the view that the increased rapport and interview practice inherent in multiple interview protocols is especially beneficial for the youngest age groups (Carnes et al., 1999; Langballe & Davik, 2017). There were no age differences in the amount of details children in the experimental group recalled in the first interview and overall.

Whilst no gender differences were found in Study 1, the reports of girls and boys differed in completeness in the present study. Between-subjects analyses revealed that girls reported significantly more details than boys in the 5-week interview, regardless of whether they have been interviewed previously or not. Within-subjects analyses revealed an added

layer of complexity to the relationship between gender and completeness, in that the difference between the completeness of girls' and boys' testimonies was larger when both interviews were considered than in the first interview only. However, follow-up tests did not confirm the significance of this difference.

Although gender difference in the completeness of recall are consistent with previous research comparing girls' and boys' narratives (Haden, Haine & Fivush, 1997; Lamb & Garretson, 2003), there is an important caveat to interpreting this result. Lamb & 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 valuable 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 the findings of Study 1, which had the benefit of high external validity but did not allow for analyses involving accuracy. Secondly, whilst in field contexts the proportion of invitations often

remains small despite guideline recommendations, the interviewer in the current study used invitations only. Results thus show the relationship between narrative coherence and accuracy in ideal, rather than typical forensic interviews.

Thirdly, the sample size for the present study was limited due to restrictions on time and financial resources. Power calculations using G*Power (Faul, Erdfelder, Bucher, & Lang, 2009) have shown that the sample size provided more than 80% power for both within-participants and between-participants analyses, assuming that the population effect size is large (η_p^2 = .26) or medium (η_p^2 = .13). However, power fell below 80% for small population effect sizes (η_p^2 = .02), suggesting that some of the analyses might be underpowered. Statisticians caution against using the sample effect size as an equivalent of the population effect size (Hoenig & Heisy, 2001; O'Keefe, 2007), however, it is difficult to estimate the population effect size for the present study due to the scarcity of previous research using similar measures and the lack of systematic reviews in the field.

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 more than a year and a half between interviews in Study 1. Finally, the current study failed to find developmental differences in the completeness, accuracy, consistency and connectedness of recall between 8- to 9-year-olds 10- to 11-year-olds. 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 all three components of the completeness, consistency and connectedness of children's recall, and the relationship between the components of narrative coherence and the accuracy of children's descriptions of a staged event elicited over the course of two interviews. Results suggest that children's recall in the second interview of the experimental condition did not differ from the recall of the control group matched for delay, providing further support for the view that non-suggestive multiple interviews do not constitute a risk factor for the accuracy of children's accounts. However, the delay between the 1-week interview and the 5-week interview of the experimental condition did compromise the accuracy of children's accounts to a small extent, suggesting that interviewers need to be mindful of the impact of delay when making decisions about the timescale of multiple interviews. Findings regarding the effects of multiple interviews on the components of narrative coherence were consistent with Study 1, as the completeness of children's accounts increased over the course of the two interviews, whilst their causal-temporal connectedness remained stable. The completeness of children's accounts was positively correlated with their accuracy, contradicting the theory of a completeness-accuracy trade-off in children's memories. Neither the connectedness, nor the consistency of children's recall was correlated with the overall accuracy of their accounts.

Chapter 4. The Impact of Narrative Coherence and Multiple Interviews on Mock Jurors' Ratings of the Credibility of Child Witnesses

4.1 Introduction

The story-telling model of legal decision-making (Pennington & Hastie, 1992) states that the narrative coherence of testimonies is a key predictor of their credibility (McAdams, 2006; Pennington & Hastie, 1988; Rideout, 2008; Voss et al., 1999). Results from Study 1 and Study 2 demonstrated that multiple interviews have an overall positive effect on the narrative coherence of children's recall, as they increase the completeness of testimonies, whilst their consistency and connectedness remain stable. However, research has not yet addressed the question of how the narrative coherence of children's testimonies elicited in multiple interviews influences adults' judgements of the credibility of child witnesses' testimony. Exploring the relationship between narrative coherence and credibility is crucial for determining whether the cognitive and linguistic benefits of multiple interviews reported by field and laboratory studies translate into positive legal outcomes for children in the courtroom.

The Story-telling Model

The story-telling model of legal decision-making has its origins in Fisher's narrative paradigm (1985), which states humans process complex information through story-telling narratives rather than rational arguments. Pennington & Hastie (1992) applied this model to the legal context through suggesting that legal decision-making, which relies on understanding and evaluating large among of complex information, is essentially based upon comparing narratives.

Pennington & Hastie describe the juror as a 'sense-making information processor who strives to create a meaningful summary of the evidence available that explains what happened in the events depicted through witnesses, exhibits and arguments at trial.' (Pennington & Hastie, 1992, p. 519). The central feature of this 'sense-making' is active story construction, which is impeded by the disjointed nature of the information presented in court. Not only is the story presented through different witnesses and material evidence, witness statements themselves follow a fragmented, question-answer format. To integrate these fragments of information into a story, jurors need to create chains of events connected by physical or motivational causation. The extent to which a story created in this manner is acceptable and jurors' level of confidence in the story is decided based on four certainty principles, three of which (coverage, uniqueness, goodness of fit) describe the ability of the story to provide a full, unambiguous and unique explanation of the evidence presented at trial (Pennington & Hastie, 1992). The fourth principle, coherence, is an internal characteristic of the story presented, providing a measure of the quality of the narrative itself. Consistent with Pennington & Hastie's theory, empirical research suggests that jurors rely on certainty principles when evaluating the defendant's guilt in murder (Voss et al., 1999; Voss & Van Dyke, 2001) and rape cases (Wilmott et al., 2018), but few studies have applied the story-telling model (Pennington & Hastie, 1992) experimentally to assess the credibility of child witnesses alleging sexual abuse.

Credibility Judgements of Children's Testimony

Past research investigating the credibility of children's statements has addressed adults' ability to distinguish true and false stories with the help of a guidelines created by experts (Vrij, 2005) and explored the factors mock jurors rely on when such guidelines are not available (e.g. Castelli, Goodman & Ghetti, 2005; Wessel, Magnussen, & Melinder, 2013).

Guideline-based or ‘structured’ approaches to credibility assessments include the Reality Monitoring approach (RM: Johnson & Raye, 1981) and Criteria-Based Content Analysis (CBCA: Steller & Kohnken, 1989). Both approaches are based on the hypothesis that truthful and false accounts can be distinguished from one another based on a range of observable characteristics (Undeutsch, 1989). Structured guidelines were created with the aim of helping evaluators make science-based decisions about the veracity of witnesses’ statements (Steller & Kohnken, 1989). Evaluations of these guidelines reached vastly different conclusions about their effectiveness, with Parker & Brown (2000) reporting an over 90% success rate in discerning between true and false statements using content-based approaches, whilst Landry & Brigham (1992) found that the accuracy of decisions made on the basis of CBCA criteria was barely above chance level.

Research on ‘unstructured’ credibility judgements, made without reliance on a set of criteria, suggests that when no guidance is provided, both laypeople and experts rely on a range of case characteristics, including the age of the witness (Tabak & Klettke, 2014), the presence or absence of emotional displays in the witness testimony (Wessel et al., 2013) and the questioning style used by the interviewer when taking the witness statement (Castelli et al., 2005). Concerningly, studies assessing the success rate of professionals working in child protection and law enforcement found that they were no more accurate in detecting children’s deception than laypeople (Chahal & Cassidy, 1995; Leach, Talwar, Lee, Bala, & Lindsay, 2004).

Structed approaches to credibility assessment

Several forensic interview guidelines, including the Extended Forensic Interview (Carnes et al., 1999) require investigators to rely on a set list of criteria when assessing the credibility of children’s statements. Whilst the effectiveness of the EFI credibility guidelines

in helping adults judge children's credibility correctly has not been empirically studied, the list of criteria shares many similarities with the Reality Monitoring Approach and CBCA, such as placing emphasis on the amount of descriptive details reported by children, the inclusion of peripheral details and the presence or absence of motivational factors which may encourage children to tell falsehoods (Carnes, Nelson-Gardell, Wilson, & Orgassa, 2001). Whilst both the Reality Monitoring approach and CBCA are based on the hypothesis that there are differences in the verbal content of truthful and deceitful statements (Undeutsch, 1989), they rely on a different set of criteria to determine whether statements are credible.

The Reality Monitoring approach states that true statements and false statements differ in both general and specific characteristics (Stromwall, Bengtsson, Leander & Granhag, 2004). In general terms, true memories are more complex and clearer than false statements, and specifically, they include more contextual, perceptual and affective information. Because made-up and imagined events originate in internal, rather than external sources, false stories include a higher proportion of statements referring to cognitive operations (Stromwall et al., 2004). Whilst these guidelines were developed using research-based evidence (Stromwall et al., 2004), studies evaluating the Reality Monitoring approach found that the proposed general and specific differences do not always help distinguish between truth-tellers and liars (Masip, Sporer, Garrido, & Herrero, 2005; Stromwall & Granhag, 2005). Consistent with Masip et al.'s (2005) suggestion that individual Reality Monitoring criteria are not diagnostic of credibility, Stromwall & Granhag (2005) found that while true stories indeed contained more affective information than false accounts, they also included less visual information, contradicting the theory that truth-tellers mention more perceptions than liars.

Criterion-based Content Analysis is more widely accepted professionally than Reality Monitoring, constituting admissible evidence in German courts (Kohnken, 2002). Similar to Reality Monitoring, CBCA relies on a list of details that are ostensibly more likely to occur in true statements than in deceitful statements (Vrij, 2005). The list contains general and specific features of the testimony which are thought to be difficult to fabricate, and motivational factors, which relate to differences in impression management between children who tell the truth and children who lie (Vrij, 2005). Vrij's (2005) review of studies evaluating the validity of CBCA suggests that results from both field studies and analogue laboratory research are often difficult to interpret due to methodological limitations. As CBCA has been developed specifically for use with children alleging sexual abuse, field studies have a much higher validity than laboratory studies, which often require children to tell stories about events that are not stressful, emotional or in other ways memorable. However, in field studies relating to child sexual abuse, it is exceedingly difficult to obtain ground truth, and many of the 'gold standards' of truthful testimonies used in credibility research are questionable (Vrij, 2005). Vrij (2005) suggests that neither confessions, nor convictions are independent from the quality of children's statements, as the more persuasive a statement is, the more likely defendants perceive their defence to be weak and the more likely they are to confess. Confessions, in turn, influence the jury's decision to convict or acquit the defendant. DNA or medical evidence would be more appropriate to use as 'proof' for the veracity of statements; however, this is rarely available to researchers (Vrij, 2005).

Keeping these limitations in mind, field studies have generally shown that CBCA scores discriminated between true and false accounts in the majority of cases (Lamb et al., 1997; Parker & Brown, 2000). Amado et al.'s meta-analysis of 18 studies applying CBCA criteria to children's true and false statements reported a positive overall effect size for the reliability of the total CBCA score, as well as each component criterion, supporting the

Undeutsch hypothesis (Amado, Arce, & Farina, 2015). Based on these findings, Amado et al. (2015) suggest that CBCA meets four out of the five criteria for court admissibility of scientific evidence, namely, 1) the hypothesis is testable, 2) the hypothesis has been tested, 3) the method has been the subject of peer review and publication, and 4) the method has a known error rate. However, it is unclear whether CBCA meets the fifth criterion, which states that the theory upon which the hypothesis and method are based should be generally accepted amongst the appropriate scientific community.

Doubts about whether CBCA meets court admissibility criteria have been expressed by Vrij (2005), who pointed out that the error rate of laboratory CBCA studies is around 30% (Amado et al., 2015; Vrij, 2005), an unacceptable rate for legal decision-making. Although the error rate for field studies is lower, around 10% (Amado et al., 2015; Vrij, 2005), the outcome of these studies is difficult to interpret due to the researchers' ignorance to ground truth. Furthermore, Vrij et al. found that while CBCA scores reliably distinguished liars and truth-tellers amongst 5-6-year-olds and undergraduate students, there were no significant CBCA differences in the true and false accounts of 10-11-year-olds and 14-15-year olds (Vrij, Akehurst, Soukara, & Bull, 2002). In addition, the lies of children who received coaching that familiarised them with CBCA criteria were indistinguishable from true stories (Vrij et al., 2002). Another important limitation of the CBCA approach is the implied assumption that testimonies are either completely true or fabricated in their entirety, disregarding the fact that true testimonies could involve imagined or misremembered elements, whilst skilful liars may opt for reporting the truth with as few alterations as possible (Vrij, 2005).

A shared limitation of CBCA and Reality Monitoring, alongside other structured approaches to detecting children's deception, is that they rely on high quality interviews based on free recall and the lack of suggestion, but most studies evaluating these approaches

used poor quality forensic interviews (Henderson & Andrews, 2018). Comparing CBCA assessments of unstructured interviews and interviews using the NICHD protocol, Hershkowitz et al. found that the use of the NICHD protocol decreased the proportion of cases investigators deemed 'no judgement possible' and increased inter-rater reliability (Hershkowitz, Fisher, Lamb, & Horowitz, 2007).

In summary, while field studies revealed encouraging results about the effectiveness of content-based structured credibility approaches, especially when used in combination with best practice interviewing methods (Hershkowitz et al., 2007), accuracy rates found in laboratory studies do not fully support the view that Reality Monitoring or CBCA scores should be admissible evidence in court (Henderson & Andrews, 2018; Vrij, 2005). However, when used with caution during investigations, structured guidelines may be preferable to unstructured approaches to credibility assessment, which can be influenced by interviewer bias and stereotypes.

Unstructured approaches to credibility assessment

As structured credibility approaches are not presently admissible in court in most jurisdictions (Vrij, 2005), research has investigated how laypeople and experts assess children's credibility in the absence of explicit guidance on how to evaluate the truthfulness of their accounts. Key factors influencing adults' decision-making regarding children's credibility, include characteristics of the witness, such as children's age (Tabak & Klettke, 2014) or emotional displays (Wessel et al., 2013), characteristics of the interview, such as the types of questions used by the interviewer (Castelli et al., 2005) and characteristics of the evaluator, such as gender (McCauley & Parker, 2001) and professional experience (Wessel et al., 2013).

Witness age

From the range of factors affecting adults' judgements of children's credibility, the age of the witness has received the most attention. However, the results of studies evaluating the influence of age have been highly inconsistent; some reported that younger children appear more credible than older children (Goodman et al., 1987), others reported the contrary (Connolly, Price, Lavoie, & Gordon, 2008), whilst a third group of studies found no effect of age on the perceived credibility of eyewitnesses (Luus, Wells & Turtle, 1995).

Goodman et al. (1987) found that when the age of the witness was manipulated between 6 years and 10 years, mock jurors rated younger witnesses as more credible, even though the accounts were identical in the two conditions. Similarly, using electronic focus groups to recreate the deliberation aspect of court trials, Tabak & Klettke (2014) reported that when the age of the witness was varied between 6 years, 11 years and 15 years, mock jurors rated younger children as more credible than older children. In contrast to the findings of Goodman et al. (1987), the latter study also found a significant difference between the proportion of guilty verdicts reached in cases involving younger versus older children (Tabak & Klettke, 2014). The higher credibility ratings given to younger witnesses in comparison to adolescents may be partially accounted for by Quas et al.'s finding that over half of adult respondents to a survey measuring attitudes to child witnesses agreed with the statement that 'a child cannot describe sexual abuse unless he or she has experienced it' (Quas, Thompson, & Clarke-Stewart, 2005, p. 442).

In contrast to Tabak & Klettke's (2014) findings, Connolly et al. (2008) reported that adults' ratings of the credibility of children's recall of a play event increased as the age of the witness increased. Similarly, Goodman et al. found that mock jurors perceived adult witnesses as more credible than children, regardless of whether the case witnesses testified about was a hit and run or a murder, and whether participants were presented with written or videotaped testimonies (Goodman, Golding, Helgeson, Haith, & Michelli, 1987). To explain

contrasting results, the relationship between age and credibility is often described in terms of the 2-factor model (Leippe & Romanczyk, 1989). This model suggests that children's perceived credibility is determined by a combination of their perceived honesty or reliability and their perceived competency. Proponents of this approach claim that older children tend to be perceived as more competent due to their more developed cognitive capacities, whilst younger children are perceived as more honest and reliable, because they have less motivation to lie (McCauley & Parker, 2001).

The 2-factor model suggests that the type of event children are asked to recall influences the relationship between age and credibility, as investigators put emphasis either on children's ability to remember and report details in a highly accurate manner, or on witnesses' motivations to make false claims (McCauley & Parker, 2001, Ross, Jurden, Lindsay, & Keeney, 2003). This theory was tested by McCauley & Parker's (2001) study comparing adults' judgements of the credibility of 6-year-olds and 13-year-olds when the children provided testimony on a sexual assault case and on a trial involving a robbery. The two-factor model predicts that younger children would be seen as more credible when testifying about the sexual assault case due to their perceived honesty, whilst the testimonies of older children would be judged more credible in the robbery case which emphasises memory accuracy. However, the results showed no interaction between the impact of witness age and case type on adults' credibility judgements (McCauley & Parker, 2001). Noting that the summaries presented to mock jurors may have provided insufficient demonstration of children's age, Ross et al. (2003) replicated the test of the 2-factor model presenting mock jurors with either a fictional witness testimony or the recreation of an entire trial. Factor loadings on a questionnaire assessing mock jurors' impressions of the honesty and cognitive ability of the child witness were consistent with the 2-factor model, however, only honesty predicted the verdict reached.

The lack of consistent results in studies investigating the relationship between children's age and mock jurors' ratings of their credibility suggest that a large variance exists in the perceived credibility of eyewitnesses belonging in the same age group, and under some circumstances, other characteristics of the witness and the testimony provided may be stronger predictors of credibility than age (Luus et al., 1995). Luus et al. (1995) asked college students to judge the credibility of a child witness or an adult witness, manipulating perceived age so that half of the participants were misinformed about the age of the witness. They found that whilst a range of factors predicted credibility judgements, including the perceived confidence of the witness, neither actual age, nor believed age were significant predictors (Luus et al., 1995).

Emotional reporting

In addition to age, a second witness characteristic predicting credibility ratings is whether victims display or verbalise emotions that conform to the social stereotypes about the behaviours and attitudes of victims of sexual abuse (Klippenstine & Schuller, 2012; Wessel et al., 2013). Comparing the perceived credibility of child actors testifying about physical abuse who displayed different emotions, Wessel et al. (2013) found that children who appeared sad received the highest credibility ratings, whilst the truthfulness of the accounts of children displaying angry or positive expressions was rated significantly lower than testimonies in both the sad and neutral conditions. Adults' use of child victims' emotional expressions when evaluating their credibility is concerning because stereotypes about child abuse do not always match the actual emotions displayed by child victims reporting abuse at court. Children who experienced sexual abuse often report feeling confused about what happened to them (Clancy, 2009) and may display relaxed or neutral expressions in court rather than sadness or anxiety (Wood, Orsak, Murphy, & Cross, 1996), indicating that relying on stereotypes of victims' emotional reporting can be a misleading method of credibility assessment.

Interview style

Research suggests that interview quality is a more reliable indicator of the accuracy of children's accounts than most witness characteristics (Henderson & Andrews, 2018). Mock jurors appear to be either implicitly or explicitly aware of this relationship, as Castelli et al. (2005) found that children's accounts elicited using leading interviewing methods were rated as less credible than testimonies provided in non-suggestive interviews. Castelli et al. (2005) systematically varied the style of interviews mock jurors were presented with between non-leading, less leading or highly leading interviews and found that credibility ratings decreased as the suggestiveness of the interview increased. Interestingly, there was an interaction between the type of case, the age of the witness and interview style (Castelli et al., 2005). Whilst in a case of non-familial sexual abuse mock jurors rated leading interviews as less credible than non-leading interviews for both 4-year-olds and 7-year-olds, interview style was a significant predictor of credibility ratings for 4-year-olds only in an incest case.

Juror characteristics

In addition to a range of variables related to the witness and the case, research has revealed that mock jurors' judgements of the credibility of child witnesses are also affected by several characteristics of the jurors themselves (Bottoms & Goodman, 1994; Davies & Rogers, 2009; McCauley & Parker, 2001; Wessel et al., 2013). Wessel et al. (2013) compared the impact of the emotional expression displayed by a child witness on the credibility ratings of laypeople and child protective workers. Although it was expected that child protective workers would be less affected by emotional displays than laypeople due to their understanding of the variety of possible emotions experienced by child victims, results instead indicated that the child protective workers gave overall higher credibility ratings, regardless of condition (Wessel et al., 2013). Taking a broader view of jurors' experience

with children, McCauley & Parker (2001) collected information about the number of children and younger siblings participants had and the amount of time they spent in the company of children and found that prior experience with children did not impact credibility ratings.

Along with jurors' personal or professional experience with children, research has analysed differences between male and female mock jurors' ratings of the credibility of child witnesses and their likelihood of convicting the defendant (e.g. Bottoms & Goodman, 1994; Davies & Rogers, 2009; McCauley & Parker, 2001; Wessel et al., 2013). Results indicate that female jurors perceive child witnesses alleging sexual abuse as more credible than male jurors (Bottoms & Goodman, 1994; Davies & Rogers, 2009; McCauley & Parker, 2001; Wessel et al., 2013), and may also be more likely to find the defendant guilty (McCauley & Parker, 2001). According to McCauley & Parker's (2001) findings, the difference in male and female jurors' ratings of child witnesses' credibility applied not only to sexual assault cases, but also in a case where the child alleged that the defendant stole her backpack.

Multiple Interviews and Children's Credibility

Although most children involved in child abuse investigations are interviewed more than once by the police, in addition to formal and informal interviews with other professionals, few studies have analysed the impact of multiple interviews on children's credibility (Waterhouse, 2016; Yozwiak, Golding, & Marsil, 2004). Yozwiak et al. (2004) compared mock jurors' credibility ratings and verdict decisions of case summaries stating either that the child witness disclosed little information in the first interview but described the full extent of the allegations in the second interview or that the child provided a complete description of the details of the alleged incident in both interviews. Results showed that mock jurors rated the child witness as more credible in the condition where the child's recall was consistent across the interviews, and they were also more likely to convict the defendant in

the consistent condition (Yozwiak et al., 2004). Yozwiak et al.'s (2004) study provided crucial insight into the potential risk of multiple interviews in decreasing children's credibility through introducing inconsistencies into their accounts, however, the briefness of the case summary might have influenced jurors' perceptions of the child's testimony. Specifically, in the consistent condition, the summary of the child's direct examination stated that the child provided the same information in the two police interviews, which does not appear to reflect tendencies in real-life sexual abuse investigations (Leander, 2010; Patterson & Pipe, 2009; Waterhouse et al., 2016).

To overcome the limitation of brief case summaries, Waterhouse (2016) investigated mock jurors' impressions of the credibility of a videotaped interview in which a child described a 4-minute recording of a theft she watched in an earlier session. Mock jurors were told either that the child provided the testimony they viewed in a single interview with a 10-minute refreshment break in the middle or in two interviews separated by a week (Waterhouse, 2016). In fact, mock jurors were presented with the exact same testimony in both conditions. In contrast to Yozwiak et al.'s (2004) results, Waterhouse (2016) found that mock jurors' credibility impressions were more favourable in the multiple interview condition, presumably due to jurors' belief that the delay between sessions in the multiple interview condition may explain contradictions in children's stories. However, mock jurors' verdict decisions were not affected by whether they thought they viewed a single interview, or two interviews.

Both Yozwiak et al. (2004) and Waterhouse (2016) took steps to increase the ecological validity of their research designs; Yozwiak et al. (2004) by presenting mock jurors with the court testimonies of additional witnesses and Waterhouse (2016) by showing participants the entire videotaped testimony of a real child. However, neither study reflected the tendency of multiple interviews to elicit overall more information than single interviews

due to reminiscence (Hershkowitz & Terner, 2007; Leander, 2010). In Study 1, children provided almost twice as much information about the allegations over the course of two interviews than in the first interview only. Although the extent of reminiscence was lower in Study 2, children who were interviewed a single time still disclosed less details about an event they participated in than participants interviewed on two separate occasions. Whilst negative impact of decreased consistency on the credibility of children's testimony has been widely studied (Connolly et al., 2008; Stromwall & Granhag, 2005; Waterhouse, 2016, Yozwiak et al., 2004) few studies addressed the question whether increased completeness may play a positive role (Kyrakidou, Zalaf, & Blades, 2014).

Consistency and Completeness

When assessing the credibility of accounts provided over multiple recall occasions, both mock jurors rely on inconsistencies as key signs of deception (Brewer, Potter, Fisher, Bond, & Luszcz; 1999; Connolly et al., 2008; Stromwall & Granhag, 2005). Brewer et al. (1999) found that undergraduate students presented with the task of assessing the credibility of stories chose inconsistencies as the strongest signs of deception. In contrast to Brewer et al.'s (1999) results, 67% of respondents to Quas et al.'s (2005) survey indicated that inconsistencies in a child's testimony do not necessarily indicate that the allegations are false. Waterhouse (2016) found that mock jurors found details mentioned in only the first interview more credible than consistent or reminisced information, providing further support for the view that inconsistencies in children's accounts do not inherently decrease their credibility.

The impact of the consistency of children's statements across interviews on their credibility in the eye of mock jurors is influenced by the type of event they experienced (Connolly et al., 2008). Noting that most children testifying about child abuse allege multiple instances of abuse, Connolly et al. (2008) compared adults' credibility judgements of the

accounts of children who experienced a single instance of an event with the recall of children who took part in multiple similar events across two phases of a single interview. Results showed that ratings of children's credibility were influenced by an interaction of children's age, the number of events they have experienced and the consistency of their accounts (Connolly et al., 2008).

The emphasis placed on inconsistency by mock jurors in some studies (Brewer et al., 1999; Stromwall & Granhag, 2005) is consistent with the 'Pinocchio hypothesis', which suggests that inconsistencies arising from the addition of new details are a key characteristic of fabricated accounts (Bruck, Ceci, & Hembrooke, 2002). Although Bruck et al. (2002) found evidence for the Pinocchio hypothesis when analysing children's true and false reports over five highly suggestive interviews, research on open-ended interviews with adults and children has shown that inconsistencies are frequently present in truthful accounts as well, and do not necessarily indicate that accounts are incorrect (Baugerud et al., 2014; Gilbert & Fisher, 2006; Krix et al., 2015). In contrast to the Pinocchio hypothesis, the 'parrot' or 'repeat versus reconstruct' hypothesis suggests that false accounts are generally more consistent than true ones, as liars are more motivated to perform 'impression management', and therefore actively avoid appearing inconsistent (Granhag, Stromwall, & Johnson, 2003). Stromwall & Granhag (2005) found evidence for this theory in a study involving 11-year-old children interviewed twice about either a real or an imagined event. Although no difference was found in the number of contradictions in the accounts of truth-tellers and liars, children who were asked to deceive evaluators omitted fewer previously mentioned details and added less new details to their stories in the second interview (Stromwall & Granhag, 2005).

The question of whether consistency is indicative of the truthfulness of testimonies is further complicated by Salykaly et al.'s (2013) results suggesting that the relationship between consistency and markers of deception may not remain stable across interviews, as

children's true descriptions of a game involving a touching element were more consistent than coached stories in the second interview, but this trend was reversed by the third interview. Whilst some studies have suggested that multiple interviews help distinguish between liars and truth-tellers through providing additional verbal cues of veracity (Ceci & Hembrooke, 2002; Granhag & Stromwall, 2005), others expressed concern that repeated recall occasions may make true and false accounts more similar, and therefore more difficult to discern between (Salykaly, et al., 2013). Stromwall & Granhag (2005) found that mock jurors' accuracy in telling true stories from false was not influenced by the number of interviews they observe, and their reliance on a step-by-step or an end-of-the-sequence method of evaluation.

Although the impact of the completeness of children's testimony on mock jurors' impressions has not been studied as widely as the relationship between consistency and credibility, research exploring the case factors underlying prosecutors' decision-making suggests that the number of details elicited in a forensic interview affects the credibility of the witness. Investigating prosecution rates in Cyprus, Kyrakidou et al. found that the only factor significantly affecting the likelihood that a case would be prosecuted was the amount of evidence provided (Kyriakidou, Zalaf, & Blades, 2014). Neither the types of questions asked by interviewers nor interviewer experience affected prosecution rates. In contrast to the results of studies focusing on consistency (Brewer et al., 1999; Stromwall & Granhag, 2005; Yozwiak et al., 2004), Kyriakidou et al.'s (2014) findings suggest that testimonies elicited over the course of multiple interviews appear more credible and may therefore be more likely to progress through the criminal justice system than single interviews due to the increase in the number of details reported.

The Present Study

The present study investigated the role of narrative coherence and multiple interviews in adult mock jurors' judgement of a child's testimony to assess whether the benefits of multiple interviews found in previous studies translate to a positive influence on children's credibility in the courtroom. In addition to comparing mock jurors' ratings of the credibility of a single interview with a vignette consisting of two interviews, the study also assessed whether children's age and the consistency of children's accounts may influence the relationship between multiple interviews and the perceived credibility of a child witness alleging sexual abuse.

In contrast to previous research that manipulated participants' beliefs regarding how a child's testimony was elicited while presenting them with the same material in each condition (Waterhouse, 2016; Yozwiak et al., 2004), in the present study, the child's testimony in the multiple interview condition included a more complete description of the allegations than the single interview condition. Field research shows that children tend to provide new information in the second interview, and new details emerging in successive interviews often have high forensic relevance (Katz & Hershkowitz, 2013; Waterhouse et al., 2016). In fact, Leander (2010) found that the number of details directly related to sexual abuse doubled in the second interview compared with the first interview. Similarly, in Study 1, the number of details reported across the two interviews was significantly higher than in the first interview only in most content categories, including crucial details related to time, location and subjective experiences. Furthermore, second interviews also included a higher proportion of sensitive details than first interviews. Thus, providing a more complete description of the allegations in the multiple interview condition than the single interview condition has higher ecological validity than matching the conditions according to the number of details.

Based on previous research (Brewer et al., 1999; Kyriakidou et al., 2014; Waterhouse, 2016, Yozwiak et al., 2004), it was predicted that 1) when interviews are consistent, mock

jurors would rate witnesses in the multiple interview condition as more credible, competent and reliable than in the single interview condition and would be more likely to indicate that the defendant is guilty due to the increased completeness of children's accounts, 2) when interviews are inconsistent, mock jurors would rate witnesses in the multiple interview condition as less credible, competent and reliable than in the single interview condition and would be less likely to indicate that the defendant is guilty due to the increased emphasis on consistency in multiple interviews. It was further expected that 3) mock jurors would rate consistent witnesses in both the single and multiple interview condition as more credible, competent and reliable and would be more likely to choose the guilty verdict.

Following the 2-factor model of the relationship between witnesses' age and credibility (McCauley & Parker, 2001), it was predicted that 4) mock jurors would rate younger children as more reliable and credible, whilst they would perceive older children as more competent. Due to the nature of the case placing emphasis on reliability rather than competent, it was expected that mock jurors would be more likely to convict the defendant when the witness was younger.

Finally, based on previous findings showing a relationship between juror characteristics and credibility ratings (Bottoms & Goodman, 1994; McCauley & Parker, 2001; Wessel et al., 2013), it was predicted that 5) female jurors would rate children's credibility, competency and reliability higher and would be more likely to indicate that the defendant is guilty than male jurors.

4.2 Method

Participants

Following a power calculation, 127 participants were recruited using social media and flyers placed in university and library buildings in Egham and London. Twenty participants

did not complete the questionnaire, and one participant did not provide informed consent. All data from these participants were excluded. The final sample of 106 respondents included 67 women (63.2%) and 39 men (36.8%). The majority of participants were under the age of 35 (81.1%), with 50 in the 18-24 age category (47.2%), 36 in the 25-34 age category (34%), 11 in the 35-44 age category (10.4%), 6 in the 45-54 age category (5.7%), 1 in the 55-64 age category (0.9%) and 2 in the 65-74 age category (1.9%).

Most respondents lived in the United Kingdom (65.1%), with the majority located in England (60.4%). Over half of participants (54.7%) had experience working with children, with 26 participants having worked with children for less than a year (24.5%), 14 for over one year (13.2%) and 18 for three years or more (17%). Participants worked in a wide range of professions involving regular contact with children, including teachers, childminders, social workers and camp counsellors. Most respondents had no direct experience of the legal system from the perspective of the jurors, with only one participant having performed jury duty in the past.

Materials

The experiment relied on a 2x2x2 experimental design, where the independent variables were interview number (single or multiple), interview consistency (high or low) and the age of the witness (8-year-old or 12-year-old). Participants were presented with an online questionnaire created using Qualtrics. The questionnaire consisted of an information sheet (Appendix F), a demographic questionnaire (Appendix G), a forensic vignette (Appendix H) and a series of questions regarding the vignette (Appendix I). Participants indicated their informed consent before taking part in the study. The demographic questionnaire included questions regarding participants' age, nationality, and gender. Participants were also asked

whether they had performed jury duty in the past and whether they had experience working with children.

In the main part of the questionnaire, participants were randomly assigned to read 1 of 8 vignettes of a fictional forensic interview with a child witness alleging sexual abuse, created using typical case patterns from the forensic transcripts analysed for Study 1. The child witness alleged that her mother's boyfriend kissed her and touched her genitals on two different occasions. The interviewer used open-ended questions consistent with NICHD guidelines (Lamb et al., 2009). Perceived witness age was manipulated by changing the introductory comment made by the interviewer to state that she is interviewing an 8-year-old or a 12-year-old. Interview number was manipulated by presenting participants with the transcript of either a single forensic interview or two forensic interviews separated by 3 months. The first interview was identical in both cases. In the multiple interview condition, the second interview included a combination of repeated and new details, consistently with patterns in real-life multiple forensic interviews revealed in Study 1. Consistency was manipulated by adding two contradictions to the child's testimony in the inconsistent condition, one related to the location where the alleged abuse took place and one related to the child's understanding of touching under or over clothes. To make sure participants are able to identify the inconsistencies, the interviewer pointed out these contradictions and asked the child to provide clarification. In the single interview condition both inconsistencies were in the first interview, whilst the multiple interview condition included one inconsistency in each interview.

After reading the interview or interviews, participants were asked to answer a series of questions about their perceptions of the case. To test the 2-factor model of children's credibility, the questionnaire was designed to include questions relating to children's reliability, competency and credibility. Participants then rated a range of factors in terms of

their importance in influencing perceptions of reliability, competency and credibility. These factors included 1) the child's verbal competency, 2) the amount of detail in the child's statement, 3) the consistency of the child's statements, 4) interviewing style, 5) logical presentation of story elements from start to finish, 6) the child's age and 7) the plausibility of the situation described by the child. At the end of the questionnaire, participants were asked to provide a binary convict/acquit judgement regarding the defendant's guilt, based on the child's testimony only, noting that the defendant denies the allegation of sexual abuse and no medical or physical evidence is available. In addition to the previous factors, participants also had to indicate the importance of 8) the defendant's denial of the abuse and 9) the lack of medical evidence in reaching their verdict, and to provide a confidence estimate for their verdict.

Upon completing the questionnaire, participants were thanked for their contribution and invited to take part in an optional prize draw.

4.3 Results

Preliminary Analyses

Conditions were well-balanced; 13 respondents were presented with the single, young, consistent condition (12.3%), 14 were presented with the single, old, consistent condition (13.2%), 13 with the single, young, inconsistent condition (12.3%), 12 with the single, old, inconsistent condition (11.3%), 13 with the repeated, young, consistent condition (12.3%), 12 with the repeated, old, consistent condition (11.3%), 14 with the repeated, young, inconsistent condition (13.2%) and 15 with the repeated, old, inconsistent condition (14.2%). There were no significant differences in the distribution of gender across conditions, $\chi^2(1) = .23$, $p = .63$.

Verdict

Descriptively, a higher percentage of female jurors chose the guilty verdict (83.6%) than of male jurors (62.5%) (Table 23). Similarly, more jurors who read both interviews (82.5%) indicated that the defendant was guilty than those who read the first interview only (65.4%) (Table 24). Although children's age seemed to have influenced jurors' decisions to a smaller extent than jurors' gender and interview number, more jurors who read the testimony of an 8-year-old chose the guilty verdict (81.1%) than those who read the testimony of a 12-year-old (69.8%) (Table 25). A slightly higher percentage of jurors chose a guilty verdict in the inconsistent condition (79.6%) than in the consistent (71.2%) condition (Table 26).

Table 23 *Verdict Choices and Confidence Ratings of Male and Female Jurors*

	Guilty		Not guilty	
	Chose verdict	Confidence	Chose verdict	Confidence
	N (%)	*M (SD)*	*N (%)*	*M (SD)*
Male	24 (62.5%)	6.38 (1.13)	15 (38.5%)	3.87 (2.26)
Female	56 (83.6%)	7.07 (2.18)	11 (16.4%)	2.63 (1.91)

Table 24 *Verdict Choices and Confidence Ratings of Participants in the One Interview and Two Interviews Conditions*

	Guilty		Not guilty	
	Chose verdict	Confidence	Chose verdict	Confidence
	N (%)	*M (SD)*	*N (%)*	*M (SD)*
One interview	34 (65.4%)	6.50 (2.22)	18 (34.6%)	3.44 (2.12)
Two interviews	46 (85.2%)	7.13 (1.69)	8 (14.8%)	3.13 (2.42)

Table 25 Verdict Choices and Confidence Ratings of Participants in the Consistent and Inconsistent Conditions

	Guilty		Not guilty	
	Chose verdict	Confidence	Chose verdict	Confidence
	N (%)	*M (SD)*	*N (%)*	*M (SD)*
Consistent	37 (71.2%)	7.11 (2.04)	15 (28.8%)	2.67 (2.02)
Inconsistent	46 (79.6%)	6.65 (1.86)	11 (20.4%)	4.27 (2.10)

Table 26 *Verdict Choices and Confidence Ratings of Participants in the 8-year old and 12-year-old Witness Conditions*

	Guilty		Not guilty	
	Chose verdict	Confidence	Chose verdict	Confidence
	N (%)	*M (SD)*	*N (%)*	*M (SD)*
8-year-old	43 (81.1%)	6.98 (2.10)	10 (18.9%)	4.00 (2.31)
12-year-old	37 (69.8%)	6.73 (1.77)	16 (30.2%)	2.94 (2.05)

A binary logistic regression analysis was conducted to assess whether juror's gender (female, male), interview number (one, two), witness consistency (consistent, inconsistent) and witness age (8-year-old, 12-year-old) predicted the verdict (guilty, not guilty). The model was significantly more effective than the null model, $B = 1.12$, $S.\ E. = .23$, $\chi^2(1) = 24.79$, $p < .001$, $e^B = 3.08$. Juror's gender, $B = 1.18$, $S.\ E. = .50$, $\chi^2(1) = 5.67$, $p = .017$, $e^B = 3.26$, 95% CI [1.32, 8.61] and interview number $B = 1.18$, $S.\ E. = .51$, $\chi^2(1) = 5.38$, $p = .02$, $e^B = 3.25$, 95% CI [1.20, 8.78] were significant predictors of verdict. Witness consistency and witness age

did not significantly predict the verdict. The Hosmer-Lemeshow test suggested that the model fitted the data well, $\chi^2(7) = 6.85$, $p = .45$. Overall, the model predicted 78.3% of verdict choices correctly. Specifically, 96.3% of guilty verdicts, and 23.1% of not guilty verdicts were predicted correctly. Results of all statistical tests relating to verdict category are presented in Table 27.

Table 27 *Summary of Analyses of Verdict choice*

Predictor	*B*	*S.E.*	*Wald's* χ^2	*df*	*p*	e^B	*CI*
Juror's gender	1.18	.49	5.67	1	.02	3.26	[1.23,8.61]
Interview number	1.18	.51	5.38	1	.02	3.25	[1.20,8.78]
Witness consistency	.23	.49	.22	1	.64	1.26	[.48,3.31]
Witness age	-.71	.50	2.04	1	.15	.49	[.19,1.30]

Confidence in the Verdict Chosen

Overall, mock jurors were more confident in their verdict decisions when they chose the guilty verdict (M = 6.86, SD = 1.95) than when they chose the not guilty verdict (M = 3.35, SD = 2.17). On average, mock jurors felt more confident in guilty verdicts when they read both interviews with the witness (Table 24) and when the interviews did not include contradictory information (Table 25). Conversely, participants were more confident in not guilty verdicts when they read a single interview and when the interview included inconsistencies. Mock jurors expressed more confidence in both their guilty and not guilty verdicts when they read the testimony of an 8-year-old than when they read the testimony of a 12-year-old (Table 26). Finally, female jurors were more confident than male jurors when a

guilty verdict was chosen, but males were more confident than females when a not guilty verdict was chosen (Table 23).

To assess whether witness and juror characteristics and the verdict chosen affected mock jurors' confidence in the verdict, a five-way ANOVA was conducted assessing the effects of verdict category (guilty, not guilty), juror's gender (male, female), interview number (one, two), consistency (consistent, inconsistent) and witness age (8-year-old, 12-year-old) on mock jurors' confidence ratings (1 = not at all confident, 10 = fully confident).

Jurors were significantly more confident when choosing a guilty verdict, $F(1,78) = 32.85$, $p < .001$, $\eta_p^2 = .30$. No main effect was found for juror's gender, $F(1,78) = 1.79$, $p = .18$, $\eta_p^2 = .02$, interview number, $F(1,78) = .004$, $p = .95$, $\eta_p^2 < .001$, consistency, $F(1,78) = 2.31$, $p = .13$, $\eta_p^2 = .03$, and witness age, $F(1, 78) = .04$, $p = .85$, $\eta_p^2 < .001$. There were no significant interactions.

Jurors' Justifications of Their Verdict Decisions

Descriptively, jurors rated the importance of the amount of detail in the witness statement, the consistency of the witness statement, the types of questions asked by the interviewer, the logical/chronological coherence of the story from start to end and the plausibility of the situation described by the witness were as important in influencing their verdict decisions. In contrast, the age and verbal competence of the child were perceived as less important. Additionally, information regarding the defendants' denial of the case and the lack of medical evidence were rated as less important than characteristics of the witness and the interview.

Table 28 *The Influence of Case Factors on Jurors' Verdict Decisions*

	Perceived importance	
	M	*SD*
Child's verbal competence	4.82	2.57
Amount of detail	6.44	2.35
Consistency of statements	6.73	2.54
The types of questions asked	6.46	2.56
Coherent story	5.91	2.60
Child's age	5.15	2.94
Plausibility of situation	6.73	2.42
Defendant's denial	3.65	2.59
No medical evidence	4.63	3.18

To assess the perceived importance of witness and interview characteristics on mock jurors' verdict choices, a RM-ANOVA was conducted assessing the effect of factor type (child's verbal competence, amount of detail, consistency of statements, type of questions asked, coherence of story, child's age, plausibility of situation) on the perceived importance witness and interview characteristics when choosing the verdict. Mauchley's test indicated that the assumption of sphericity was violated for factor type, $\chi^2(35) = 120.97$, $p < .001$. Consequently, a Greenhouse-Geisser corrections was applied ($\varepsilon = .75$).

There was a significant main effect of factor type, $F(6.03, 633.14) = 24.74$, $p < .001$, $\eta_p^2 = .19$. The main effect of factor type was followed up by 21 paired t-tests tests (adjusted $\alpha = .0014$). The amount of detail in the witness statement, the consistency of the witness statement, the type of question asked, the coherence of the story and the plausibility of the situation described by the child were deemed significantly more important than the defendant's denial and the lack of medical evidence. The amount of detail in the witness statement, the consistency of the witness statement, the type of question asked, and the plausibility of the situation were more important than the child's age and the child's verbal competence. The consistency of the witness statement was more important than the coherence of the story. The child's age was more important than the defendant's denial. There were no other significant differences. All pairwise comparisons are presented in Table 29.

Table 29 *Summary of Analyses of the Perceived Importance of Case Factors on Verdict Choices*

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Main effect	Factor Type	6.03, 633.14	$F = 24.74$	<.001*	$\eta_p^2 = .19$	
Follow up	Verbal Competence vs. Amount of detail	105	$t = -6.34$	<.001**	$d = .66$	[-2.13,-1.12]
Follow up	Verbal competence vs. Consistency	105	$t = -7.00$	<.001**	$d = .75$	[-2.45,-1.36]
Follow up	Verbal competence vs. Question type	105	$t = -5.64$	<.001**	$d = .64$	[-2.22,-1.06]
Follow up	Verbal competence vs. Coherence	105	$t = -3.87$	<.001**	$d = .43$	[-1.66,-.53]
Follow up	Verbal competence vs. Age	105	$t = -1.20$	=.23	$d = .12$	[-.87,.21]
Follow up	Verbal competence vs. Plausibility	105	$t = -6.68$	<.001**	$d = .77$	[-2.47,-1.34]
Follow up	Verbal competence vs. Denial	105	$t = 3.82$	<.001**	$d = .43$	[.56,1.77]
Follow up	Verbal competence vs. Medical evidence	105	$t = .55$	=.59	$d = .07$	[-.50,.87]

(continued)

(continued)

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Follow up	Amount of detail vs. Consistency	105	*t* = -1.30	=.20	*d* = .12	[-.72,.15]
Follow up	Amount of detail vs. Question type	105	*t* = -.06	=.95	*d* = .01	[-.64,.60]
Follow up	Amount of detail vs. Coherence	105	*t* = 2.01	=.05	*d* = .21	[.01,1.05]
Follow up	Amount of detail vs. Age	105	*t* = 4.55	<.001**	*d* = .48	[.73,1.86]
Follow up	Amount of detail vs. Plausibility	105	*t* = -1.13	=.26	*d* = .12	[-.78,.21]
Follow up	Amount of detail vs. Denial	105	*t* = 8.39	<.001**	*d* = 1.07	[2.12,3.45]
Follow up	Amount of detail vs. Medical evidence	105	*t* = 5.27	<.001**	*d* = .65	[.34,1.13]
Follow up	Consistency vs. Question type	105	*t* = .97	=.33	*d* = .11	[-.27,.80]
Follow up	Consistency vs. Coherence	105	*t* = 3.42	=.001**	*d* = .32	[.34,1.28]
Follow up	Consistency vs. Age	105	*t* = 5.02	<.001**	*d* = .58	[.95,2.20]
Follow up	Consistency vs. Plausibility	105	*t* <.001	=1.00	*d* = 0	[-.49,.49]

(continued)

(continued)

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Follow up	Consistency vs. Denial	105	*t* = 8.93	<.001**	*d* = 1.14	[2.39,3.76]
Follow up	Consistency vs. Medical evidence	105	*t* = 5.56	<.001**	*d* = .73	[1.34,2.84]
Follow up	Question type vs. Coherence	105	*t* = 1.91	=.06	*d* = .21	[-.02,1.12]
Follow up	Question type vs. Age	105	*t* = 3.80	<.001**	*d* = .48	[.63,1.99]
Follow up	Question type vs. Plausibility	105	*t* = -.96	=.34	*d* = .11	[-.81,.28]
Follow up	Question type vs. Denial	105	*t* = 8.30	<.001**	*d* = 1.04	[2.14,3.48]
Follow up	Question type vs. Medical evidence	105	*t* = 4.72	<.001**	*d* = .63	[1.06,2.60]
Follow up	Coherence vs. Age	105	*t* = 2.11	=.04	*d* = .28	[.05,1.48]
Follow up	Coherence vs. Plausibility	105	*t* = -2.88	=.005	*d* = .32	[-1.37,-.25]
Follow up	Coherence vs. Denial	105	*t* = 6.61	<.001**	*d* = .83	[1.58,2.94]
Follow up	Coherence vs. Medical Evidence	105	*t* = 3.38	=.001**	*d* = .44	[.53,2.04]
Follow up	Age vs. Plausibility	105	*t* = -5.50	<.001**	*d* = .59	[-2.14,-1.01]
Follow up	Age vs. Denial	105	*t* = 4.48	<.001**	*d* = .52	[.86,2.16]

(continued)

(continued)

Analysis			Results			
Type	Variables	Df	test value	*p*	effect size	95% CI
Follow up	Age vs. Medical evidence	105	*t* = 1.37	=.17	*d* = .17	[-.23,1.27]
Follow up	Plausibility vs. Denial	105	*t* = 9.41	<.001**	*d* = 1.16	[2.43,3.72]
Follow up	Plausibility vs. Medical evidence	105	*t* = 5.61	<.001**	*d* = .74	[1.35,2.83]
Follow up	Denial vs. Medical evidence	105	*t* = -3.18	=.002	*d* = .32	[-1.59,-.37]

Note: * denotes significance at $p < .05$ ** denotes significance at $p < .001$

Perceived Reliability, Competence and Credibility ratings

Descriptively, jurors rated the reliability, competence and credibility of the 8-year-old witness higher than the 12-year-old witness (Table 30). Jurors in the consistent condition gave higher reliability, competence and credibility scores than jurors in the inconsistent condition (Table 31). Jurors who read both interviews rather than the first interview only rated the reliability, competence and credibility of the witness higher (Table 32). Finally, males gave lower scores than females when rating the reliability, competence, and credibility of the child witness (Table 33).

Table 30 *The Perceived Reliability, Competence and Credibility of the 8-year old Witness and the 12-year-old Witness*

	8-year-old		12-year-old	
	M	*SD*	*M*	*SD*
Reliability	7.47	1.89	7.21	1.94
Competence	7.40	2.20	6.79	2.07
Credibility	7.96	1.78	7.49	1.85

Table 31 *The Perceived Reliability, Competence and Credibility of the Witness in the Consistent and Inconsistent Conditions*

	Consistent		Inconsistent	
	M	*SD*	*M*	*SD*
Reliability	7.54	1.63	7.15	2.14
Competence	7.42	1.98	6.78	2.26
Credibility	8.04	1.87	7.43	1.73

Table 32 *The Perceived Reliability, Competence and Credibility of the Witness in the One Interview and Two Interviews Conditions*

	One		Two	
	M	*SD*	*M*	*SD*
Reliability	6.94	1.91	7.72	1.84
Competence	6.77	2.09	7.41	2.17
Credibility	7.56	1.76	7.89	1.87

Table 33 *Male and Female Jurors' Perceived Reliability, Competence and Credibility Ratings of The Witness*

	Male		Female	
	M	*SD*	*M*	*SD*
Reliability	6.56	1.97	7.79	1.73
Competence	6.44	2.19	7.48	2.04
Credibility	6.95	2.03	8.18	1.53

Reliability

To assess whether witness and juror characteristics affected the perceived reliability of the witness, a four-way ANOVA was conducted assessing the effects of juror's gender (male, female), interview number (one, two), consistency (consistent, inconsistent) and witness age (8-year-old, 12-year-old) on the reliability scores given by mock jurors (1 = not at all reliable, 10 = fully reliable).

Female jurors rated the reliability of the witness significantly higher than male jurors, $F(1,90) = 13.56$, $p < .001$, $\eta_p^2 = .13$ and reliability ratings were significantly higher in the two interviews condition, $F(1,90) = 6.43$, $p = .01$, $\eta_p^2 = .07$. No main effect was found for consistency $F(1,90) = 3.56$, $p = .06$, $\eta_p^2 = .04$, and witness age, $F(1, 90) = .46$, $p = .50$, $\eta_p^2 = .01$. There were no significant interactions.

Competence

To assess whether witness and juror characteristics affected the perceived competence of the witness testimony, a four-way ANOVA was conducted assessing the effects of juror's gender (male, female), interview number (one, two), consistency (consistent, inconsistent) and witness age (8-year-old, 12-year-old) on the competence scores given by mock jurors (1 = not at all competent, 10 = fully competent).

Female jurors rated the competence of the witness significantly higher than male jurors, $F(1,90) = 32.76$, $p = .007$, $\eta_p^2 = .08$. Competence ratings were significantly higher in the two interviews condition, $F(1,90) = 20.34$, $p = .03$, $\eta_p^2 = .05$, and in the consistent condition, $F(1,90) = 20.49$, $p = .03$, $\eta_p^2 = .05$. No main effect was found for witness age, $F(1, 90) = 3.35$, $p = .07$, $\eta_p^2 = .04$. A significant interaction was found between juror's gender and interview number, $F(1, 90) = 4.89$, $p = .03$, $\eta_p^2 = .05$, which was further explored using post-hoc tests. There were no other significant interactions.

The significant interaction between juror's gender and interview number was followed up by two independent samples t-tests (adjusted $\alpha = .025$). In the single interview condition, females rated the competence of the witness significantly higher ($M = 7.47$, $SD = 1.87$) than males ($M = 5.65$, $SD = 1.98$), $t(50) = 3.34$, $p = .002$, 95% CI [.72,2.91]. However, when jurors read two interviews with the child witness, there was no difference in the

competence ratings of females (M = 7.49, SD = 2.21) and males (M = 7.63, SD = 2.13), $t(52) = .36$, $p = .72$, 95% CI [1.02,1.47].

Credibility

To assess whether witness and juror characteristics affected the perceived credibility of the witness, a four-way ANOVA was conducted assessing the effects of juror's gender (male, female), interview number (one, two), consistency (consistent, inconsistent) and witness age (8-year-old, 12-year-old) on the credibility scores given by mock jurors (1 = not at all credible, 10 = fully credible).

Female jurors rated the competence of the witness significantly higher than male jurors, $F(1,90) = 15.33$, $p < .001$, $\eta_p^2 = .15$ and competence ratings were higher in the consistent condition than in the inconsistent condition, $F(1,90) = 5.63$, $p = .02$, $\eta_p^2 = .06$. No main effect was found for interview number $F(1,90) = 1.41$, $p = .24$, $\eta_p^2 = .02$, and witness age, $F(1, 90) = 2.48$, $p = .12$, $\eta_p^2 = .03$. There were no significant interactions.

4.4 Discussion

Overall, findings supported most of the hypotheses, but also revealed previously unexplored relationships between case characteristics and mock jurors' impressions of the credibility of child witnesses alleging sexual abuse.

As expected, mock jurors who read two interviews with the witness were more likely to choose a guilty verdict than those who read a single interview when the interviews were consistent. However, contrary to expectations, the advantage of two interviews applied to the inconsistent condition as well. Predictions regarding the influence of the age of the witness on verdict decisions and mock jurors' impressions were not supported by the findings. Witness age did not affect participants' ratings of the reliability, competence and credibility of the witness or the percentage of mock jurors who chose a guilty versus a not guilty verdict.

Juror characteristics had the largest and most pervasive effect on verdict decisions and credibility ratings, as female jurors gave higher reliability, competence and credibility ratings to witnesses than male jurors and were also more likely to choose a guilty verdict.

Overall, jurors were more confident in guilty verdicts than not guilty verdicts. Descriptively, in conditions where children's credibility was rated higher, jurors felt more confident in their guilty verdicts than in conditions where children's credibility was rated lower, but this effect did not reach significance. The perceived importance of case characteristics, indicated by mock jurors, were remarkably consistent with mock jurors' actual verdict decisions. In particular, mock jurors indicated that the age and verbal competency of witnesses was less important to them than other case characteristics, and accordingly, the age of the witness did not significantly affect their credibility ratings or verdict decisions. According to self-report, the case characteristics that jurors found most important were the type of questions asked by the interviewer, the consistency of statements in the interviews, the amount of detail reported by the witness and the plausibility of the situation described by the witness.

Interview Number and Consistency

Following jurors' gender, interview number was the strongest predictor of mock jurors' verdict decisions and credibility judgements. Consistent with Waterhouse's (2016) findings, multiple interviews resulted in more positive perceptions of the child witness, even if recall over the two interviews included contradictions. Jurors in the two-interviews condition gave higher competence and reliability ratings to the child witness than jurors in the one interview condition. In contrast to Waterhouse (2016), interview number also affected verdict decisions, as a larger percentage of jurors who read both interviews gave a guilty verdict than of those who read a single interview. Although the consistency of the witness

statements did not affect the verdict, jurors did appear to be sensitive to differences in consistency, as they rated the credibility and competence of the child witness higher in the consistent condition than in the inconsistent condition.

The present results do not support Yozwiak et al.'s (2004) findings suggesting that increased disclosure over the course of two interviews leads to a decrease in the credibility of child witnesses. The contrasting results are partially explained by differences in the research methods of the present experiment and Yozwiak et al.'s (2004) study. Yozwiak et al.'s (2004) participants were presented with the same information in both conditions, with the only difference being whether the full extent of information was disclosed in the first or second interview. In contrast, participants in the present experiment were either presented only with children's partial disclosure provided in the first interview or the full disclosure gathered over the course of the two interviews.

Following tendencies found in field studies, the key difference between the single and multiple interview conditions was in the amount of detail children reported (Hershkowitz & Terner, 2007; Leander, 2010; Waterhouse et al., 2016). Overall, results are consistent with a narrative coherence framework, as they suggest that an increase in completeness or consistency affects witness credibility positively, whilst a decrease in either completeness or consistency compromises credibility. However, in this study, consistency and completeness were not weighed equally by mock jurors, as a more complete account was judged more credible than a less complete account, even if it was less consistent. When asked about factors that influenced their decision-making, jurors rated both the amount of details reported by the witness and the consistency of the witness statement as highly important. This is consistent with previous research showing that jurors rely on inconsistencies when assessing children's credibility (Brewer et al., 1999; Bond, & Luszcz; 1999; Connolly et al., 2008; Stromwall & Granhag, 2005).

Jurors' Gender

In accordance with previous research (Bottoms & Goodman, 1994; Davies & Rogers, 2009; McCauley & Parker, 2001; Wessel et al., 2013) jurors' gender emerged as the strongest predictor of mock jurors' verdict decisions and their impressions of the witness. A higher percentage of female jurors convicted the defendant than male jurors, and female jurors' ratings of witness reliability, competence and credibility were higher than male jurors' ratings. However, despite these differences, both male and female jurors were more likely to convict the defendant than not to and both groups felt overall more confident in guilty verdicts than not guilty verdicts.

Research has indicated that gender difference in verdict choice might decrease as a result of deliberation (Crowley, O'Callaghan, & Ball, 1994; Gabora, Spanos, & Joab, 1993). Using gender-balanced jury panels, Crowley et al. (1994) found that while female jurors were more likely to choose a guilty verdict in a child sexual abuse case, the difference between male and female jurors' verdict preferences became narrower post-deliberation. Although the current study did not include a deliberation component, it is possible that mock jurors' confidence ratings provide an indication of how likely they are to be persuaded by arguments for a different verdict decision. As jurors of both genders felt more confident in guilty verdicts than not guilty verdicts, the levelling of gender differences post-deliberation could lead to a guilty consensus in this case. However, a range of other factors could play a role in group decision-making during deliberation, such as relative dominance within the group, the persuasiveness of jurors' arguments and jurors' understanding of the concept of reasonable doubt (Crowley et al., 1994; Gabora et al., 1993).

Considering the pervasiveness of the gender difference in verdict preference across sexual assault cases involving child witnesses (Bottoms & Goodman, 1994; Crowley et al.,

1994; Davies & Rogers, 2009; Gabora et al., 1993; McCauley & Parker, 2001; Wessel et al., 2013), the potential influence of the evaluator's gender must be evaluated in all situations where decisions are made regarding the extent to which children's accounts are believable. This implies not only that the gender distribution of juries needs to be considered in cases involving children alleging sexual abuse, but also that the gender of individuals making credibility decisions could be influential earlier in the investigative process, such as when investigators need to determine whether a case is worthy of an official investigation or suitable to refer for prosecution.

Previous research by Lamb & Garretson (2003) has revealed an interaction between the gender of the interviewer and the gender of the child witness that affected the quality of children's testimony. It is possible that such an interaction exists between the gender of mock jurors and the gender of the child witness as well, although this has not been explored. In the present study, mock jurors were intentionally left uncertain about the gender of the child witness, who had a gender-neutral name that is commonly given to both boys and girls in the United Kingdom. However, several participants commented that they immediately assumed that the witness was a girl, reflecting gender differences in rates of reported child sexual abuse (Radford et al., 2011). Thus, it is possible that female jurors' increased trust of the child witness in this study was related to their ability to empathise with a child they presumed to be female.

Witness Age

In this study, the 'believed' age of the witness was manipulated, with participants led to believe that they are reading the testimony of either an 8-year-old or a 12-year-old. The 'actual' age of the witness was not manipulated, as participants read the exact same fictional account in both conditions. Overall, witness age had no significant effect on jurors' choice of

verdict, their confidence in the verdict chosen or their perception of the reliability, competence and credibility of the witness. Remarkably, jurors appeared to consciously avoid the age of the child witness influencing their credibility decisions, as they rated the age and verbal competence of the witness as the least important factors affecting their verdict decisions and credibility judgements. Their results are consistent with McCauley & Parker's (2001) findings showing that the believed age of the witness did not affect participants' impressions of their credibility.

The lack of influence of witness age on participants' verdict decisions and credibility ratings may appear encouraging, as it demonstrates that participants made decisions based on the actual testimony they were presented with, rather than age-related stereotypes. The present results provide further support for Luus et al.'s (1995) findings suggesting that the manner in which children narrated their testimony predicted how credible they appeared to mock jurors, rather than children's age per se. Accordingly, although the results of Study 1 and previous research (Baugerud et al., 2014; Hubbard et al., 2016; Peterson, 2011) showed age differences in the accuracy and completeness of children's testimonies, Study 2 demonstrated that large individual differences exist within age groups.

Limitations and Further Research

This study was the first to assess the impact of completeness and consistency in multiple interviews on children's credibility, but some questions regarding the relationship between narrative coherence and credibility remain open to future research.

Firstly, whilst some previous research used short vignettes to describe cases (Tabak & Klettke, 2014; Yozwiak et al., 2004), the current study used fictional verbatim forensic interviews, developed based on the results of field research on actual forensic interviews with child witnesses alleging sexual and physical abuse. The decision to use full interviews rather

than case descriptions led to higher ecological validity, however, it may have affected mock jurors' ratings in several ways. Firstly, some case characteristics that were previously shown to influence jurors' perception of the case, such as the child's verbal skills (Luus et al., 1995), the child's emotional expression (Wessel et al., 2013), the interview style (Castelli et al., 2005) and the identity of the suspect (Castelli et al., 2005) remained identical across conditions and may have led to a generally more positive or negative impression of the case.

Secondly, in contrast to previous research (Waterhouse, 2016; Yozwiak et al., 2004), the multiple interview condition in the present study included a higher number of details than the single interview condition. Whilst this decision reflects the general tendencies found in field research on multiple interviews, (Katz & Hershkowitz, 2013; Leader, 2010; Waterhouse et al., 2016), it also means that findings cannot be generalised to interviews where children do not report new forensically relevant details in the second interview. Furthermore, changes between conditions might not have been comparable in importance, as the amount of details reported differed more dramatically between conditions than the number of inconsistencies. The two-interviews condition included 19 more details than the single interview condition, whilst the inconsistent condition included only 2 contradictions, although these affected details with high forensic relevance.

Although care was taken to render the interviews as realistic as possible, they differed from real-life forensic interviews in two key aspects. Firstly, to avoid inattention and drop-out by mock jurors, the interviews were much shorter than the average forensic interview in Study 1. Secondly, previous research has shown that despite recommendations, forensic interviews conducted in the United Kingdom still include a high proportion of closed-ended and leading questions (Waterhouse et al., 2016). In contrast, the interview depicted in the present study involved open-ended questions only, as the emphasis was on revealing the impact of witness and juror characteristics on the credibility of child witnesses, rather than on

interview techniques. Mock jurors in the present study also differed from jurors in real-life trials, primarily due to the skewed age distribution of participants. More than 80% of participants were younger than 35-years-old, therefore, the study was not able to explore potential differences in perception between older and younger generations.

Future studies relying on different experimental designs could shed light on remaining questions regarding the influence of jurors' gender, witness age and interview number on the perceived credibility of child witnesses. Mock jurors in the current study were asked to rate their impressions of the case individually, thus, the study did not involve a deliberation stage. With previous research suggesting that the gender difference between male and female jurors' ratings of witness credibility is decreased post-deliberation (Crowley et al., 1994; Gabora et al., 1993), the inclusion of the deliberation stage in future studies, through a focus group methodology (e.g. Tabak & Klettke, 2014), would be highly valuable. Participants in the current study were led to believe that the child witness was either 8 years old or 12 years old. The inclusion of a fictional witness of pre-school age or an older teenager may be advisable in future studies, as the reliability and competence of these groups may be more compromised in the eyes of the jury, than that of witnesses in middle childhood. Finally, participants in the current study were presented with two interviews with the witness, however, as Study 1 revealed, many witnesses in the United Kingdom are interviewed more than twice, with a significant proportion of witnesses interviewed 3 or more times. No previous study has investigated the impact of multiple interviewing protocols involving more than two sessions on the credibility of child witnesses in the eye of a mock jury.

Finally, the current study only investigated two of the three components of narrative coherence: completeness and consistency. The third component, connectedness or the extent to which the testimony is connected through causal and temporal links, has not been impacted by multiple interviews in Study 1 and Study 2, therefore, it was not included in the current

study. However, when participants were asked to rate the importance of a range of factors in influencing their credibility ratings, causal-temporal connectedness was rated relatively high, above children's age and verbal competence. Future research could investigate whether the self-reported importance of this factor translates into actual differences in participants' verdict decisions and credibility assessments.

Conclusion

This study was the first to explore the relationship between multiple interviews, witness age and witness consistency in influencing mock jurors' judgements of the credibility of child witnesses. Results suggest that multiple interviews have an overall positive effect on the credibility of child witnesses, as participants who read two interviews rated the credibility of a fictional child witness higher and were more likely to choose a guilty verdict. These findings are consistent with a narrative coherence framework, in that an increase in either completeness or consistency resulted in more positive impressions of the child witness, whereas a decrease in either component compromised the credibility of the child.

Considered together with the benefits of multiple interviews revealed in Study 1 and Study 2, the use of multiple interviews may be encouraged in some investigations, as a second interview increases the completeness of children's testimony without compromising their credibility. However, in cases when the delay between interviews is long, there may be concerns about inflation of credibility through details gathered in multiple interviews when doubt is cast on the accuracy on those details.

Jurors' gender was the strongest predictor of verdict decisions and credibility ratings, suggesting that research investigating juror characteristics is as important as studies on characteristics of the case and the witness in illuminating the underlying aspects of courtroom decisions. The present study did not provide support for the view that jurors evaluate

testimonies differently based on the ‘believed’ age of the witness. Future research is needed to establish whether these tendencies apply to witness testimonies provided over the course of three or more interviews as well, and whether the deliberation phase of trials may diminish the influence of witness characteristics on the verdict.

Chapter 5. Implications and Conclusion

The aim of this thesis was to investigate the effect of multiple interviews on children's testimony, using narrative coherence as the measure of the quality of children's reports. The goals were thus twofold: to explore narrative coherence as an overarching, forensically relevant measure of the quality of children's testimony, and to assess the risks and benefits of multiple interviewing in the field, and under experimental conditions. This chapter will consider the implications of the results of Study 1, Study 2 and Study 3 on the prevailing view of the impact of multiple interviews on children's memory and the credibility of their testimony. Following the discussion of the key findings, the suitability of the present model of narrative coherence for research in the field of investigative interviewing, and the strengths and limitations of the methodologies of the three studies will be assessed. Finally, currently used guidelines for conducting multiple interviews with child witnesses will be evaluated based on the present findings and the literature review.

5.1 The Impact of Multiple Interviews on the Narrative Coherence, Accuracy and Credibility of Children's Testimony

Multiple interviews have been the focus of considerable attention and controversy for several decades. Previous systematic reviews of research on multiple interviews (Goodman & Quas, 2008; La Rooy et al., 2009) were able to ascertain key evidence-based facts about the effects of multiple interviews on children's memory. Firstly, they found that multiple interviews can be beneficial to children's memory, as they result in the recall of new, previously unmentioned information without inherently decreasing the accuracy of children's testimony. However, they also reported that the effect of suggestive questioning methods can be exacerbated by multiple interviews. Contradictory findings on the effects of multiple interviews on the accuracy of children's recall can be partially accounted for by the impact of

delay between the event and the investigation and between the interviews themselves. Both reviews (Goodman & Quas, 2008, La Rooy et al., 2009) suggested that further research is needed on individual differences between witnesses, the delay between interviews and the effect of different questioning methods before clear recommendations can be made about when and how multiple interviews should be used in forensic investigations involving child witnesses.

The present literature review built upon and extended previous reviews, supporting their main conclusions but also revealing that recent research has begun to address previously unexplored topics in relation to multiple interviews. One crucial change from the 2008 reviews was the emergence of new field studies investigating real-life forensic interviews (Katz & Hershkowitz, 2013; Leander, 2010; Waterhouse et al., 2016), which were rare 10 years ago. The present review supports the conclusion that multiple interviews can aid child witnesses in recalling previously unmentioned details (Dietze et al., 2013; Knutsson et al., 2014; Leander, 2010; Waterhouse et al., 2016) and that this increase in completeness is not inherently associated with a decrease in accuracy (Baugerud et al., 2014). The risks revealed in previous reviews were also supported by more recent research, which provided further evidence that the effects of suggestive questioning on children's recall can be exacerbated in multiple interviews (Melinder et al., 2010). However, the present review also revealed two further benefits of multiple interviews. Firstly, recent studies indicated that newly recalled information in multiple interviews can have high forensic relevance in field studies (Leander, 2010; Waterhouse et al., 2016). Secondly, in experimental studies, the effects of leading questions in multiple interviews were, under some circumstances, ameliorated by non-leading questioning in an additional interview (Melinder et al., 2010; Righarts et al., 2014). Whilst delay was still found to be an influential factor in decreasing the accuracy of children's testimony, the review also suggested that the effects of delay are not uniform across

laboratory and field studies (Baugerud et al., 2014; Katz & Hershkowitz, 2013; Peterson, 2011, 2015; Salmon et al., 2012).

Altogether, the literature review indicated that although the field has diversified, and recent studies began to address previously unexplored topics related to multiple interviewing, it was difficult to reach overarching conclusions due to the lack of forensically meaningful measures of the quality of children's testimony that could be used in both field research and experiments. To overcome this limitation, narrative coherence was chosen as an overarching measure of the quality of children's testimony, drawing on the legal (Pennington & Hastie, 1992; MacCormick, 2005; Rideout, 2008) and psychological literature (Kulkofsky et al., 2008; Reese et al., 2011) on memory and credibility. In addition to measuring the effect of multiple interviews on the completeness, consistency and connectedness of children's accounts, the thesis also investigated the impact of multiple interviews on the accuracy and credibility of children's testimony, aiming to explore the real-life impact of multiple interviews throughout the investigative process and in the courtroom.

The Impact of Multiple Interviews on the Narrative Coherence of Children's Testimony

In this thesis, the concept of narrative coherence was used to capture an overarching measure of the quality of children's testimony. Using a variation of the model of narrative coherence proposed by the story-telling model of legal decision-making (Pennington & Hastie, 1992, MacCormick, 2005; Rideout, 2008), narrative coherence was operationalised as a combination of the completeness, consistency and connectedness of children's testimony, each measured through a range of codes in Study 1 and Study 2.

Consistent with Erdelyi's (1996) theory of reminiscence as well as previous laboratory and field research (Dietze et al., 2013; Knutsson et al., 2014; Leander, 2010; Waterhouse et al., 2016), results from both Study 1 and Study 2 clearly demonstrated that

multiple interviews increase the completeness of children's testimony through allowing witnesses a second opportunity to recall new details. In Study 1, completeness codes included the number of details recalled, their relevance to the investigation, and their content. Not only did multiple interviews result in the recall of a significantly higher number of details than a single interview, newly recalled details included a large proportion of forensically sensitive information, as well as a significant number of new details relating to the victim and the alleged perpetrator, actions and objects related to the abuse and information about the location and timing of the alleged abuse. In fact, the second interview included a higher proportion of sensitive information than the first interview. Within the content categories, location and time may be particularly important, as these are necessary for the particularisation of the alleged crime (Powell & Thompson, 1997) and are rarely reported spontaneously by child witnesses (Westcott & Kynan, 2004). Additionally, two interviews included a significantly higher number of subjective details relating to children's emotions, cognitions and perceptions than a single interview. Previous research suggested that children's testimonies generally include only a small number of subjective details (Lyon et al., 2012, Westcott & Kynan, 2004), despite their impact on the credibility of child witnesses in court (Wessel et al., 2013). Altogether, the results of Study 1 suggest that multiple interviews can effectively increase the completeness of children's testimony, with particular emphasis on details considered to have high forensic relevance.

Similarly to methods used in previous laboratory research, completeness was measured as the proportion of correctly recalled details from a list of all possible details in Study 2. This method did not allow for the finely tuned analysis of each detail performed in Study 1 but had the advantage of discerning between correct and incorrect details. Results provided further support for the effectiveness of multiple interviews in increasing the completeness of children's testimony. However, the extent of reminiscence was far lower in

Study 2 than Study 1, suggesting that children may provide more new details in multiple interviews when describing events that are more complex or more emotionally meaningful. This difference also suggests that previous laboratory research may have underestimated the extent to which multiple interviews increase the completeness of children's testimony.

Based on previous research, the expectation was that multiple interviews would result in a decrease in consistency in children's accounts (Baugerud et al., 2014; Connolly et al., 2008). When inconsistencies are defined in the strictest sense, including any details that were mentioned in the first interview only, or the second interview only as well as contradictory details, a decrease in consistency is an inherent result of the increase in completeness that almost uniformly characterises multiple interviews. However, only contradictions are certain signs of inaccuracies (by definition, they can't be correct in both interviews), omitted and reminisced details can remain highly accurate under some circumstances (Baugerud et al., 2014; La Rooy et al., 2005). Therefore, Study 1 focused only on contradictory details. However, unlike previous studies, both contradictions within each interview and contradictions across interviews were considered. Results did not support the hypothesis that multiple interviews increase the proportion of contradictory details in children's accounts. As contradictions are often the only available proxy for the accuracy of children's accounts in the field, this result may have important implications for the use of multiple interviews in the field.

Finally, the third component of narrative coherence, connectedness, was coded as the proportion and number of linguistically connected details in children's accounts in Study 1 and Study 2. Due to the dearth of previous research on the impact of multiple interviews on the extent to which children provide causal-temporal links connecting the details mentioned (Habermas & Silveira, 2008), these analyses were fully exploratory. On the one hand, it was possible that multiple interviews would increase the connectedness of testimonies through

providing children with additional opportunities to establish causal-temporal links in their accounts. On the other hand, the increased number of details in multiple interviews compared with single interviews may make it more difficult for children to integrate details mentioned in different interviews on the same causal-temporal framework. Results from both Study 1 and Study 2 indicated that connectedness remains stable across multiple interviews, with the proportion of linguistically connected details from all details being almost equal in the first interview, the second interview, and information reported overall.

Whilst Study 1 and Study 2 have clearly demonstrated that multiple interviews can effectively increase the narrative coherence of children's testimony, the mechanisms underlying the benefits of multiple interviews need further exploration. La Rooy et al. (2009) suggested that the increase in completeness in multiple interviews is the result of reminiscence, as repeated recall occasions allow children to remember new details that they previously forgot (Ballard, 1913; Erdelyi, 1996). This explanation is consistent with Erdelyi's (1996) experiments showing that the extent of reminiscence across recall occasions is highest when the material to be recalled is complex, such as information presented in visual or narrative form. As forensic interviews require children to recall complex real-life experiences in fine detail, it is plausible that reminiscence contributes to the increase in completeness found across multiple interviews.

However, analyses of interviews with children who do not disclose abuse in forensic interviews despite corroborating evidence, suggest that in some cases, children may be unwilling, rather than unable to describe the allegations (Hershkowitz, Orbach, Lamb, Sternberg, & Horowitz, 2006). Maltreated children may find it difficult to develop rapport with the interviewer (Eltz, Shirk, & Sarlin, 1995), thus, some forensic interviewing guidelines suggest that the first interview should focus primarily on rapport building, with the allegations explored in later interviews once rapport has been established (Carnes et al.,

1999). The aim of rapport building is to increase the child's comfort and facilitate communication between the child and the interviewer (Collins, Doherty-Sneldon, & Doherty, 2013). Hershkowitz et al. (2006) found that premature transition to the allegation phase of the interview before rapport has been established may result in resistance and denial that can lead to non-disclosure. Thus, Hershkowitz et al. (2006) recommend that when interviewers encounter difficulties with communication during the pre-substantive phase, they may need to focus on further rapport building in the initial interview and return to the allegations on a later occasion. According to this view, the increased completeness of recall observed in multiple interviews is the result of successful rapport building across interviews (Carnes et al., 1999). The rapport building explanation of the benefits of multiple interviews is consistent with findings from Study 1 and Leander's (2010) field research showing that children report more sensitive and sexual details in the second interview than the first interview, presumably due to their increased comfort with the interviewer.

An alternative explanation of the increased proportion of sensitive and sexual details in further interviews (Leander, 2010) proposes that across multiple recall occasions, children become familiar with the requirements of forensic interviews and this enables them to focus increasingly on the details investigators want to hear about. Forensic interviews require a different manner of communication than conversations children engage in under typical conditions, thus, practice and scaffolding across multiple interviews may improve children's skill in reporting. For instance, Graffam Walker and Warren (1995) suggest that interviewers can scaffold children's narratives through providing context and chronology for open-ended questions. Consistent with this view, research on adult-child communication has demonstrated that children are able to learn conversational rules from more skilled individuals and improve the quality of their own narratives as a result (Langley, Coffman, & Ornstein, 2017; Reese, Haden, & Fivush, 1993).

Although whether the increase in completeness across interviews is the result of reminiscence, rapport building, or scaffolding is yet to be established, results from Study 1 and Study 2 revealed that when data from field studies and laboratory research is analysed using the forensically meaningful, multidimensional measure of narrative coherence, the benefits of multiple interviews outweigh the risks, as they increase completeness without decreasing the consistency and connectedness of children's testimony. However, as narrative coherence measures only the internal characteristics of a testimony, it is vital to clarify the relationship between components of narrative coherence and external outcome measures of forensic investigations, most importantly, the accuracy and credibility of testimonies in multiple interviews.

The Impact of Multiple Interviews on the Accuracy of Children's Testimony

Whilst internal measures of the quality of children's testimony are essential for field research, it is important to consider the aim of forensic investigations when evaluating methods of interviewing. Investigators aim to gain a complete picture of the allegations through collecting as many details as necessary to substantiate the witness' claims. However, additional details provided are only advantageous to the investigation if they constitute an accurate description of the events that took place. Therefore, the main aim of Study 2 was to establish the effect of multiple interviews on the accuracy of children's testimony.

A review of the literature on the effect of multiple interviews on the accuracy of children's testimony has revealed that multiple interviews can have a positive, negative or neutral effect on the accuracy of children's testimony, partially determined by the delay between the event and the interviews and the temporal spacing of the interviews (Baugerud et al., 2014; Brown et al., 2015; La Rooy et al., 2005). The results of Study 2 provide support for the view that multiple interviews do not inherently affect the accuracy of children's

testimony in either a positive or negative way, as the accuracy of children's recall 5-weeks after the event was the same regardless of whether they have also been interviewed 1-week after the event or not. However, the delay between the 1-week interview and the 5-week interview did lead to a small but significant decrease in the accuracy of children's testimony. A small proportion of the decrease in accuracy from children's recall in the 1-week interview to their overall recall in the two interviews can also be accounted for by the recall of different incorrect details in the first and second interviews.

Altogether, these results suggest that concerns about accuracy should not altogether prevent interviewers from conducting multiple interviews, but that the timing of both single and multiple interviews needs to be considered very carefully to avoid lengthy delays, which result in a decrease in the accuracy of children's testimony under laboratory conditions (La Rooy et al., 2015).

The Impact of Multiple Interviews on the Credibility of Children's Testimony

Alongside seeking the truth, forensic investigations also serve the aim of providing justice through convicting criminals if their crime has been substantiated and a jury finds them guilty. Thus, when evaluating methods of forensic interviewing, the impact of those methods on the credibility of children's testimony and thus, on the likelihood that a case will progress through the criminal justice system and the perpetrator will be found guilty also needs to be considered. Previous studies regarding the effect of multiple interviews on the credibility of children's testimony reported contrasting results, with Yozwiak et al. (2004) suggesting that multiple interviews may compromise credibility if children's recall is inconsistent across the interviews, while Waterhouse (2016) found that child witnesses interviews on two separate occasions were perceived more positively by mock jurors than children interviewed a single time, regardless of the consistency of their accounts.

Results from Study 3 did not support the view that conducting multiple interviews constitute a risk in the courtroom, as multiple interviews affected the credibility of a fictional testimony positively, even when it included inconsistencies in the form of contradictory details. Mock jurors rated the child as more competent and reliable when they read two interviews than when they were provided with a single interview, and they were also more likely to convict the defendant. Although the child was perceived as less credible in the inconsistent than in the consistent condition, inconsistent recall in two interviews was still rated as more credible than inconsistent recall in a single interview. Whilst previous research has emphasised the importance of consistency in determining children's credibility (Brewer et al., 1999; Yozwiak et al., 2004), these results suggest that completeness is at least as influential in determining credibility impressions and verdict decisions as consistency.

Altogether, findings indicate that multiple interviews that help children provide more complete descriptions of the allegation also increase the credibility of their testimonies in the eyes of a mock jury.

Individual Differences in the Impact of Multiple Interviews on Children's Testimony

Whilst multiple interviews were found to have an overall positive effect on the narrative coherence and credibility of children's testimony and were only associated with a slight decrease in accuracy due to delay, not all witnesses benefitted from multiple interviews in the same way.

In Study 1, witnesses of all ages reported more details overall than in a single interview, but the difference in completeness grew proportionally greater as the age of the witness increased. This suggests that older children may especially benefit from multiple interviews, perhaps due to remembering an overall larger amount of details, which may be difficult to describe fully in a single interview. In Study 2, children's age did not influence

the extent of reminiscence from the first to the second interview. However, a comparison the second interview with the single interview of the control group matched for delay revealed that younger children recalled more details 5 weeks after the event if they had not been questioned previously, whereas older children's reports were more complete if they had also participated in a previous interview 1 week after the event. Unlike Study 1, Study 2 also revealed significant gender differences between participants, as girls reported more details about an event that they participated in than boys, both in a single interview and in two interviews.

In conclusion, the main contribution of the thesis was to reveal that multiple interviews provide an effective method of increasing the completeness of children's testimony, and that an increase in completeness is associated with both increased accuracy and increased credibility. Results also provided support for the view that the decrease in accuracy observed in successive interviews is primarily the result of delay between interviews rather than multiple recall occasions. The results of Study 3 provide particularly valuable information for researchers and practitioners studying the impact of multiple interviews, as no previous study has directly compared the credibility of testimonies elicited in a single interview with the credibility of recall provided over two interviews. Another novel contribution of the thesis was the use of the concept of narrative coherence to provide an overarching measure of the quality of children's testimony. The following section will assess the suitability of the present model of narrative coherence for research on multiple forensic interviews with child witnesses based on the results of Study 1, Study 2 and Study 3.

5.2 Assessing the Present Model of Narrative Coherence

For the purpose of this thesis, narrative coherence was defined as a combination of the completeness, consistency and connectedness of children's accounts. This definition was based on previous research in law and psychology on the story-telling model of legal

decision-making (Pennington & Hastie, 1992). According to the story-telling model of jury decisions, the perceived credibility of a story is determined partly by its coherence. In Pennington & Hastie's original model (1992), narrative coherence is described as consisting of three parts: completeness, consistency and plausibility. However, as the aim of the thesis was to explore measures of the quality of the testimony itself, without regards to other evidence, plausibility was not included as a component of narrative coherence. Instead, connectedness, or the extent to which details in a story are connected through causal-temporal links was added as the third component to concept of narrative coherence (MacCormick, 2005; Rideout, 2008).

Previous research on the relationship between narrative coherence and children's memory relied on different approaches to operationalise the concept, including narrative cohesion (Chae et al., 2016; Kulkofsky et al., 2008), story schema (Brown et al., 2018; Feltis et al., 2010; Stein & Nezworski, 1978; Westcott & Kynan, 2004) and multidimensional models (Haberman & Silveira, 2008; Peterson et al., 2013; Reese et al., 2011). In the present studies, the three components of narrative coherence were translated into frequently used measures of children's testimony and measures developed specifically for this research. Completeness was measured through the number of details reported, the relevance of the details for the investigation, the content of the details and their subjective content. Consistency was operationalised as the number of contradictions in Study 1 and the number of omissions, additions and contradictions in Study 2. For connectedness, a variation of the Narrative Cohesion Coding Scheme (Kulkofsky et al., 2008) was used to identify local markers of causal-temporal connectedness. Although narrative coherence was defined based on the story-telling model (Pennington & Hastie, 1992), components can also be mapped onto multidimensional models of coherence (Reese et al., 2011).

Results revealed that narrative coherence is a useful concept for evaluating the quality of children's testimony, although components did not play equal roles in contributing to the relationship between narrative coherence and the accuracy and credibility of children's accounts. Whilst completeness was correlated with both consistency and connectedness, there was no evidence of a relationship between connectedness and consistency. Consistency and completeness, but not connectedness, were also linked in terms of their positive impact on the accuracy and credibility of children's testimony. Altogether, findings do not suggest an overall relationship between consistency, completeness and connectedness. Whilst consistency and completeness were found to clearly contribute to the quality of children's testimony, more research is needed to determine the role of connectedness in forensic interviews.

Narrative Coherence and Memory

Previous findings on the relationship between narrative coherence and memory have been inconsistent, with some research suggesting that children who provide coherent narratives are also more accurate in their recall (Chae et al., 2016), whilst others reported a completeness-accuracy trade-off (Kulkofsky et al., 2008). The results of Study 2 support the theory of a positive relationship between narrative coherence and accuracy. In contrast to Kulkofsky's results, the strongest positive correlation was found between completeness and accuracy, demonstrating that when children are able to describe the full extent of the events they are asked about, they also describe details more accurately than when they provide less information about them.

Although no overall relationship was found between the consistency of children's accounts and their accuracy, consistency was a strong predictor of accuracy at the level of individual details. On average, 97% of details mentioned in both interviews were accurate,

whilst details mentioned in the first interview or the second interview only were 90% and 82% correct, respectively. Previous studies have found that accounts rich in causal-temporal connections were more accurate, as connections allowed children to recall details in a more structured way (Chae et al., 2016; Kulkofsky et al., 2008). In contrast to these findings, no correlation was found between connectedness and accuracy in Study 2. Altogether, results showed a positive correlation between narrative coherence and accuracy, mainly driven by a strong relationship between the number of details reported and the proportion of correct details.

Narrative Coherence and Credibility

Whilst previous research in cognitive psychology focused primarily on the relationship between narrative coherence and memory accuracy, from a legal perspective, the emphasis was on the influence of narrative coherence on the credibility of stories (Bennett & Feldman, 1981; Pennington & Hastie, 1992; Willmott et al., 2018). The story-telling model of legal decision-making suggests that coherent narratives are perceived as more credible than narratives which do not fulfil the requirement of internal coherence (Pennington & Hastie, 1992). The results of Study 3 provide support for the expected relationship between narrative coherence and credibility. When presented with more complete testimonies, mock jurors rated the reliability and competence of the child witness higher than when the testimonies contained less details, and they were also significantly more likely to convict the alleged perpetrator.

Although the consistency of the testimony had no significant effect on the proportion of convict decisions, mock jurors rated the witness as more credible and competent in the consistent condition than when the testimony included contradictory details. Additionally, in the self-report part of the questionnaire, mock jurors rated the completeness and consistency

of children's testimony as the two most influential aspects of the testimony affecting their decision to convict or acquit the defendant.

Unfortunately, connectedness was not manipulated experimentally in Study 3, therefore, it is not known whether this component of narrative coherence affected mock jurors' impressions and decision-making. The importance of causal-temporal connectedness was rated higher than that of previously established predictors of credibility, such as children's age and verbal competence, suggesting that connectedness may play a role in determining the credibility of narratives.

Altogether, results supported the hypothesis that narrative coherence influences credibility, with the completeness of the testimony emerging as the most important factor, above its consistency.

The Impact of Individual Differences

Research on multidimensional models of narrative coherence has suggested that the components of coherence develop at different rates across the lifespan (Reese et al., 2011). In line with these findings, children's age was associated with some, but not all components of the narrative coherence of their recall in Study 1. Older children provided more complete and more connected accounts than younger children, but there was no relationship between children's age and the consistency of their testimony. Age, however, did not significantly affect any component of the narrative coherence of 8- to 11-year old children's recall in Study 2, suggesting that in middle childhood, other individual differences may outweigh the influence of age on narrative coherence. In contrast to previous findings (Goodman et al., 1987; Tabak & Klettke, 2014), the perceived age of the child witness also did not affect the credibility of the fictional testimony in Study 3, regardless of the consistency and completeness of the story.

In addition to age, children's gender also affected the narrative coherence of their testimony in Study 2, but not Study 1. Girls recalled significantly more details than boys, both in a single interview and across two interviews. However, the gender difference found in the completeness of children's recall could be a result of an interaction between the gender of the interviewer and the gender of the child (Lamb & Garretson, 2003), as all participants were interviewed by a female interviewer. Gender did not affect narrative coherence in Study 1, suggesting that the difference between girls and boys is less prominent when more prominent differences in age, background and case type are present.

Assisting Children in Providing Coherent Narratives

In addition to exploring the correlations between the three components of narrative coherence, and their relationship with the accuracy and credibility of children's testimony, the present studies have also revealed methods which may be used by interviewers to help children provide coherent narratives. Most importantly, the present studies revealed that multiple interviews are an effective method to increase the narrative coherence of children's testimony. In both Study 1 and Study 2, the completeness of children's accounts increased significantly when they were interviewed twice, rather than a single time. Remarkably, in the forensic interviews analysed in Study 1, almost twice as many new details were reported over the course of the two interviews than in a single interview only. In contrast to the hypotheses, multiple interviews did not decrease the consistency of children's testimony in Study 1, when both contradictions within each interview and across the two interviews were considered. The proportion of causal-temporal connections in children's account also remained stable across the two interviews, in both Study 1 and Study 2. Therefore, multiple interviews had an overall positive effect on the narrative coherence of children's accounts through increasing their completeness without affecting their consistency and connectedness.

In addition to multiple interviews, an additional method of increasing the connectedness of children's testimony was explored in Study 1. Based on research on the influence of the types of questions used by interviewers (Feltis et al., 2010; Lamb et al., 2003), it was expected that the connectedness of interviewers' own questions would affect the connectedness of children's responses. Consistent with this expectation, results demonstrated that when interviewers asked causally or temporally connected questions, children were more likely to provide linguistically connected responses than in response to unconnected questions. Interestingly, a second interviewing strategy was also associated with a higher proportion of linguistically connected responses. When interviewers followed the child's lead by requesting elaborations on details mentioned by the child, the proportion of causal-temporal links was significantly higher than for both connected and unconnected questions. It appears that interviewers can help children provide connected responses either through providing a causal-temporal framework for the child to use, or through allowing the child to connect their own responses. Considering the positive relationship between narrative coherence and memory accuracy and the impact of narrative coherence on the credibility of children's accounts, methods that increase the coherence of children's testimony may provide valuable tools for interviewers.

In conclusion, results from Study 1, Study 2 and Study 3 suggest that a model of narrative coherence consisting of the two original components included in the story-telling model (Pennington & Hastie, 1991) provides a useful measure of the quality of children's testimony in field research. Both completeness and consistency were found to have a significant impact on the two key outcome measures of forensic investigations: accuracy and credibility. Whilst the present research provided no evidence that connectedness is associated with the other two measures, or that it influences the accuracy of children's recall, the relationship between connectedness and credibility has not yet been explored and future

research may reveal associations between completeness, consistency and connectedness in courtroom decisions.

5.3 Strengths and Limitations of the Methodology

This thesis included three studies with different methodologies: a field study, an experiment and a mock jury survey. The rationale behind choosing these methods was that results from the three studies would be complementary, and the limitations of one method would partially be overcome by another. This approach was deemed advantageous, because the literature review revealed that the lack of compatibility between measures used in field research and laboratory studies prevents researchers and practitioners from drawing strong conclusions about the effect of multiple interviews on children's memory.

This complementary methodological approach was successful in overcoming some inherent limitations in field research and experiments. Analysing real-life forensic interviews resulted in high external validity for Study 1, however, this method did not allow for analyses of accuracy due to the researcher's ignorance to the 'ground truth' of the investigation. This shortcoming was addressed in Study 2, where children were interviewed about a staged event they participated in, thus, researchers were able to ascertain the accuracy of their recall. Although findings from laboratory studies should be interpreted with caution due to crucial differences in the types of events children are asked to recall in experiments and real-life forensic interviews, taken together with Study 1, results provide convincing evidence for the beneficial impact of forensic interviews on children's memory. An additional advantage of the diverse methodological approach was that whilst Study 1 focused on establishing the impact of multiple interviews on narrative coherence, Study 2 and Study 3 provided crucial information on the influence of changes in narrative coherence on external measures of the quality multiple interviews, which are not usually available in field research. Furthermore,

the thesis was able to address both crucial outcome measures of the real-life impact of forensic interviewing methods: credibility and accuracy.

Despite the advantages of the complementary methodologies, methods in all three studies were limited by restrictions on time, financial possibilities and access to sensitive forensic materials. The literature review highlighted some pervasive limitations of previous research on multiple interviews; thus, certain research elements were prioritised over others. Specifically, the inclusion of a control group in Study 2 and the use of a between-subjects design involving 8 different conditions in Study 3 both led to a decrease in power due to sample size restrictions. In addition to limitations on sample size, the researcher's access to forensic materials was also restricted, leading to the use of a convenience sample, rather than a systematic sampling method in Study 1. Despite these shortcomings, the studies addressed previously unexplored questions regarding the impact of multiple forensic interviews on children's testimony and revealed results which may motivate future research.

5.4 Future Research

Although recent research has begun to address the role of narrative coherence in children's memory and communication (e.g. Brown et al., 2018; Chae et al., 2016; Feltis et al., 2010; 2011; Kulkofsky et al., 2008; Reese et al., 2011; Westcott & Kynan, 2004), there is currently no consensus on how the concept of narrative coherence is best operationalised in research investigating the quality of forensic interviews with child witnesses. While the story-telling model of legal decision-making (Pennington & Hastie, 1992) suggests that the coherence of testimonies is determined by their completeness, consistency and connectedness (MacCormick, 2005; Rideout, 2008), within the field of developmental psychology, story grammar approaches (Brown et al., 2018; Feltis et al., 2010, 2011; Westcott & Kynan, 2004), linguistic markers of coherence (Chae et al., 2016; Kulkofsky et al., 2008) and the multidimensional model of narrative coherence (Brown et al., 2018, Reese et al., 2011) have

been more widely studied than strict interpretations of the story-telling model. Consistent with Brown et al.'s (2018) research, in Study 2, the number of local markers of temporal and causal connectedness in children's accounts correlated with the extent of global connections established in the narrative, suggesting that there is an overlap between different models of narrative coherence.

However, the application of different models of narrative coherence to children's autobiographical reports has also led to contrasting results in some respects. For instance, using linguistic and story grammar measures of narrative coherence, Brown et al. (2018) and Kulkofsky et al. (2008) found that components of narrative coherence correlate with one another, while Reese et al. (2008) suggested that each element of the multidimensional model is independent from other elements. In Study 2, significant correlations were found between the completeness and consistency and between completeness and connectedness, but not between consistency and connectedness. To further improve understanding of the role of narrative coherence in children's recall, future research should establish the relationship between different models of narrative coherence and assess the strengths and limitations of each approach in the context of evaluating the quality of forensic interviews with child witnesses.

One specific question highlighted for future research on the story-telling approach to coherence (Pennington & Hastie, 1992; MacCormack, 2005; Rideout, 2008) is the role of connectedness in the relationship between narrative coherence and the credibility of children's testimony. Whilst the results of Study 3 indicate that consistency and completeness are key factors in determining adults' impressions of the credibility of a fictional child witness alleging sexual abuse, connectedness was not experimentally manipulated in this study. As both previous research (MacCormick, 2005; Rideout, 2008) and participants' responses to a self-report questionnaire in Study 3 suggest that the perceived credibility of

witnesses is influenced by causal and temporal connections in the testimony, further research on this topic is necessary to fully understand the factors affecting credibility decisions in the courtroom and at earlier points in the investigation.

In addition to exploring the relationship between models of narrative coherence used in forensic and developmental psychology and between the connectedness and credibility of children's testimonies, future research also needs to investigate the impact of individual differences between witnesses and cases, on the narrative coherence, accuracy and credibility of children's recall. Whilst in Study 1, children's age was the only factor affecting the relationship between multiple interviews and narrative coherence, the results of Study 2 suggested that gender differences can also be influential under certain conditions. Findings from Study 2 also clearly indicated that differences between investigations in terms of the temporal spacing of the interviews influence the accuracy of children's descriptions elicited in multiple interviews. Although individual differences did not affect the credibility of the child witness in Study 3, previous research reported the contrary, suggesting that case type and children's age can also influence the perceived credibility of the witness statement (Goodman et al., 1987; Tabak & Klettke, 2014). Thus, further research systematically analysing the effects of the age and gender of the witness and the delays involved in investigations on the narrative coherence, accuracy and credibility of testimonies elicited in multiple interviews is essential for establishing whether successive interviews should be conducted in any specific case.

Finally, although children in the sample of Study 1 were interviewed up to 11 times, only data from the first two interviews was analysed. Likewise, in Study 2, the narrative coherence and accuracy of children interviewed twice was compared with the recall of a control group interviewed on a single occasion and in Study 3, the perceived credibility of a fictional case involving either one or two interviews was assessed. Therefore, results from the

present studies allow researchers to draw conclusions about the benefits of two interviews, without any clear implications on the effectiveness of investigations involving three or more interviews. Whilst some studies suggest that the completeness of children's accounts is significantly increased in each interview even in 6-session protocols (Carnes et al., 2001; Faller & Nelson-Gardell, 2010), a field study found that no new information was reported in third interviews (Waterhouse et al., 2016). The impact of multiple interviews on the accuracy of children's testimony may also cause more concern in longer interview protocols, as an increase in the number of interviews often means an increase in the length of delay between the first interview and the last interview, which may compromise children's ability to recall details accurately based on the results of Study 2. Although field research found that children's scores on structured credibility questionnaires were higher after 6 interviews than 4 sessions (Faller & Nelson-Gardell, 2010), no studies have yet analysed the credibility of testimonies elicited in more than 2 interviews under controlled conditions. Therefore, it is vital that future field and laboratory research addresses the impact of 3 or more interviews on children's memory and credibility.

Although these limitations highlight that crucial questions remain unexplored, the body of evidence that accumulated following decades of research on the benefits and risks of multiple interviews may provide sufficient evidence to make tentative suggestions about when and how multiple interviews should be conducted.

5.5 Evaluating Guidelines for Multiple Interviews

Despite the prevalence of multiple forensic interviews with children alleging physical and sexual abuse, there are currently no protocols designed specifically for police officers conducting investigations involving multiple interviews with children in the United Kingdom. *Achieving Best Evidence* (ABE) guidelines (Ministry of Justice, 2011) state that the video-recording of interviews should minimise the need for witnesses to repeat their

accounts, but acknowledge that further interviews may be necessary when the full extent of the allegations could not be covered in a single interview and when new details about the case emerge after the interview from the witness, the accused or a third party. Additionally, ABE guidelines suggest that interviewers may consider conducting pre-planned multiple interview sessions in investigations involving particularly vulnerable witnesses but include no instructions on how such interviews should be conducted (Ministry of Justice, 2011). Even in countries where specific guidelines exist (e.g. in the United States, Canada and Nordic countries), there has been little empirical research evaluating these protocols (Connell, 2009). This section will describe and evaluate existing guidelines for multiple interviews with children alleging sexual and physical abuse to make four key recommendations for conducting multiple interviews based on the results of the present studies and a review of the literature on the risks and benefits of multiple interviews.

All existing multiple interview protocols originate in the Extended Forensic Evaluation (now called Extended Forensic Interview), developed in Child Advocacy Centres in the United States for interviewing children who are not able to disclose the full extent of their allegations in a single forensic interview (Carnes et al., 1999; 2001). The Extended Forensic Interview (EFI) ‘spreads the phases of a forensic interview with the child over time’ (Carnes et al., 2001, p. 2.), thus, rather than repeating questions in successive interviews, investigators aim to progress through the allegation or allegations over several interview sessions. The term ‘extended interview’ is used to differentiate these sessions from ‘repeated interviews’, where every phase of the forensic interview is present in each session, and ‘subsequent interviews’, which also refer to full interviews, usually conducted to explore additional allegations (Duron & Remko, 2018).

EFI-based guidelines all emphasise collaboration within multi-disciplinary teams (MDTs) and the integration of a range of non-investigative elements (such as behavioural

assessments and body safety information) in the forensic interview schedule. However, more recent guidelines, such as the Sexual Abuse Forensic Evaluation (SAFE) Program developed in Toronto (Azzopardi et al., 2014), the Multiple Session Forensic Interviews (MSFI) used in Children's Advocacy Centres in Texas (Duron & Remko, 2018) and the Sequential Interview of the Nordic Barnahus model (Langballe & Davik, 2017) introduce variations from the initial EFI protocol in the recommended number, structure and timing of the forensic interviews. This review will compare and evaluate the original EFI, SAFE, MSFI and Sequential Interview models based on the types of cases for which they are recommended, the length and structure of the interview process, the methods used to determine the outcome of the case, and the use of props to enhance recall in multiple interviews.

What Types of Cases are Referred for Multiple Interviews?

Generally, multiple interview guidelines suggest that successive interviews should be conducted when the child is not able disclose or describe the full extent of the allegation in the first interview, despite a strong suspicion of abuse (Azzopardi et al., 2014; Faller et al., 2010). However, cases involving specific types of victims may be recommended for multiple interview sessions even before a single interview has been conducted, in the planning stage of the investigation. In the EFI model, for instance, it is suggested that multiple interviews should be considered when interviewing victims who are very young (Hewitt, 1999), have disabilities (Davies & Faller, 2007) or belong to a cultural minority (Fontes, 2000), as well as in cases where the allegations are complicated (Faller et al., 2010). The MSFI further suggests that children who experienced trafficking or extreme trauma may also benefit from multiple interviews (Duron & Remko, 2018). The MSFI also makes it clear that there are cases where the use of multiple interviews should not be encouraged, such as when the interviewer has already asked specific abuse-related questions from the child in the first interview, and in cases where the child's environment might make multiple interviews

impractical due to safety factors, lack of caregiver support or the possibility that the suspect constitutes a flight risk (Duron & Remko, 2018).

In the EFI, SAFE and MSFI models, the decision to conduct multiple forensic interviews is generally made during or after the first interview, when the interviewer recognises that a single forensic interview will not be sufficient to resolve concerns and conclude the investigation (Azzopardi et al., 2014; Duron & Remko, 2018; Faller et al., 2010). In contrast, Barnahus Sequential Interview guidelines emphasise preparation, therefore, the decision to offer multiple interviews is often made in advance to starting the interview process (Langballe & Davik, 2017). Specifically, when children between the ages of 3 and 6 are interviewed in Barnahus centres, multiple interviews are the default option for investigators. Although the EFI and SAFE guidelines also recommend using multiple interviews primarily with younger children (3- to 12-year-olds), Barnahus guidelines differ from these protocols due to their sole emphasis on age rather than inconclusive investigations or case types. Multiple interviews allow interviewers to provide children with breaks between interviews which is seen as particularly beneficial to preschool children due to their shorter attention spans (Langballe & Davik, 2007).

Evaluation of the types of cases referred for multiple interviews

The use of multiple interviews to gain more complete information about the alleged abuse in cases where the child is reluctant or unable to disclose is strongly supported by research. In Study 1, the proportion of forensically relevant information was higher in the second interview than the first one, and a large number of key details were first mentioned in the second interview, including details related to the location and timing of abuse and the cognitions, perceptions and emotions of the victim during the abuse. Consistent with these findings, field studies have shown that second and subsequent interviews frequently yield

new disclosures of forensically relevant information (Azzopardi et al., 2014; Carnes et al., 2001; Faller et al., 2010; Leander, 2010; Waterhouse et al., 2016).

Interestingly, in a laboratory context (Study 2), the proportion of new information in the second interview was far lower than in real forensic interviews (Study 1), suggesting that the more complex and emotionally impactful the events children are asked to describe, the more beneficial subsequent interviews are in terms of increasing the completeness of their accounts. Based on the results of Study 1, which involved a sample mainly consisting of witnesses alleging multiple instances of abuse, multiple interviews might be particularly advantageous in increasing the completeness of children's accounts when children need to describe more than one instance of abuse. In general, forensic interviews with victims alleging multiple instances of abuse are challenging, as these victims may find it difficult to provide descriptions of specific instances of abuse rather than script-based reports (Brubacher et al., 2014; Connolly et al., 2008). Consistent with previous research (Hershkowitz & Terner, 2007; Patterson & Pipe, 2009), second interviews in Study 1 yielded a large proportion of new information about different instances, or different aspects of the alleged abuse than the first interview, suggesting that multiple interviews help interviewers explore the full extent of the allegations when cases are more complex. Using multiple interviews in this manner, however, involves 'subsequent interviews' rather than extending the phases of a single interview over several recall occasions, as suggested by EFI guidelines.

Although guidelines suggest that multiple interviews should preferably or only be used with the youngest age groups, there is little empirical evidence supporting the idea that multiple interviews are more suitable for pre-schoolers than older children. In Study 1, older children appeared to benefit more from multiple interviews than younger children, as they reported a higher proportion of new details in the second interview. However, as younger children reported less details overall, a small increase in the completeness of their accounts

may have been more forensically relevant than a larger increase for older children. Altogether, research on age differences in the effect of multiple interviews on the accuracy and narrative coherence of children's recall has been inconclusive (e.g. Baker-Ward et al., 1993; Flin et al., 1992; Peterson et al., 2005; Peterson & Bell, 1996; Pipe et al., 1999; Ornstein et al., 1992; Salmon & Pipe, 1997), and individual differences between children outweighed age differences in many studies, suggesting that a flexible approach to eligibility for multiple interviews may be more advantageous than strict age-based criteria.

In summary, it may be suggested that multiple interviews are particularly useful for victims who are unable to report the full extent of the alleged abuse in the first interview for a multitude of reasons, including the complexity of the case or individual differences.

How are Investigations Involving Multiple Interviews Structured and Scheduled?

The extended forensic interview was initially conceived as a 12-session interview protocol, involving 11 sessions with the child and one with the non-offending caregiver. The interview process was then reduced to 7 sessions with the child, due to the lack of new information reported after the 8th interview in the pilot study (Carnes et al., 1999). Further research comparing 4-session interview protocols with 8-session versions has shown that 3 sessions with the child are often insufficient to explore the full extent of the allegations. However, the 6th and 7th session rarely yielded new disclosures (defined as new details related to the alleged abuse), indicating that a 6-session format is the best way to balance the need to maximise information gained with the aim to minimize the number of necessary interviews (Carnes et al., 2001). The SAFE model was designed to involve slightly fewer interviews than the EFI, with investigators aiming to interview children between 2 and 4 times (Azzopardi et al., 2014). In the evaluation of the SAFE model, 86% of children disclosed some forensically relevant information by the third interview, however, new forensically

relevant details were provided by almost all children in each successive interview. The Barnahus Sequential Interview model emphasises flexibility with regards to the number of sessions necessary to gain a complete picture of the allegation, as the multi-disciplinary team decides on the number of sessions and their pacing as the interviews progress.

Interview guidelines are less prescriptive about the delay between the interviews than about the number of interviews conducted. All guidelines suggest that successive interviews should be conducted within a week of the previous interview (Carnes et al., 2001; Azzopardi et al., 2014), with the MSFI and the Sequential Interview model prioritising same-day interviews, whenever possible (Duron & Remko, 2018; Langballe & Davik, 2017). Importantly, the SAFE model also clarifies that no further interviews should be conducted after the interviewer has transitioned to asking specific abuse-related questions from the child (Azzopardi et al., 2014).

The recommended structure of multiple interviews in each protocol is based on variations of the original EFI structure. Using the ideal 6-session investigation as an example (Carnes et al., 2001), the EFI would generally begin with an interview with the non-offending caregiver, during which the child is not present. This would be followed by rapport building, developmental assessment and a discussion of ground rules in the second interview, and the completion of behavioural checklists in order to assess the social and cognitive functioning of the child in the third interview. Abuse-focused questioning would only begin in the fourth interview, during which the interviewer would start probing with life context and abuse context questions, then move on to the allegation, using the most open-ended interviewing methods. In the fifth interview, the interviewer would continue asking open-ended questions, followed up with more specific questioning, using memory retrieval cues and props, when necessary. Finally, the interviewer would ask any remaining questions in the 6th interview before providing safety information to the child and closing the interview process.

More recent guidelines tend to be more flexible in their approach to the structure of the interview process, and usually involve less components (Azzopardi et al., 2014; Duron & Remko, 2018; Langballe & Davik, 2017). The SAFE protocol, for example, concludes in fewer sessions because it does not involve social-behavioural assessment or body safety teaching (Azzopardi et al., 2014). The MSFI is also less prescriptive in terms of the stage each interview should address, simply stating that the second interview should start where the previous interview has ended (Duron & Remko, 2018).

As the Barnahus Sequential Interview model can include an unspecified number of interviews, spread over the same day or several days, the structure of the interview process also differs from case to case. Taking a 3-session process with all interviews taking place on the same day as an example, Langballe & Davik (2017) suggest that the first session should include rapport building and an exploration of the child's cognitive abilities, as well a discussion of the ground rules for the interview. Interviewers are explicitly recommended to avoid questions that might prompt an allegation in the first interview. After a break, the second interview would involve open-ended and focused questions, whilst the third and final interview would be a short session, providing the interviewer with an opportunity to ask questions suggested by the multi-disciplinary team during the break between the second and third interviews.

Evaluation of the structure of multiple interviews

Although a large body of research has focused on first and second interviews, studies involving more than 2 interview sessions are scarce. Consistent with research evaluating the number of interviews in EFI protocols, case studies have shown that details of high forensic relevance do emerge after the second interview (Leander, 2010), although not in every case

(Waterhouse et al., 2016). However, previous research indicates that the delay between interviews is at least as important a predictor of their utility as their number.

In laboratory research, the accuracy of newly recalled information decreased from 72% to 56% when the interviews were separated by a 6-month delay rather than conducted less than a day apart (La Rooy et al., 2007). In contrast, field studies show that newly recalled information can remain highly accurate when interviews are months apart (Baugerud et al., 2014). In Study 2, a 1-month delay did compromise the accuracy of newly recalled details, although the second interview still led to a significant increase in the completeness of children's accounts, and their overall accuracy decreased only to a very small extent. Although all EFI guidelines aim to conduct interviews much less than a month apart, in 6- or 8-session interview protocols, the delay between the first and last interview can be substantial, potentially leading to a decrease in the consistency and accuracy of information reported. In Brown et al.'s (2015) research involving children with intellectual delay, interviews between the first and last sessions provided children with an 'inoculation' against the effects of delay, however, the existence of the inoculation effect has not generally been supported by Peterson's (2015) study involving typically developing children.

The structure of investigations involving multiple forensic interviews has not been sufficiently explored by previous research. The explicit or implicit aim of guidelines to avoid abuse-related questioning in the first interview appears consistent with findings from Study 1 showing that children alleging physical and sexual abuse provide far less sensitive information in the first interview than in the second one. However, the increase in the proportion of sensitive information in the second interview may be due to a better understanding of what the interviewer is interested in, rather than increased rapport, which could mean that the benefits of repeated recall occasions may decrease if sensitive information is only introduced in the second interview. Furthermore, most guidelines

advocate using the first interview as an extended rapport building session rather than or in addition to trying to establish rapport at the beginning of each session, but studies suggest that lengthy rapport-building phases can be counter-productive in some circumstances (Davies, Westcott, & Horan, 2000; Teoh & Lamb, 2010). Davies et al. (2000) found that when rapport building lasted more than 8 minutes in forensic interviews, children's answers in the substantive part of the interview contained less CBCA criteria than following a shorter rapport phase.

How is the Outcome of the Investigation Determined?

Achieving Best Evidence guidelines (Ministry of Justice, 2011) state that the information obtained should be objectively evaluated after each interview by the interview team, in light of the case as a whole. The guidelines caution against using behavioural signs as indicates of reliability, but do not provide specific instructions on how the credibility of the witness should be assessed. In contrast, the EFI protocol provides structured guidance for credibility assessments (Carnes et al., 1999). The 'desk guide' used in EFI investigations is described as a tool to assist interviewers in analysing the outcome of the interview process rather than a test of the veracity of the witness statement. In the EFI model, the credibility assessment is the final step of the interview process, taking place after the last interview with the child has concluded. The assessment guide consists of 8 main categories, each of which lists several factors which affect the credibility of the statement. The guide should be used with caution, as the presence or absence of any factors does not confirm whether the testimony is true or false. The eight categories include disclosure factors, specific details obtained, developmental factors, emotional content, the results of the behavioural checklist, corroborative information, motivational factors and alternative explanations (Carnes et al., 1999; Faller & Nelson-Gardell, 2010). Based on the results of the credibility checklist, cases are categorised into credible disclosures, credible nondisclosures, noncredible disclosures and

unclear cases, where the allegations have not been resolved (Faller & Nelson-Gardell, 2010). During the interviews, interviewers are encouraged to gently challenge the credibility of children's statements, to allow witnesses to clarify or elaborate on statements which may compromise their credibility. The outcome of the investigation is decided by the multidisciplinary team, taking all evidence into consideration, including the interviewer's credibility impressions. The SAFE protocol relies on the same credibility assessment guidelines as the EFI, but the result of the credibility assessment is determined by a consensus between the forensic interviewer and a peer reviewer (Azzopardi et al., 2014).

Evaluation of the process of determining the outcome of the investigation

The credibility assessment checklist used in EFI investigations has not been evaluated by empirical research, and the results are not admissible in court (Carnes et al., 1999). The criteria described in the guideline are similar to factors listed in other structured protocols, in particular, Criterion-based Content Analysis (CBCA). For instance, both the EFI guidelines and the CBCA criteria emphasise the presence or absence of a range of specific details in children's accounts, as well as motivational factors which may point to alternative explanations (Steller & Kohnken, 1989). Whilst research has shown a clear relationship between some CBCA factors and the veracity of children's statements, a meta-analysis of laboratory studies evaluating the effectiveness of CBCA guidelines in distinguishing between liars and truth-tellers has found a 30% error rate in decision-making (Vrij et al., 2005). Furthermore, coaching has been found to affect children's scores on structured credibility assessment checklists (Vrij et al., 2002). Altogether, research suggests that current structured credibility assessment guidelines do not pass court admission criteria (Vrij et al., 2005), therefore, they should be used with caution during investigations.

Whilst the type of structured credibility assessment recommended by EFI guidelines has not been universally supported by empirical research, it could be argued that using a set list of criteria is still preferable to unstructured approaches, which may be influenced by interviewer bias to a greater extent. In particular, the EFI checklist includes several criteria in relation to the child's age and emotional displays during the forensic interviews, which may be helpful in decreasing interviewer bias (Faller & Nelson-Gardell, 2010). Although in Study 3, the perceived age of the witness did not significantly affect the credibility impressions of adults, previous research suggests that under some circumstances, older children alleging sexual abuse may be perceived as less credible than younger children (Goodman et al., 1987; Tabak & Klettke, 2014). Encouraging interviewers to explicitly consider factors related to the child's developmental level may help counter unconscious age-related biases regarding the honesty and credibility of witnesses. Similarly, the EFI guidelines include 10 emotional content categories for interviewers to consider when assessing whether the child's affective display was congruent with the allegations (Faller & Nelson-Gardell, 2010). As previous research has shown that adults find the accounts of children who cry or display other signs of sadness when describing alleged sexual or physical abuse significantly more credible than the testimonies of witnesses with a neutral, positive or angry emotional display (Wessel et al., 2013), instructing interviewers to pay attention to a wider range of emotions may be beneficial for avoiding common biases.

However, the EFI credibility assessment guidelines also include criteria which seem to endorse widely held beliefs about factors affecting children's credibility which are not supported by empirical research, such as the consistency of children's statements. In line with previous research (Berman et al., 1995; Brewer et al., 1999; Myers et al., 1999; Stromwall & Granhag, 2005), in Study 3, adults indicated that consistency is one of the key factors they used when assessing the credibility of children's statements, rating the credibility of

consistent witnesses significantly higher than inconsistent witnesses. As Study 2 and previous research (Baugerud et al., 2014; Gilbert & Fisher, 2008) suggest a lack of correlation between the consistency and overall accuracy of children's statements, emphasising the consistency of statements in credibility assessment guidelines might strengthen, rather than avoid the impact of bias.

In all EFI-based models, including the SAFE, MSFI and Barnahus Sequential Interview protocols, case outcomes are determined by a multidisciplinary team. In Study 3, female evaluators rated the fictional account of a child witness alleging sexual abuse significantly more credible, competent and reliable than male evaluators, suggesting that the beliefs of individual adults tasked with assessing children's credibility may influence case outcomes. Considering the impact of the evaluator's characteristics on their impressions of the credibility of children's accounts, involving multiple team members in outcome decisions appears preferable. Previous research has indicated that group deliberations decrease the gender difference in perceived credibility (Crowley et al.,1994; Gabora et al., 1993), suggesting that team discussions may help avoid bias resulting from individual differences between evaluators.

Assessing the credibility of the witness statement is the final step of the EFI protocol, thus, interviewers' credibility impressions are only recorded once all interviews with the child witness have been conducted. As the results of Study 3 indicate that multiple interviews increase the perceived credibility of child witnesses through introducing additional details which make the child's testimony more complete, it might be beneficial to record credibility impressions after each forensic interview, to draw conscious awareness to the impact of completeness on credibility. Saykaly et al.'s (2013) research suggesting that multiple interviews result in more similar scores for truth-tellers and liars on structured credibility assessments further supports the importance of increased awareness between the relationship

of completeness and credibility. However, in Stromwall & Granhag's (2005) study, adults' performance in distinguishing true and false accounts has not been improved when assessing credibility in a sequential manner rather than at the end of the interview process in previous research.

How are Props and Memory Cues Used to Enhance Children's Recall?

To facilitate children's recall, the EFI recommends the use of several types of props and memory cues, especially when interviewing very young children (Carnes et al., 2001). These may include the body parts inventory and the touch survey (Hewitt, 1999), as well as dolls and drawings with or without anatomical detail. The SAFE guideline allows the use of similar materials, although the use of the body touch survey was discontinued after the early years of the model's implementation (Azzopardi et al., 2004). Props and drawing materials play a central role in the Barnahus Sequential Interview as well, as they are seen as child-friendly methods to use when questioning pre-schoolers (Langballe & Davik, 2017). Before the interview begins, picture books, puzzles and drawing might be used to evaluate the child's cognitive and social functioning, and the child is encouraged to draw and use props during the interview alongside providing a verbal description of the allegation.

Although *Achieving Best Evidence* guidelines (Ministry of Justice, 2011) do not provide specific guidance about using props in multiple interviews, the document does discuss the use of interview aids in general. It is suggested that props may be used in a variety of ways, including as a method of assessing children's language and understanding, and to enable children to provide details through alternative communication methods in which they might be more competent than in verbal conversation. However, ABE guidelines acknowledge that props may encourage children to engage in fantasy or report inaccurate details, and this could ultimately damage the credibility of the case. Thus, the ABE states that

props should be used with caution, with care taken to ensure that interviewers avoid asking leading questions whilst children are engaging with interview aids.

Evaluation of the use of props and memory cues

Reviews of past research on multiple interviews demonstrate that providing children with multiple recall occasions does not compromise the accuracy of their accounts when the interviews are non-suggestive, however, the use of leading interviewing methods can lead to sharp decreases in accuracy over multiple interviews (Goodman & Quas, 2008; La Rooy et al., 2009). Although all multiple interviewing protocols emphasise the importance of using non-leading questioning, some of the props and methods used in the EFI model and its successors are controversial. For instance, a study comparing the accuracy of accounts elicited by trained police officers using structured interviewing protocols with prop-assisted interviews conducted by clinicians found that whilst children's recall remained highly accurate in the structured interview condition, the accuracy of prop-assisted recall decreased in the second interview (Melinder et al., 2010). Considering that multiple interviews are most frequently used when interviewing pre-schoolers, the most suggestible age group according to some studies (Ceci & Crotteau Huffman, 1997; Leichtman & Ceci, 1994), investigators should exercise a high degree of caution when using props in multiple interviews.

Conclusion and Recommendations

Altogether, EFI guidelines successfully address all three components of narrative coherence in multiple interviews. Allowing children with multiple opportunities to report details related to the allegation increases the completeness of their accounts, using the last session (in both the EFI and the Barnahus Sequential Interview) to clarify unclear details helps avoid contradictions in children's accounts, and progressing through the allegations in multiple interviews rather than repeating the same questions about the allegations in separate

interviews may help children narrate their accounts in a causally and temporally connected manner.

However, multiple interviewing guidelines clearly prioritise completeness over consistency and connectedness, emphasising the importance of gaining a picture of the full extent of the allegations (Carnes et al., 1999; 2001). The emphasis on completeness is consistent with results from Study 2, which suggest that from the three components of narrative coherence, completeness is the most strongly correlated with accuracy. Thus, ensuring that children are able to provide a complete account of the events they are questioned about does not risk a decrease in the accuracy of the witness statement, but on the contrary, helps them describe the events accurately. The prioritisation of completeness over the other two components of narrative coherence is also supported by the results of Study 3 suggesting that completeness is the most important aspect of the witness statement in terms of determining adults' impressions of the credibility of children's statements.

Altogether, current guidelines' focus on increasing the completeness of children's accounts appears effective both in gaining accurate information and in ensuring that the credibility of the witness statement is not compromised in court. However, based on the results of the current studies and a review of the literature on the effects of multiple interviews on the accuracy, completeness and credibility of children's accounts, four key recommendations can be suggested to be implemented in research-driven guidelines to ensure that interviews are conducted in an effective but non-leading manner.

Recommendation 1: Multiple interviews are an effective method of increasing the completeness of the account of witnesses who are unable to report the full extent of the allegations in a single interview. Victims' inability to provide a complete narrative of the alleged abuse can stem from multiple reasons, including their age, gender or experience of

multiple instances of abuse. Multiple interviews should be available to all victims in need of a second recall occasion, not only the youngest ones.

Recommendation 2: Investigators conducting multiple interviews should be mindful of the effects of delay on the accuracy of children's testimony. This is especially important in the case of many-session interview protocols, such as the 6-session or 8-session EFI model, as these might result in lengthy delays between the first and last interviews, and evidence does not unanimously support the view that the in-between interviews 'inoculate' children against the damaging effects of delay.

Recommendation 3: Credibility assessments should always be conducted by a team rather than an individual to minimize the effect of bias introduced by individual differences between evaluators, such as gender. Additionally, guidelines and checklists used when assessing the credibility of child witnesses following the interview process should adhere to court admissibility criteria, as credibility decisions made at this stage determine whether the case will progress through the criminal system to be tried in court.

Recommendation 4: Investigators should exercise a high degree of caution when using props of any kind in multiple interviews with very young children, as multiple interviews may exacerbate the effects of leading questioning methods on the accuracy of children's testimony.

5.6 Conclusion

The present thesis was the first to investigate the impact of multiple interviews on the narrative coherence, accuracy and credibility of children's testimony using complementary field and experimental methodologies. Results show that multiple interviews increase the narrative coherence of children's testimonies both in real forensic interviews and under controlled conditions, through providing children with an opportunity to report more details

related to crucial aspects of the allegations. Whilst multiple interviews increased the completeness of children's testimonies, their consistency and connectedness remained constant. The narrative coherence of children's accounts also increased when interviewers used questioning strategies that promoted connectedness either through providing linguistic scaffolding or through allowing children to rely on their own causal-temporal framework to connect details in the narrative.

Study 2 showed that the completeness and consistency of children's reports of a staged experiment correlated with the proportion of accurate details in their narratives. In addition, mock jurors developed more positive impressions of a fictional child witness in Study 3 when the child's testimony was higher in completeness and consistency and were more likely to convict the defendant if they were presented with two interviews than in the single interview condition. Verdict decisions were not influenced by the age of the child witness or the consistency of the testimony. Further research is needed to establish whether the connectedness of children's testimonies influences their perceived credibility.

Whilst all types of witnesses benefitted from multiple interviews, the increase in completeness was most substantial when the witness was older in Study 1, and when the witness was female in Study 2. Altogether, findings from the three studies indicate that multiple interviews effectively increase the completeness, and therefore, the credibility of children's testimony, without inherently compromising their accuracy. However, Study 2 showed that a 1-month delay between interviews led to a decrease in the proportion of correct information reported by 8-11-year-olds, suggesting that the timing of interviews is crucial in avoiding the potential risks associated with conducting multiple interviews.

Whilst the key points of EFI-based multiple interview guidelines are supported by the present studies and previous research, changes are recommended to the protocols in the areas

of availability for different types of witnesses, scheduling interviews, determining case outcomes and using props to enhance children's recall. Future research investigating the impact of multiple interviews consisting of three or more sessions and the influence of individual differences on the narrative coherence, accuracy and credibility of children's testimonies is necessary to make recommendations about the suitability of multiple interviews for specific investigations.

References

Ackil, J. K., Van Abbema, D. L., & Bauer, P. J. (2003). After the storm: Enduring differences in mother-child recollections of traumatic and nontraumatic events. *Journal of Experimental Child Psychology*, *84*(4), 286–309. http://dx.doi.org/10.1016/s0022-0965(03)00027-4.

Ahern, E. C., & Lyon, T. D. (2013). Facilitating maltreated children's use of emotional language. *Journal of Forensic Social Work*, *3*(2), 176–203. https://doi.org/10.1080/1936928X.2013.854124

Amado, B. G., Arce, R., & Fariña, F. (2015). Undeutsch hypothesis and Criteria Based Content Analysis: A meta-analytic review. *The European Journal of Psychology Applied to Legal Context*, *7*(1), 3-12. https://doi.org/10.1016/j.ejpal.2014.11.002

Andrews, S. J., Lamb, M. E., & Lyon, T. D. (2015). The effects of question repetition on responses when prosecutors and defense attorneys question children alleging sexual abuse in court. *Law and Human Behavior*, *39*(6), 559–570. doi:10.1037/lhb0000152

Azad, A., Christianson, S. Å., & Selenius, H. (2014). Children's reporting patterns after witnessing homicidal violence – the effect of repeated experience and repeated interviews. *Psychology, Crime & Law*, *20*(5), 407–429. https://doi.org/10.1080/1068316X.2013.793337

Azzopardi, C., Madigan, S., & Kirkland-Burke, M. (2014). Sexual abuse forensic evaluation with young children: program outcomes and predictors of disclosure. *Journal of Child Custody*, *11*(4), 304–324. https://doi.org/10.1080/15379418.2014.988901

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. https://doi.org/10.2307/1131550

Baker-Ward, Lynne, Hess, T. M., & Flannagan, D. A. (1990). The effects of involvement on children's memory for events. *Cognitive Development*, *5*(1), 55–69. https://doi.org/10.1016/0885-2014(90)90012-I

Ballard, P. B. (1913). Obliviscence and reminiscence. *British Journal of Psychology*, *1*(2), 82–82.

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

Bennett, W. L., & Feldman, M. S. (1981). *Reconstructing Reality in the Courtroom*. London, UK: Tavistock Publications Ltd.

Black, J. B., & Bern, H. (1981). Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior*, *20*(3), 267–275. https://doi.org/10.1016/S0022-5371(81)90417-5

Bottoms, B. L., Shaver, P. R., & Goodman, G. S. (1996). An analysis of ritualistic and religion-related child abuse allegations. *Law and Human Behavior*, *20*(1), 1–34. https://doi.org/10.1007/BF01499130

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology*, *71*(2), 81–129. https://doi.org/10.1006/jecp.1998.2464

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

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

Brubacher, S. P., Powell, M. B., & Roberts, K. P. (2014). Recommendations for interviewing children about repeated experiences. *Psychology, Public Policy, and Law*, *20*(3), 325–335. https://doi.org/10.1037/law0000011

Bruck, M., Ceci, S. J., & Hembrooke, H. (2002). The nature of children's true and false narratives. *Developmental Review*, *22*(3), 520–554. https://doi.org/10.1016/S0273-2297(02)00006-0

Carnes, C. N., Nelson-Gardell, D., Wilson, C., & Orgassa, U. C. (2001). Extended forensic evaluation when sexual abuse is suspected: A multisite field study. *Child Maltreatment*, *6*(3), 230–242. https://doi.org/10.1177/1077559501006003004

Carnes, C. N., Wilson, C., & Nelson-Gardell, D. (1999). Extended forensic evaluation when sexual abuse is suspected: A model and preliminary data. *Child Maltreatment*, *4*(3), 242–254. https://doi.org/10.1177/1077559599004003005

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

Castelli, P., Goodman, G. S., & Ghetti, S. (2005). Effects of interview style and witness age on perceptions of children's credibility in sexual abuse cases. *Journal of Applied Social Psychology*, *35*(2), 297–317. https://doi.org/10.1111/j.1559-1816.2005.tb02122.x

Ceci, S. J., & Crotteau Huffman, M. L. (1997). How suggestible are preschool children? Cognitive and social factors. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*(7), 948–958. https://doi.org/10.1097/00004583-199707000-00017

Ceci, Stephen 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

Clancy, S. A. (2009). *The trauma myth: The truth about the sexual abuse of children—and its aftermath*. New York, New York, US: Basic Books.

Collins, K., Doherty-Sneddon, G., & Doherty, M. J. (2014). Practitioner perspectives on rapport building during child investigative interviews. *Psychology, Crime & Law*, *20*(9), 884–901. https://doi.org/10.1080/1068316x.2014.888428

Connell, M. (2009). The extended forensic evaluation. In *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 451–487). Hoboken, New Jersey, US: John Wiley & Sons Inc.

Connolly, D. A., & Gordon, H. M. (2014). Can order of general and specific memory prompts help children to recall an instance of a repeated event that was different from the others? *Psychology, Crime & Law*, *20*(9), 852–864. https://doi.org/10.1080/1068316X.2014.885969

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

Connolly, D. A., & Read, J. (2006). Delayed prosecutions of historic child sexual abuse: Analyses of 2064 Canadian criminal complaints. *Law and Human Behavior*, *30*(4), 409–434. https://doi.org/10.1007/s10979-006-9011-6

Cross, T. P., Jones, L. M., Walsh, W. A., Simone, M., & Kolko, D. (2007). Child forensic interviewing in Children's Advocacy Centers: Empirical data on a practice model. *Child Abuse & Neglect*, *31*(10), 1031–1052. https://doi.org/10.1016/j.chiabu.2007.04.007

Crowley, M. J., O'Callaghan, M. G., & Ball, P. J. (1994). The juridical impact of psychological expert testimony in a simulated child sexual abuse trial. *Law and Human Behavior*, *18*(1), 89–105. https://doi.org/10.1007/BF01499146

Davies, G. (1991). Children on Trial? Psychology, videotechnology and the law. *The Howard Journal of Criminal Justice*, *30*(3), 177–191. https://doi.org/10.1111/j.1468-2311.1991.tb00707.x

Davies, G. M., Westcott, H. L., & Horan, N. (2000). The impact of questioning style on the content of investigative interviews with suspected child sexual abuse victims. *Psychology, Crime & Law*, *6*(2), 81–97. https://doi.org/10.1080/10683160008410834

Davies, M., & Rogers, P. (2009). Perceptions of blame and credibility toward victims of childhood sexual abuse: Differences across victim age, victim-perpetrator relationship, and respondent gender in a depicted case. *Journal of Child Sexual Abuse*, *18*(1), 78–92. https://doi.org/10.1080/10538710802584668

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

Ebbinghaus, H. (1948). Concerning memory, 1885. In *Century Psychology Series*. *Readings in the history of psychology* (pp. 304–313). https://doi.org/10.1037/11304-034

Eltz, M. J., Shirk, S. R., & Sarlin, N. (1995). Alliance formation and treatment outcome among maltreated adolescents. *Child Abuse & Neglect*, *19*(4), 419–431. https://doi.org/10.1016/0145-2134(95)00008-v

Erdelyi, M. H. (1996). *The recovery of unconscious memories: Hypermnesia and reminiscence*. Chicago, Illinois, US: University of Chicago Press.

Faller, K. C. (2015). Forty years of forensic interviewing of children suspected of sexual abuse, 1974–2014: Historical Benchmarks. *Social Sciences*, *4*(1), 34–65. https://doi.org/10.3390/socsci4010034

Faller, K. C., Cordisco-Steele, L., & Nelson-Gardell, D. (2010). Allegations of sexual abuse of a child: What to do when a single forensic interview isn't enough. *Journal of Child Sexual Abuse*, *19*(5), 572–589. https://doi.org/10.1080/10538712.2010.511985

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149

Feltis, B. B., Powell, M. B., & Roberts, K. P. (2011). The effect of event repetition on the production of story grammar in children's event narratives. *Child Abuse & Neglect*, *35*(3), 180–187. https://doi.org/10.1016/j.chiabu.2010.11.004

Feltis, B. B., Powell, M. B., Snow, P. C., & Hughes-Scholes, C. H. (2010). An examination of the association between interviewer question type and story-grammar detail in child witness interviews about abuse. *Child Abuse & Neglect*, *34*(6), 407–413. https://doi.org/10.1016/j.chiabu.2009.09.019

Fivush, R., Haden, C. A., & Reese, E. (2006). Elaborating on elaborations: Role of maternal reminiscing style in cognitive and socioemotional development. *Child Development*, *77*(6), 1568–1588. https://doi.org/10.1111/j.1467-8624.2006.00960.x

Fivush, R., Hazzard, A., McDermott Sales, J., Sarfati, D., & Brown, T. (2003). Creating coherence out of chaos? Children's narratives of emotionally positive and negative events. *Applied Cognitive Psychology*, *17*(1), 1–19. https://doi.org/10.1002/acp.854

Flin, R., Boon, J., Knox, A., & Bull, R. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology*, *83*(3), 323–336. https://doi.org/10.1111/j.2044-8295.1992.tb02444.x

Fogliati, R., & Bussey, K. (2014). The effects of cross-examination on children's reports of neutral and transgressive events. *Legal and Criminological Psychology*, *19*(2), 296–315. https://doi.org/10.1111/lcrp.12010

Gabora, N. J., Spanos, N. P., & Joab, A. (1993). The effects of complainant age and expert psychological testimony in a simulated child sexual abuse trial. *Law and Human Behavior*, *17*(1), 103–119. https://doi.org/10.1007/BF01044540

Gee, S., & Pipe, M.-E. (1995). Helping Children to Remember: The Influence of Object Cues on Children's Accounts of a Real Event. *Developmental Psychology*, *31*(5), 746–758. https://doi.org/10.1037/0012-1649.31.5.746

Ghetti, S., Goodman, G. S., Eisen, M. L., Qin, J., & Davis, S. L. (2002). Consistency in children's reports of sexual and physical abuse. *Child Abuse & Neglect*, *26*(9), 977–995. https://doi.org/10.1016/S0145-2134(02)00367-8

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. (2006). Children's eyewitness memory: A modern history and contemporary commentary. *Journal of Social Issues*, *62*(4), 811–832. https://doi.org/10.1111/j.1540-4560.2006.00488.x

Goodman, G. S., Bottoms, B. L., Herscovici, B. B., & Shaver, P. (1989). Determinants of the child victim's perceived credibility. In S. J. Ceci, D. F. Ross, & M. P. Toglia (Eds.), *Perspectives on Children's Testimony* (pp. 1–22). New York, New York, US: Springer-Verlag Publishing.

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., Golding, J. M., Helgeson, V. S., Haith, M. M., & Michelli, J. (1987). When a child takes the stand. *Law and Human Behavior*, *11*(1), 27–40. https://doi.org/10.1007/BF01044837

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

Gordon, B. N., & Follmer, A. (1994). Developmental issues in judging the credibility of children's testimony. *Journal of Clinical Child Psychology*, *23*(3), 283–294. https://doi.org/10.1207/s15374424jccp2303_6

Graffam Walker, A., Kenniston, J., Inada, S. S., & Caldwell, C. (2013). *Handbook on Questioning Children: A Linguistic Perspective*. ABA Center on Children and the Law.

Graffam Walker, A., & Warren, A. R. (1995). The language of the child abuse interview: Asking the questions, understanding the answers. In T. Ney (Ed.) *True and false allegations of child sexual abuse: Assessment and case management* (pp. 153–162). New York, New York, US: Brunner/Mazel.

Greenhoot, A. F., Ornstein, P. A., Gordon, B. N., & Baker-Ward, L. (1999). Acting out the details of a pediatric check-up: The impact of interview condition and behavioral style on children's memory reports. *Child Development*, *70*(2), 363–380. https://doi.org/10.1111/1467-8624.00027

Gudjonsson, G., Vagni, M., Maiorano, T., & Pajardi, D. (2016). Age and memory related changes in children's immediate and delayed suggestibility using the Gudjonsson Suggestibility Scale. *Personality and Individual Differences*, *102*, 25–29. https://doi.org/10.1016/j.paid.2016.06.029

Habermas, T., & Bluck, S. (2000). Getting a life: The emergence of the life story in adolescence. *Psychological Bulletin*, *126*(5), 748–769. https://doi.org/10.1037//0033-2909.126.5.748

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

Henderson, H. M., & Andrews, S. J. (2018). Assessing the veracity of children's forensic interviews: Implications for the courtroom. In H. Otgaar & M. L. Howe (Eds.) *Finding the truth in the courtroom: Dealing with deception, lies, and memories* (pp. 103–135). Oxford, UK: Oxford University Press.

Hershkowitz, I., Fisher, S., Lamb, M. E., & Horowitz, D. (2007). Improving credibility assessment in child sexual abuse allegations: The role of the NICHD investigative interview protocol. *Child Abuse & Neglect*, *31*(2), 99–110. https://doi.org/10.1016/j.chiabu.2006.09.005

Hershkowitz, I., Orbach, Y., Lamb, M. E., Sternberg, K. J., & Horowitz, D. (2006). Dynamics of forensic interviews with suspected abuse victims who do not disclose abuse. *Child Abuse & Neglect*, *30*(7), 753–769. https://doi.org/10.1016/j.chiabu.2005.10.016

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

Hewitt, S. K. (1999). *Assessing allegations of sexual abuse in preschool children: Understanding small voices*. Thousand Oaks, California: Sage Publications, Inc.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *The American Statistician*, *55*(1), 19-24. https://doi.org/10.1198/000313001300339897

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*, *23*(6), 849-862. https://doi.org/10.1080/13218719.2016.1256015

Jack, F., & Zajac, R. (2014). The effect of age and reminders on witnesses' responses to cross-examination-style questioning. *Journal of Applied Research in Memory and Cognition*, *3*(1), 1–6. https://doi.org/10.1016/j.jarmac.2013.12.001

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

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67–85. https://doi.org/10.1037/0033-295X.88.1.67

Katz, C., & Hershkowitz, I. (2012). Repeated interviews with children who are the alleged victims of sexual abuse. *Research on Social Work Practice*, *23*(2), 210–218. https://doi.org/10.1177/1049731512467511

Klippenstine, M. A., & Schuller, R. (2012). Perceptions of sexual assault: Expectancies regarding the emotional response of a rape victim over time. *Psychology, Crime & Law*, *18*(1), 79–94. https://doi.org/10.1080/1068316X.2011.589389

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

Krix, A. C., Sauerland, M., Lorei, C., & Rispens, I. (2015). Consistency across repeated eyewitness interviews: contrasting police detectives' beliefs with actual eyewitness performance. *PLoS ONE*, *10*(2). https://doi.org/10.1371/journal.pone.0118641

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. https://curve.coventry.ac.uk/open/file/7ee04f6f-6f2e-43c5-8958-61d7015c8078/1/prosecution%20reliance.pdf

La Rooy, D. L., Brubacher, S. P., Aromäki-Stratos, A., Cyr, M., Hershkowitz, I., Korkman, J., … 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 eyewitness memory. *Journal of Experimental Child Psychology*, *90*(3), 235–254. https://doi.org/10.1016/j.jecp.2004.11.002

La Rooy, D., Lamb, M. E., & Pipe, M.-E. (2009). Repeated interviewing: A critical evaluation of the risks and potential benefits. In Kuehnle, K. & Connell, M. (Eds.) *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 327–361). Hoboken, NJ, US: John Wiley & Sons Inc.

Lamb, M. E. (1994). The investigation of child sexual abuse: An international, interdisciplinary consensus statement. *Family Law Quarterly*, *28*(1), 151–162. https://doi.org/10.1016/0145-2134(94)90127-9

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

Lamb, M. E., Orbach, Y., Hershkowitz, I., Esplin, P. W., & Horowitz, D. (2007). 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., Sternberg, K. J., Aldridge, J., Pearson, S., Stewart, H. L., … Bowler, L. (2009). Use of a structured investigative protocol enhances the quality of investigative interviews with alleged victims of child sexual abuse in Britain. *Applied Cognitive Psychology*, *23*(4), 449–467. https://doi.org/10.1002/acp.1489

Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., Stewart, H., & Mitchell, S. (2003). Age differences in young children's responses to open-ended invitations in the course of forensic interviews. *Journal of Consulting and Clinical Psychology*, *71*(5), 926–934. https://doi.org/10.1037/0022-006X.71.5.926

Landry, K. L., & Brigham, J. C. (1992). The effect of training in Criteria-Based Content Analysis on the ability to detect deception in adults. *Law and Human Behavior*, *16*(6), 663–676. https://doi.org/10.1007/BF01884022

Langballe, Å., & Davik, T. (2017). Sequential interviews with preschool children in Norwegian Barnahus. In S. Johansson, K. Stefansen, E. Bakketeig, & A. Kaldal (Eds.), *Collaborating Against Child Abuse: Exploring the Nordic Barnahus Model* (pp. 165–183). London, UK: Palgrave macmillan.

Langley, H. A., Coffman, J. L., & Ornstein, P. A. (2017). The socialization of children's memory: Linking maternal conversational style to the development of children's autobiographical and

deliberate memory skills. *Journal of Cognition and Development*, *18*(1), 63–86. https://doi.org/10.1080/15248372.2015.1135800

Larkina, M., & Bauer, P. J. (2010). The role of maternal verbal, affective, and behavioral support in preschool children's independent and collaborative autobiographical memory reports. *Cognitive Development*, *25*(4), 309–324. https://doi.org/10.1016/j.cogdev.2010.08.008

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

Leippe, M. R., & Romanczyk, A. (1989). Reactions to child (versus adult) eyewitnesses. *Law and Human Behavior*, *13*(2), 103–132. https://doi.org/10.1007/BF01055919

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*(4), 361–366. https://doi.org/10.1101/lm.94705

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

Luus, C. A. E., Wells, G. L., & Turtle, J. W. (1995). Child eyewitnesses: Seeing is believing. *Journal of Applied Psychology*, *80*(2), 317–326. https://doi.org/10.1037/0021-9010.80.2.317

Lyon, T. D., Scurich, N., Choi, K., Handmaker, S., & Blank, R. (2012). "How did you feel?": Increasing child sexual abuse witnesses' production of evaluative information. *Law and Human Behavior*, *36*(5), 448–457. https://doi.org/10.1037/h0093986

MacCormick, N. (2005). *Rhetoric and the rule of law: A theory of legal reasoning*. Oxford, UK: Oxford University Press.

Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law*, *11*(1), 99–122. https://doi.org/10.1080/10683160410001726356

McAdams, D. P. (2006). The Problem of Narrative Coherence. *Journal of Constructivist Psychology*, *19*(2), 109–125. https://doi.org/10.1080/10720530500508720

McCauley, M. R., & Parker, J. F. (2001). When will a child be believed? The impact of the victim's age and juror's gender on children's credibility and verdict in a sexual-abuse case. *Child Abuse & Neglect*, *25*(4), 523–539. https://doi.org/10.1016/S0145-2134(01)00224-1

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

Merritt, K. A., Ornstein, P. A., & Spicker, B. (1994). Children's memory for a salient medical procedure: Implications for testimony. *Pediatrics*, *94*(1), 17–23. https://pediatrics.aappublications.org/content/94/1/17.long

Ministry of Justice. (2011). *Achieving best evidence in criminal proceedings: Guidance on interviewing victims and witnesses, and guidance on using special measures*. Retrieved from http://www.justice.gov.uk/downloads/victims-and-witnesses/vulnerablewitnesses/achieving-bestevidence-criminal-proceedings.pdf.

Morris, G., Baker-Ward, L., & Bauer, P. J. (2010). What remains of that day: The survival of children's autobiographical memories across time. *Applied Cognitive Psychology*, *24*(4), 527–544. https://doi.org/10.1002/acp.1567

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

O'Kearney, R., Speyer, J., & Kenardy, J. (2007). Children's narrative memory for accidents and their post-traumatic distress. *Applied Cognitive Psychology*, *21*(7), 821–838. https://doi.org/10.1002/acp.1294

O'Keefe, D. J. (2007). Brief report: post hoc power, observed power, a priori power, retrospective power, prospective power, achieved power: sorting out appropriate uses of statistical power analyses. *Communication methods and measures*, *1*(4), 291-299. https://doi.org/10.1080/19312450701641375

O'Neill, S., & Zajac, R. (2013). The role of repeated interviewing in children's responses to cross-examination-style questioning: Cross-examination and repeated interviewing. *British Journal of Psychology*, *104*(1), 14–38. https://doi.org/10.1111/j.2044-8295.2011.02096.x

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

Otgaar, H., Verschuere, B., Meijer, E. H., & van Oorsouw, K. (2012). The origin of children's implanted false memories: Memory traces or compliance? *Acta Psychologica*, *139*(3), 397–403. https://doi.org/10.1016/j.actpsy.2012.01.002

Parker, A. D., & Brown, J. (2000). Detection of deception: Statement Validity Analysis as a means of determining truthfulness or falsity of rape allegations. *Legal and Criminological Psychology*, *5*(2), 237–259. https://doi.org/10.1348/135532500168119

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. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 521–533. https://doi.org/10.1037/0278-7393.14.3.521

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. (2011). Children's memory reports over time: Getting both better and worse. *Journal of Experimental Child Psychology*, *109*(3), 275–293. https://doi.org/10.1016/j.jecp.2011.01.009

Peterson, C. (2015). A decade later: Adolescents' memory for medical emergencies: A decade later. *Applied Cognitive Psychology*, *29*(6), 826–834. https://doi.org/10.1002/acp.3192

Peterson, C., & Bell, M. (1996). Children's memory for traumatic injury. *Child Development*, *67*(6), 3045–3070. https://doi.org/10.1111/j.1467-8624.1996.tb01902.x

Peterson, C., & Biggs, M. (1998). Stitches and casts: Emotionality and narrative coherence. *Narrative Inquiry*, *8*(1), 51–76. https://doi.org/10.1075/ni.8.1.04pet

Peterson, C., & McCabe, A. (1983). *Developmental Psycholinguistics: Three Ways of Looking at a Child's Narrative*. New York, NY, US: Springer Publishing Co.

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

Peterson, C., Pardy, L., Tizzard-Drover, T., & Warren, K. L. (2005). When initial interviews are delayed a year: Effect on children's 2-Year recall. *Law and Human Behavior*, *29*(5), 527–541. https://doi.org/10.1007/s10979-005-6833-6

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

Pipe, M.-E., Sutherland, R., Webster, N., Jones, C., & Rooy, D. L. (2004). Do early interviews affect children's long-term event recall? *Applied Cognitive Psychology*, *18*(7), 823–839. https://doi.org/10.1002/acp.1053

Plotnikoff, J., & Woolfson, R. (2009). *Measuring Up?: Evaluating Implementation of Government Commitments to Young Witnesses in Criminal Proceedings*. London: NSPCC.

Plotnikoff, J., & Woolfson, R. (2001). *An Evaluation of Child Witness Support*. London: Nuffield Foundation.

Powell, M. B., & Roberts, K. P. (2002). The effect of repeated experience on children's suggestibility across two question types. *Applied Cognitive Psychology*, *16*(4), 367–386. https://doi.org/10.1002/acp.801

Powell, M. B., & Thomson, D. M. (1997). The effect of an intervening interview on children's ability to remember one occurrence of a repeated event. *Legal and Criminological Psychology*, *2*(2), 247–262. https://doi.org/10.1111/j.2044-8333.1997.tb00346.x

Price, H. L., & Connolly, D. A. (2013). Suggestibility effects persist after one year in children who experienced a single or repeated event. *Journal of Applied Research in Memory and Cognition*, *2*(2), 89–94. https://doi.org/10.1016/j.jarmac.2013.03.001

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

Principe, G. F., & Ceci, S. J. (2002). “I saw it with my own ears”: The effects of peer conversations on preschoolers’ reports of nonexperienced events. *Journal of Experimental Child Psychology*, *83*(1), 1–25. https://doi.org/10.1016/S0022-0965(02)00120-0

Quas, J. A., Goodman, G. S., Ghetti, S., & Redlich, A. D. (2000). Questioning the child witness. *Trauma, Violence, & Abuse*, *1*(3), 223–249. doi:10.1177/1524838000001003002

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

Quas, J. A., & Schaaf, J. M. (2002). Children’s memories of experienced and nonexperienced events following repeated interviews. *Journal of Experimental Child Psychology*, *83*(4), 304–338. https://doi.org/10.1016/s0022-0965(02)00150-9

Radford, L., Corral, S., Bradley, C., Fisher, H., Bassett, C., Howat, N., & Collishaw, S. (2011). *Child Abuse and Neglect in the UK Today*. London: NSPCC.

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

Reese, E., Haden, C. A., & Fivush, R. (1993). Mother-child conversations about the past: Relationships of style and memory over time. *Cognitive Development*, *8*(4), 403–430. https://doi.org/10.1016/S0885-2014(05)80002-4

Rideout, C. (2008). Storytelling, Narrative Rationality, and Legal Persuasion. *Legal Writing: The Journal of the Legal Writing Institute*, *14*, 53-86. Retrieved from: https://pdfs.semanticscholar.org/c2a0/b6b9e94d7a4b3e5bb2e80cfdf909f260f028.pdf?_ga=2.245881860.1697372375.1567528946-424814861.1567528946

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*(3), 274–296. https://doi.org/10.1080/1068316X.2014.951650

Roberts, K. P., & Powell, M. B. (2006). The consistency of false suggestions moderates children's reports of a single instance of a repeated event: Predicting increases and decreases in suggestibility. *Journal of Experimental Child Psychology*, *94*(1), 68–89. https://doi.org/10.1016/j.jecp.2005.12.003

Ross, D. F., Jurden, F. H., Lindsay, R. C. L., & Keeney, J. M. (2003). Replications and limitations of a two-factor model of child witness credibility. *Journal of Applied Social Psychology*, *33*(2), 418–431. https://doi.org/10.1111/j.1559-1816.2003.tb01903.x

Rudy, L., & Goodman, G. S. (1991). Effects of participation on children's reports: Implications for children's testimony. *Developmental Psychology*, *27*(4), 527–538. https://doi.org/10.1037/0012-1649.27.4.527

Salmon, K., & Pipe, M.-E. (1997). Props and children's event reports: The impact of a 1-year delay. *Journal of Experimental Child Psychology*, *65*(3), 261–292. https://doi.org/10.1006/jecp.1996.2362

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

Salmon, K., Pipe, M.-E., Malloy, A., & Mackay, K. (2012). Do non-verbal aids increase the effectiveness of 'Best Practice' verbal interview techniques? An experimental study. *Applied Cognitive Psychology*, *26*(3), 370–380. https://doi.org/10.1002/acp.1835

Saykaly, C., Talwar, V., Lindsay, R. C. L., Bala, N. C., & Lee, K. (2013). The influence of multiple interviews on the verbal markers of children's deception. *Law and Human Behavior*, *37*(3), 187–196. https://doi.org/10.1037/lhb0000023

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

Schneider, B. A., Avivi-Reich, M., & Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00474

Schneider, P., & Winship, S. (2002). Adults' judgments of fictional story quality. *Journal of Speech, Language, and Hearing Research: JSLHR*, *45*(2), 372–383. https://doi.org/10.1044/1092-4388(2002/029)

Stein, N. L., & Nezworski, T. (1978). The effects of organization and instructional set on story memory. *Discourse Processes*, *1*(2), 177–193. https://doi.org/10.1080/01638537809544433

Steller, M., & Koehnken, G. (1989). Criteria-based statement analysis. In Raskin, D. C. (Ed.) *Psychological methods in criminal investigation and evidence* (pp. 217–245). New York, NY, US: Springer Publishing Co.

Strömwall, L. A., Bengtsson, L., Leander, L., & Granhag, P. A. (2004). Assessing children's statements: The impact of a repeated experience on CBCA and RM ratings. *Applied Cognitive Psychology*, *18*(6), 653–668. https://doi.org/10.1002/acp.1021

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., Lamb, M., 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/10.17863/CAM.8180

Tabak, S. J., & Klettke, B. (2014). Mock jury attitudes towards credibility, age, and guilt in a fictional child sexual assault scenario: Credibility in child sexual assault. *Australian Journal of Psychology*, *66*(1), 47–55. https://doi.org/10.1111/ajpy.12035

Teoh, Y.-S., & Lamb, M. E. (2010). Preparing children for investigative interviews: Rapport-building, instruction, and evaluation. *Applied Developmental Science*, *14*(3), 154–163. https://doi.org/10.1080/10888691.2010.494463

The Scottish Executive. (2011). *Guidance on joint investigative interviewing of child witnesses in Scotland*. Retrieved from http://www.scotland.gov.uk/Publications/2011/12/

Thomas, A. K., Chen, C., Gordon, L. T., & Tenbrink, T. (2015). Choose your words wisely: What verbal hesitation indicates about eyewitness accuracy. *Applied Cognitive Psychology*, *29*(5), 735–741. https://doi.org/10.1002/acp.3157

Tizzard-Drover, T., & Peterson, C. (2004). The influence of an early interview on long-term recall: A comparative analysis. *Applied Cognitive Psychology*, *18*(6), 727–743. https://doi.org/10.1002/acp.1028

Undeutsch, U. (1989). The development of Statement Reality Analysis. In J. C. Yuille (Ed.), *Credibility Assessment* (pp. 101–119). https://doi.org/10.1007/978-94-015-7856-1_6

Voss, J. F., & Van Dyke, J. A. (2001). Narrative structure, information certainty, emotional content, and gender as factors in a pseudo jury decision-making task. *Discourse Processes*, *32*(2–3), 215–243. https://doi.org/10.1080/0163853X.2001.9651599

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). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Vrij, A. (2005). Criteria-Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, *11*(1), 3–41. https://doi.org/10.1037/1076-8971.11.1.3

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behavior*, *26*(3), 261–283. https://doi.org/10.1023/A:1015313120905

Wang, Q., Bui, V.-K., & Song, Q. (2015). Narrative organisation at encoding facilitated children's long-term episodic memory. *Memory*, *23*(4), 602–611. https://doi.org/10.1080/09658211.2014.914229

Waterhouse, G. (2016). *Investigating the forensic interviewing of children: Multiple interviews and social support* (Phd, London South Bank University). Retrieved from https://openresearch.lsbu.ac.uk/item/873yq

Waterhouse, G. F., Ridley, A. M., Bull, R., La Rooy, D., & Wilcock, R. (2016). Dynamics of repeated interviews with children. *Applied Cognitive Psychology*, *30*(5), 713–721. https://doi.org/10.1002/acp.3246

Wessel, E., Magnussen, S., & Melinder, A. M. D. (2013). Expressed emotions and perceived credibility of child mock victims disclosing physical abuse: Child expressed emotions and perceived credibility. *Applied Cognitive Psychology*, *27*(5) 611-616. https://doi.org/10.1002/acp.2935

Westcott, H. L., & Kynan, S. (2004). The application of a 'story-telling' framework to investigative interviews for suspected child sexual abuse. *Legal and Criminological Psychology*, *9*(1), 37–56. https://doi.org/10.1348/135532504322776843

Willmott, D., Boduszek, D., Debowska, A., & Woodfield, R. (2018). Introduction and validation of the Juror Decision Scale (JDS): An empirical investigation of the Story Model. *Journal of Criminal Justice*, *57*, 26–34. https://doi.org/10.1016/j.jcrimjus.2018.03.004

Wood, B., Orsak, C., Murphy, M., & Cross, H. J. (1996). Semistructured child sexual abuse interviews: Interview and child characteristics related to credibility of disclosure. *Child Abuse & Neglect*, *20*(1), 81–92.

Yale, R. N. (2013). Measuring narrative believability: Development and validation of the narrative believability scale (NBS-12). *Journal of Communication*, *63*(3), 578–599. https://doi.org/10.1111/jcom.12035

Yozwiak, J. A., Golding, J. M., & Marsil, D. F. (2004). The impact of type of out-of-court disclosure in a child sexual assault trial. *Child Maltreatment*, *9*(3), 325–334. https://doi.org/10.1177/1077559504266518

Appendices

Appendix A

As the probability of Type 1 errors is elevated in within-subject designs involving an interaction between the covariate and the dependent variable (Schneider et al., 2015), alternative statistical tests were conducted corresponding to each ANCOVA, without the covariate.

Number of Details

To assess whether investigators were aware of significantly more details after the second interview than after the first one, a paired t-test was conducted comparing the number of new details children recalled in the first interview and across the two interviews. Significantly less new details were recalled in a single interview (M = 121.75, SD = 147.48) than across the two interviews (M = 221.04, SD = 250.70), $t(27) = -4.27$, $p < .001$, $d = .48$, 95% CI = [-147.03, -51.54].

Consistency

To assess whether the proportion of consistent responses was affected by interview number, a paired t-test was conducted comparing the consistency of repeated information in the first interview and across the two interviews. There was no significant difference between the proportion of consistent responses in the first interview (M = .87, SD = .15) and overall (M = .84, SD = .16), $t(27) = .61$, $p = .55$, $d = .08$, 95% CI = [-.03, .06].

Connectedness

To assess whether children’s accounts were significantly more connected in the second interview than in the first one, a RM-ANOVA was conducted assessing the potential effect of interview number (first interview, overall) on the frequency of the four types of

connectedness markers (simple temporal, complex temporal, causal, optional) witnesses reported. A significant main effect was found for connectedness, $F(3, 81) = 4.04$, $p = .01$, $\eta_p^2 = .13$. No main effect was found for interview number, $F(1, 27) = 1.39$, $p = .25$, $\eta_p^2 = .05$. No significant interactions were found.

Appendix B

Would pupils know how to help in an emergency?

Royal Holloway, University of London is offering a free first aid course for Key Stage 2 pupils as part of a Psychology study investigating children's memory

- The Children's Memory Research Team at Royal Holloway, University of London is recruiting children to take part in a short first aid training session and two 10-minute interviews about their experience
- Give your pupils the opportunity to learn how to help in an emergency situation...
- ... While also contributing to psychological research on how children remember their experiences

For more information:
Zsofia.Szojka.2016@live.rhul.ac.uk
07871541649
Royal Holloway,
University of London
Egham, Surrey, TW20 0EX

ROYAL HOLLOWAY UNIVERSITY OF LONDON

What will happen during the first aid course and interviews?

The Life.Live it first aid course

- The Life.Live it first aid course was developed by The British Red Cross to teach children aged 7 to 11 how to stay safe in risky situations and how to help others who are hurt or are feeling ill
- In this 15-minute training session, a qualified first aider will introduce children to basic first aid, including practising calling 999 in case of an emergency and learning how to help someone who is choking or unresponsive
- At the end of the training, children will receive a first aid certificate

The interviews

- First 1 week then 5 weeks after the training session, children will meet a trained interviewer who will ask them to recall everything they can remember about the session in 10 minutes or less
- At the end of the interviews, children will receive a small gift as a token of our appreciation for taking part in the study

The aim of the study

- The Children's Memory Research Team at Royal Holloway, University of London investigates how children recall events they experienced when questioned by a trained interviewer
- Our primary goal is to improve the methods the police use when questioning children who are victims or witnesses of crime, but our findings might also benefit teachers, social workers and parents

ROYAL HOLLOWAY UNIVERSITY OF LONDON

First aid can save lives in an emergency

Give your pupils the opportunity to learn how to help in a fun, interactive way

If you are interested, contact Zsofia Szojka or Dr David La Rooy for further information

Zsofia Szojka
PhD student
Zsofia.Szojka.2016@rhul.ac.uk
07871541649

Dr David La Rooy
Senior Lecturer of Psychology
David.LaRooy@rhul.ac.uk
07954401711

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Appendix C

Dear Parents and Caregivers,

The Children's Memory Research Team of the Department of Law, Criminology and Sociology at Royal Holloway, University of London would like to invite your child to participate in a study that will investigate how children remember their experiences.

The project involves 6- to 11-year-old children. They will be invited to participate in pairs in a short first aid training session designed by The British Red Cross specifically for children in primary school. A female research assistant will introduce your child to the contents of a first aid kit and demonstrate how to help in case of an emergency. The training session will be playful, and the research assistant will use child-friendly language. This activity will take place on the 11th and 12th of June and will take about 12 minutes for each pair.

Following the British Red Cross first aid training session your child will also be asked to participate in two interviews during which they will be asked to recall all that they can about what happened during the training session. The interviews will take place on the 18th and 19th of June and the 16th, 19th and 20th of July at [redacted] and will last no longer than 10 minutes. Each child will receive a first aider certificate and a small gift as a token of our appreciation after participating in the training and the interviews.

If you would like your child to take part, you can return the tear-off slip at the bottom of the page to your child's teacher or at the reception by the 11th of June.

Thank you,

Zsofia Szojka and Dr David La Rooy
Royal Holloway, University of London
Egham Hill
Egham
TW20 0EX

I acknowledge that I have read the attached information sheet and give permission for my child to participate in the first aid training and the follow-up interviews, under the direction of Zsofia Szojka and Dr. David La Rooy.

Child's name _____ Child's date of Birth _____

Parent's signature _____

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Children's Memory Study

INFORMATION SHEET FOR PARENTS AND CAREGIVERS

Thank you for showing an interest in this project. Please read this information sheet carefully before deciding whether or not you consent to your child taking part in the project. If you decide to consent to your child's participation, we thank you. If you decide not to give your consent for your child to take part there will be no disadvantage to you or your child of any kind and we thank you for considering our request.

What is the Aim of the Project?

This is a project conducted by a PhD student at Royal Holloway, University of London, on children's memory of experienced events. The purpose of this project is to understand how children's description of a past event changes when they are asked to recall what happened more than once. In some cases, researchers have found that repeated interviews about a past experience helped children remember more information. We would like to invite your child to take part.

What will Participants be Asked to Do?

Should you give your consent for your child to take part in this project, a female research assistant will invite your child to take part in a short first aid training session which will focus on how children can help in an emergency. The session will last for about ten minutes and will include playful activities and child-friendly language.

After the first aid session, your child will be interviewed twice about what happened over the following two months. The interviews will be conducted by a trained interviewer and take place during school time by arrangement with your child's teacher. The interviews will last for approximately ten minutes. As a small token of our appreciation, your child will receive a first aider certificate and a small gift such as a pencil or a rubber.

Can Participants Change their Mind and Withdraw from the Project?

You may withdraw your child's participation in the project at any time, without disadvantage to you or your child of any kind. You may consent to your child taking part in the first aid training without taking part in the study, in this case, no data will be collected about your child. Your child may also express that they do not wish to participate at any time during the training session or the interviews, in which case the research assistant will take them back to their class.

ROYAL HOLLOWAY UNIVERSITY OF LONDON

What Data or Information will be Collected and What Use will be Made of it?

During the interviews your child will be encouraged to provide an account of what happened during the first aid training session in their own words. The interviews will be audio recorded so that we can record the information that your child recalled. Your child will only be identified on the recordings by a number, to protect their identity. Only the researchers in Dr La Rooy's laboratory will have access to the information reported by your child, for the purposes of analysing the information. Results of the project may be published, but any data included in publications will in no way be linked to any specific child. We will provide you with a summary of the results of the project when it has been completed. You are most welcome to request a copy of the complete results of the project should you wish.

After the interviews, the two personal pieces of information collected about your child (their name and date of birth) will be kept securely for 6 months to allow you to withdraw your child's participation in the project retrospectively. After 6 months, these personal pieces of information will be destroyed. The raw data on which the results of the project depend, and which cannot be linked to an individual, will be retained in secure storage at Royal Holloway, University of London for ten years after which it will be destroyed.

What If Participants have any Questions?

If you have any questions about our project, either now or in the future, please feel free to contact

Zsofia Szojka
PhD student
Zsofia.Szojka.2016@live.rhul.ac.uk
07871541649

Dr David La Rooy
Senior Lecturer of Psychology
David.LaRooy@rhul.ac.uk
07954401711

This project has been reviewed and approved by the Ethics Committee of Royal Holloway, University of London.

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Appendix D

Event protocol – First Aid Training session

Materials:

- First aid for eyes wall poster
- First Aid Kit box (Contents: Kool Kidz Instant Cold Pack, HSE Eyepad dressings, children's plasters)
- "Emergency" label (with picture of ambulance), "Not emergency" label (with picture of ambulance crossed out), small pictures depicting injuries; nosebleed, unresponsive person, grazed elbow, very bad bleeding from a wound, minor sunburn
- Laptop for playing the two British Red Cross videos; how to help someone who is choking, how to help someone who is unresponsive but breathing
- Cards containing picture of bike accident / boy falling down / unconscious girl, map of route from hospital to location and a phone number.
- First aider certificates

Stages:

1. First aid box
2. Emergency or not
3. How to help a person who is choking or unconscious
4. Emergency call

Introduction

Hi *(children's names)*. I'm *(research assistant's name)*. Today we're going to learn what to do in an emergency. If you know what to do in an emergency, you can help your family and friends or other people you meet if they are in have been hurt or if they feel ill. First aid is really easy to learn – it's just a few simple steps and you can make a real difference!

Part 1: First aid kit

But first let's look into this box! What do you think this box is for? (*Wait for response)* Yes, it is a first aid box. Can you tell me what is inside? *(Wait for response)* Alright, these are some really good suggestions, let's look inside and see! *(Open box and take out contents)*

(Point plasters) Can you tell me what these are for? (*Wait for response)* Yes, we can use this if someone has a small injury and they are bleeding. Do you want to try putting this on your finger? (*Allow children to put plaster on finger)*

(Point at eye pad) And what do you think these are for? (*Wait for response)* Yes, we can use this if someone's eye is hurt. Do you want to try putting this over your eye? (*Allor children to put eye pad on)*

(Point at cold pack) How about this one? Can anyone guess what it's for? If you have hurt yourself, this can stop the swelling by cooling down the area. (*Allow children to touch it)* It doesn't feel cold, does it? Try to shake it! *(Help children shake up the cold pack)* It's cold now!

Part 2: Emergency or not

The first aid box can be really useful but sometimes people are hurt badly and need more than a plaster or some bandages to get better. When someone is seriously hurt, there is a medical emergency and you need to call 999 for an ambulance. Let's look at these pictures and try to decide which ones are emergencies and which ones are not.

(*First show children each picture and name the injuries they depict then let them sort the pictures into two categories – a choking person, an unresponsive person and a very bad bleeding are emergencies, a nosebleed, a grazed elbow and a minor sunburn are not)*

Very well done. If there is a medical emergency, it is important to tell an adult, call 999 and try to help the injured person if you can.

Part 3: How to help

So these children (*point at the pictures under the medical emergency label)* are hurt badly and need help. (*point at picture of girl choking)* Do you know what you can do to help someone who is choking? (*wait for response)* Very good, let's watch a video.

(*British Red Cross video showing how children can help someone who is choking – 56 seconds)*

So you can help someone who is choking by hitting them 5 times on the back. (*point at picture of boy helping unconscious girl)* Do you know what you can do to help someone who seems to be unconscious and is not responding to you? (*wait for response)* Very good, let's watch a video.

(British Red Cross video showing how children can help someone who is unconscious – 1 minute 28 seconds)

So you can help an unconscious person by turning them on their side and tilting their head back to make sure they can breathe. Well done, we can now call an ambulance!

Part 4: Calling for help

Do you remember what number you have to call when there is an emergency? (*wait for response)* Yes, that's right, it's 999. You can call 999 for free even if your mobile phone is out of credit or if you call from a pay phone.

Let's pretend I'm a phone operator. I'll give you each a card which shows you what the emergency is, where it happened at what number you are calling from. (*show example card #1)* So for example, if I have this card, then I would need to tell the operator that there is an emergency and someone is unconscious. Then, if the operator asks I will need to tell her my address and phone number I can tell her that the ambulance will need to take a left turn to the park where the accident happened and that she can call me back on 07645372815.

(*Choose which child goes first)* I'll give you these cards, dial 999 and tell me what happened, where I should send the ambulance and what number I can call you back on.

(*Give child card #2 with picture of bike accident, map and phone number and the toy phone)*

Hello, emergency service operator. Which service do you require? Fire, police or ambulance?

(*Wait for response)*

What has happened?

(*Wait for response)*

What number are you calling from?

(*Wait for response)*

Where did the accident happen? How can the ambulance get there?

(*Wait for response)*

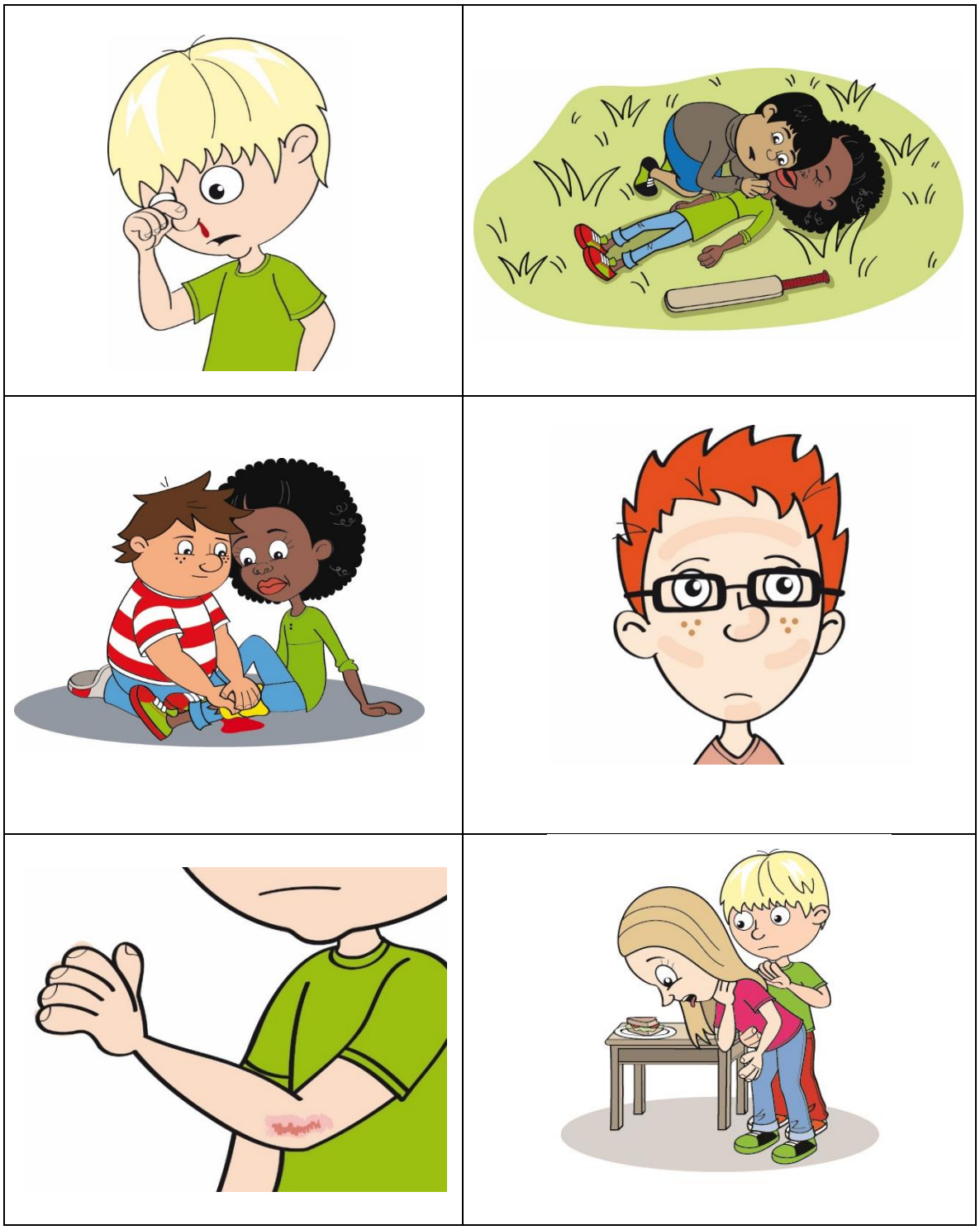
I can hear the sirens, I think the paramedics will take over now, you can put the phone down.

Very well done, you helped someone who was injured and called an ambulance!

(Give the other child card #2 with picture of boy falling down, map and phone number. Repeat questions.)

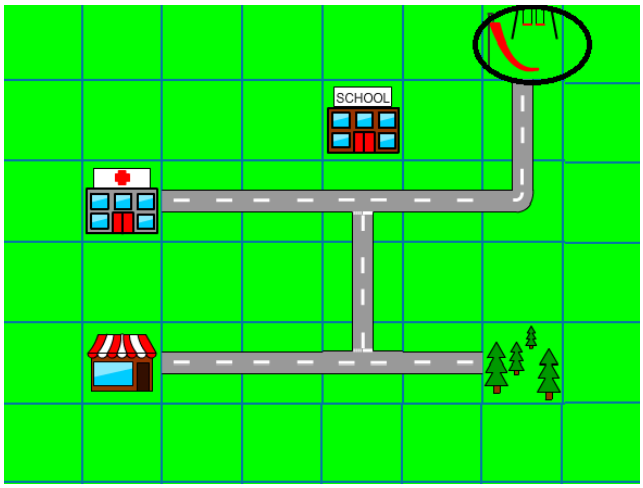
Well done, both of you! Thank you for learning about emergencies with me. Now you know how to help someone who is hurt and how to call an ambulance. Because you did such a good job, I am giving you both a First Aider certificate (*Give children the certificate and take them back to class.)*

Materials for the 'Emergency or not' activity



Materials for the 'Calling 999' activity

CALL 999!



07645372815

Appendix E

First Aid Training Interview Protocol – Developed on the basis of the NICHD protocol

Introduction and rapport building

1 **Hello, my name is _____ and I am from Royal Holloway University. My job is to listen to children and today I will listen to you. I might write down some notes so I can remember what you say. This video recorder will record everything what you say so I need you to talk loud, okay?**

2 **What class did you come from?**

[Wait for child to respond.]

3 **Tell me what you were doing in X class today**.

[Wait for response]

[Use open-ended questions to elicit a narrative, try to talk about the activity for about 2 minutes]

3a Tell me more about that.

[Wait for an answer.]

[Note: Use this question as often as needed throughout this session]

3b Then what happened? / Tell me more about that.

[Wait for an answer.]

[Note: Use this question as often as needed throughout this session]

3c Think back and tell me everything that happened from the beginning to the end.

[Wait for an answer.]

[Note: Use this question as often as needed throughout this session]

3d Tell me more about [person/object/activity mentioned by the child.] / You mentioned [person/object/activity mentioned by the child, tell me everything about that.

[Wait for an answer.]

[Note: Use this question as often as needed throughout this session]

Abbreviated ground rules

4 **Now I just have a few rules to tell you for when we're talking today.**

5 **If I ask you something and you don't understand, please tell me you don't understand.**

6 **And if I make a mistake, please tell me I made a mistake because I don't know what has happened.**

7 **If I ask a question and you don't know the answer, just say, "I don't know."**

8 **But if I ask you something and you do know the answer, you should tell me, ok?**

9 **The last rule is that we tell the truth. Will you promise to tell me the truth when we're talking today?**

[Wait for child to respond.]

10 **Thank you.**

The substantive part of the interview

11 **Now that I know you a little better, I want to talk about the time when you learnt about first aid at school. Do you remember learning about first aid?**

[If yes, skip to question 12.]

[If no, prompt.]

11a Do you remember someone came from Royal Holloway and you learnt how to help someone who is ill or injured?

[Use open-ended questions to elicit more details.]

12 Tell me everything about the time when you learnt about first aid in school.

[Wait for an answer.]

[Note: Use this question as often as needed throughout this session]

4b Then what happened? / Tell me more about that.

[Wait for an answer.]

[Note: Use this question as often as needed throughout this session]

4c Think back and tell me everything that happened from the beginning to the end.

[Wait for an answer.]

[Note: Use this question as often as needed throughout this session]

4d Tell me more about [person/object/activity mentioned by the child.] / You mentioned [person/object/activity mentioned by the child, tell me everything about that.

[Wait for an answer.]

[Note: Use this question as often as needed throughout this session]

Closure

13 Great job, you have told me lots of things today, thank you for helping me. Is there anything else you think I should know?

[Wait for an answer.]

14 Are there any questions you want to ask me?

[Wait for an answer.]

15 What are you going to do today after you leave here?

[Take the child back to class.]

Note: the second interview should be conducted in the same manner as the first one. If the child asks why the interview is repeated, say:

I had to come back and talk to you again about what happened when you learnt about first aid at school. Tell me everything about what happened.

Appendix F

Participant Information Sheet

Dear participant,

We would like to invite you to take part in a study investigating the credibility of child witnesses, designed by researchers at the School of Law, Royal Holloway University of London. Before you take part in this study, it is important for you to understand what it will involve.

Purpose of the study

The study aims to investigate adults' perceptions of the credibility of children alleging sexual abuse at court. Results will be used for the main author's PhD thesis.

Who is invited to take part in this study?

We are recruiting adults above the age of 18 to take part in the study.

Do I have to take part?

It is entirely your decision whether you wish to take part in the study.

What will happen if I take part?

If you decide to take part in the study, you will be asked to imagine you are part of a jury tasked with making a decision about the guilt of the accused in a trial involving a child witness alleging sexually abuse. You will read a fictional transcript of the forensic interview with the child witness and answer a series of questions regarding your perceptions of the case.

What are the risks of taking part?

The study includes a fictional forensic interview with a child witness alleging sexual abuse. For some people, this material might be distressing or triggering.

What are the benefits of taking part?

The results of mock jury studies advance psychologists' and legal professionals' understanding of how people make decisions at court. By taking part in this study, you are contributing to a body of research which may benefit a vulnerable population.

Additionally, to thank participants for their time, we are offering you the option to take part in a prize draw for a £50 Amazon voucher upon completion of the survey.

Will the data collected in this study be kept confidential?

The survey does not include any questions asking for information that could be used to identify you. If you wish to take part in the optional prize draw, you will be asked to enter your email address on a separate webpage. Your email address will not be linked with your responses.

For further information about the study, you can contact Zsofia Szojka (Zsofia.Szojka.2016@rhul.ac.uk) or Dr David La Rooy (David.LaRooy@rhul.ac.uk).

- I confirm that I have read and understand the Participant Information Sheet.
- I confirm that I understand the sensitive nature of the material I will be presented with.
- I agree that data gathered in this study will be kept anonymously and securely and may be used for future research.

If you or someone you know is affected by child sexual abuse, you can contact the NSPCC helpline 24/7 at help@nspcc.org.uk or 0808 800 5000.

Debrief form

Dear Participant,

Thank you for taking part in our research investigating mock jurors' perceptions of the credibility of child witnesses. If you would like to learn more information about this study, please do not hesitate to get in touch via email (Zsofia.Szojka.2016@rhul.ac.uk). Thank you for your time.

Best wishes,

Zsofia Szojka

If you or someone you know is affected by child sexual abuse, you can contact the NSPCC helpline 24/7 at help@nspcc.org.uk or 0808 800 5000.

Appendix G

Demographics questionnaire

1.) How old are you?

   a) Under 18

   b) 18-24

   c) 25-34

   d) 35-44

   e) 45-54

   f) 55-64

   g) 65-74

   h) 75 or older

2.) Do you live in the United Kingdom?

   a) Yes

   b) No

3.) What is your gender?

   a) Male

   b) Female

   c) Prefer not to say

4.) Have you performed jury duty in the past?

   a) Yes

   b) No

5.) Do you have experience working with children?

   a) Yes

   b) No

IF 1.) = a), the survey ends.

Follow up questions:

IF 2.) = a)

6.) What is your citizenship status?

a) UK citizen

b) EU/Commonwealth/Irish citizen

c) Other

7.) Where do you live?

a) England

b) Wales

c) Scotland

d) Northern Ireland

e) Other UK

IF 5.) = a)

8.) For how long have you worked in an occupation which included regular contact with children?

a) Less than a year

b) 1 year

c) 2 years

d) 3 years

e) More than 3 years

9.) What position did you work in when you gained experience working with children?

[Text box]

Appendix H

Condition 1 – single, consistent, young

POLICE	My name is Julie and I am a Police officer. Today is the 13th of January 2019 and it is now 10.35 in the morning. I'm interviewing Alex Franklin, age 8, at Pryport Police station.
POLICE	Alex, I want to talk about why you are here today. Do you know why you are here today?
CHILD	To talk about Paul.
POLICE	Tell me everything about Paul.
CHILD	He's mum's boyfriend.
POLICE	Mum's boyfriend, okay. And why do you want to talk about Paul today?
CHILD	Because he did something bad.
POLICE	Tell me everything about what he did.
CHILD	Something dirty. I don't know how to say it.
POLICE	What do you mean by dirty?
CHILD	I mean something wrong. He touched me.
POLICE	Tell me more about Paul touching you.
CHILD	He said he just wants to give me a hug. But he touched my privates.
POLICE	You say Paul touched your privates. What do you mean by privates?
CHILD	I don't know.
POLICE	Do you know another word for privates?
CHILD	My… my front bum, I think.
POLICE	Tell me what happened from the beginning to the end, when Paul touched your privates.
CHILD	He came into my room and he said he wants to hug me, because I'm so good. He was hugging me but then he was lying on me. And he was hugging me everywhere and he touched my privates… my front bum. And then he got up and went out.
POLICE	How did he touch your privates?
CHILD	I don't understand.
POLICE	Did he touch you over your clothes or under your clothes?
CHILD	He put his hands in my trousers.
POLICE	Okay.
CHILD	But not my underwear.
POLICE	So just to make sure I understand, he touched you under your trousers but over your underwear?
CHILD	Yes.
POLICE	How did you feel when Paul touched your private?
CHILD	I was scared. It's a dirty thing.
POLICE	When did this happen?
CHILD	Last week. But I didn't tell my mum until yesterday, cause I was scared.
POLICE	Who was in the house when Paul touched you?
CHILD	Just me and Paul. Mum was at Granny's, I think Paul picked me up from school.
POLICE	Okay. Has this happened once or more than once?

CHILD	Once. But… he hugged me and kissed me another time.
POLICE	How did he kiss you?
CHILD	He kissed my mouth. He said that's how adults do it.
POLICE	Tell me everything that happened when Paul kissed you on the mouth.
CHILD	It was in my room, he was hugging me on my bed and then he kissed my mouth. And he said I should learn to do it like the adults.
POLICE	When did this happen?
CHILD	A long time ago. Before my birthday, maybe in October.
POLICE	Was anyone in the house when Paul kissed you?
CHILD	I don't remember. We were alone in my room.
POLICE	Okay, Alex, so just to summarise. Paul touched you on your privates once, in your room, when Mum was at Granny's?
CHILD	Yes.
POLICE	And he kissed you on the mouth another time, in your room?
CHILD	Yes.
POLICE	Is there anything else you would like to tell us about, Alex?
CHILD	No. I told you everything. What will happen now? Can I go home?
POLICE	Yes, you can go home with Mum. Thank you for talking to us today.
POLICE	The time is now 11.12 and I am concluding the interview with Alex Franklin.

Condition 2 – single, consistent, old

POLICE	My name is Julie and I am a Police officer. Today is the 13th of January 2019 and it is now 10.35 in the morning. I'm interviewing Alex Franklin, age 12, at Pryport Police station.
POLICE	Alex, I want to talk about why you are here today. Do you know why you are here today?
CHILD	To talk about Paul.
POLICE	Tell me everything about Paul.
CHILD	He's mum's boyfriend.
POLICE	Mum's boyfriend, okay. And why do you want to talk about Paul today?
CHILD	Because he did something bad.
POLICE	Tell me everything about what he did.
CHILD	Something dirty. I don't know how to say it.
POLICE	What do you mean by dirty?
CHILD	I mean something wrong. He touched me.
POLICE	Tell me more about Paul touching you.
CHILD	He said he just wants to give me a hug. But he touched my privates.
POLICE	You say Paul touched your privates. What do you mean by privates?
CHILD	I don't know.
POLICE	Do you know another word for privates?
CHILD	My… my front bum, I think.
POLICE	Tell me what happened from the beginning to the end, when Paul touched your privates.
CHILD	He came into my room and he said he wants to hug me, because I'm so good. He was hugging me but then he was lying on me. And he was hugging me everywhere and he touched my privates… my front bum. And then he got up and went out.
POLICE	How did he touch your privates?
CHILD	I don't understand.
POLICE	Did he touch you over your clothes or under your clothes?
CHILD	He put his hands in my trousers.
POLICE	Okay.
CHILD	But not my underwear.
POLICE	So just to make sure I understand, he touched you under your trousers but over your underwear?
CHILD	Yes.
POLICE	How did you feel when Paul touched your private?
CHILD	I was scared. It's a dirty thing.
POLICE	When did this happen?
CHILD	Last week. But I didn't tell my mum until yesterday, cause I was scared.
POLICE	Who was in the house when Paul touched you?
CHILD	Just me and Paul. Mum was at Granny's, I think Paul picked me up from school.
POLICE	Okay. Has this happened once or more than once?
CHILD	Once. But… he hugged me and kissed me another time.
POLICE	How did he kiss you?
CHILD	He kissed my mouth. He said that's how adults do it.
POLICE	Tell me everything that happened when Paul kissed you on the mouth.

CHILD	It was in my room, he was hugging me on my bed and then he kissed my mouth. And he said I should learn to do it like the adults.
POLICE	When did this happen?
CHILD	A long time ago. Before my birthday, maybe in October.
POLICE	Was anyone in the house when Paul kissed you?
CHILD	I don't remember. We were alone in my room.
POLICE	Okay, Alex, so just to summarise. Paul touched you on your privates once, in your room, when Mum was at Granny's?
CHILD	Yes.
POLICE	And he kissed you on the mouth another time, in your room?
CHILD	Yes.
POLICE	Is there anything else you would like to tell us about, Alex?
CHILD	No. I told you everything. What will happen now? Can I go home?
POLICE	Yes, you can go home with Mum. Thank you for talking to us today.
POLICE	The time is now 11.12 and I am concluding the interview with Alex Franklin.

Condition 3 – Single, inconsistent, young

POLICE	My name is Julie and I am a Police officer. Today is the 13th of January 2019 and it is now 10.35 in the morning. I'm interviewing Alex Franklin, age 8, at Pryport Police station.
POLICE	Alex, I want to talk about why you are here today. Do you know why you are here today?
CHILD	To talk about Paul.
POLICE	Tell me everything about Paul.
CHILD	He's mum's boyfriend.
POLICE	Mum's boyfriend, okay. And why do you want to talk about Paul today?
CHILD	Because he did something bad.
POLICE	Tell me everything about what he did.
CHILD	Something dirty. I don't know how to say it.
POLICE	What do you mean by dirty?
CHILD	I mean something wrong. He touched me.
POLICE	Tell me more about Paul touching you.
CHILD	He said he just wants to give me a hug. But he touched my privates.
POLICE	You say Paul touched your privates. What do you mean by privates?
CHILD	I don't know.
POLICE	Do you know another word for privates?
CHILD	My… my front bum, I think.
POLICE	Tell me what happened from the beginning to the end, when Paul touched your privates.
CHILD	He came into my room and he said he wants to hug me, because I'm so good. He was hugging me but then he was lying on me. And he was hugging me everywhere and he touched my privates… my front bum. And then he got up and went out.
POLICE	How did he touch your privates?
CHILD	I don't understand.
POLICE	Did he touch you over your clothes or under your clothes?
CHILD	Over my clothes.
POLICE	Over your clothes?
CHILD	He put his hands in my trousers.
POLICE	First you said that he touched you over your clothes but then you said he put his hands in your trousers. So, which one is it, under your clothes or over your clothes?
CHILD	Under my clothes.
POLICE	Under your clothes. How did you feel when Paul touched your private?
CHILD	I was scared. It's a dirty thing.
POLICE	When did this happen?
CHILD	Last week. But I didn't tell my mum until yesterday, cause I was scared.
POLICE	Who was in the house when Paul touched you?
CHILD	Just me and Paul. Mum was at Granny's, I think Paul picked me up from school.
POLICE	Okay. Has this happened once or more than once?
CHILD	Once. But… he hugged me and kissed me another time.
POLICE	How did he kiss you?
CHILD	He kissed my mouth. He said that's how adults do it.

POLICE	Tell me everything that happened when Paul kissed you on the mouth.
CHILD	It was in my room, he was hugging me on my bed and then he kissed my mouth. He said I should learn to do it like the adults.
POLICE	When did this happen?
CHILD	A long time ago. Before my birthday, maybe in October.
POLICE	Was anyone in the house when Paul kissed you?
CHILD	I don't remember. We were alone in my room.
POLICE	Okay, Alex, so just to summarise. Paul touched you on your privates once, in your room, when Mum was at Granny's?
CHILD	Yes.
POLICE	And he kissed you on the mouth another time, in your room?
CHILD	No, that time was in the living room. We were watching TV and he kissed me.
POLICE	Earlier, you said that you were in your room when he kissed you. Was it in your room or in the living room?
CHILD	It was in the living room, I remember now.
POLICE	Is there anything else you would like to tell us about, Alex?
CHILD	No. I told you everything. What will happen now? Can I go home?
POLICE	Yes, you can go home with Mum. Thank you for talking to us today.
POLICE	The time is now 11.12 and I am concluding the interview with Alex Franklin.

Condition 4 – Single, inconsistent, old

POLICE	My name is Julie and I am a Police officer. Today is the 13th of January 2019 and it is now 10.35 in the morning. I'm interviewing Alex Franklin, age 12, at Pryport Police station.
POLICE	Alex, I want to talk about why you are here today. Do you know why you are here today?
CHILD	To talk about Paul.
POLICE	Tell me everything about Paul.
CHILD	He's mum's boyfriend.
POLICE	Mum's boyfriend, okay. And why do you want to talk about Paul today?
CHILD	Because he did something bad.
POLICE	Tell me everything about what he did.
CHILD	Something dirty. I don't know how to say it.
POLICE	What do you mean by dirty?
CHILD	I mean something wrong. He touched me.
POLICE	Tell me more about Paul touching you.
CHILD	He said he just wants to give me a hug. But he touched my privates.
POLICE	You say Paul touched your privates. What do you mean by privates?
CHILD	I don't know.
POLICE	Do you know another word for privates?
CHILD	My… my front bum, I think.
POLICE	Tell me what happened from the beginning to the end, when Paul touched your privates.
CHILD	He came into my room and he said he wants to hug me, because I'm so good. He was hugging me but then he was lying on me. And he was hugging me everywhere and he touched my privates… my front bum. And then he got up and went out.
POLICE	How did he touch your privates?
CHILD	I don't understand.
POLICE	Did he touch you over your clothes or under your clothes?
CHILD	Over my clothes.
POLICE	Over your clothes?
CHILD	He put his hands in my trousers.
POLICE	First you said that he touched you over your clothes but then you said he put his hands in your trousers. So, which one is it, under your clothes or over your clothes?
CHILD	Under my clothes.
POLICE	Under your clothes. How did you feel when Paul touched your private?
CHILD	I was scared. It's a dirty thing.
POLICE	When did this happen?
CHILD	Last week. But I didn't tell my mum until yesterday, cause I was scared.
POLICE	Who was in the house when Paul touched you?
CHILD	Just me and Paul. Mum was at Granny's, I think Paul picked me up from school.
POLICE	Okay. Has this happened once or more than once?
CHILD	Once. But… he hugged me and kissed me another time.
POLICE	How did he kiss you?
CHILD	He kissed my mouth. He said that's how adults do it.

POLICE	Tell me everything that happened when Paul kissed you on the mouth.
CHILD	It was in my room, he was hugging me on my bed and then he kissed my mouth. He said I should learn to do it like the adults.
POLICE	When did this happen?
CHILD	A long time ago. Before my birthday, maybe in October.
POLICE	Was anyone in the house when Paul kissed you?
CHILD	I don't remember. We were alone in my room.
POLICE	Okay, Alex, so just to summarise. Paul touched you on your privates once, in your room, when Mum was at Granny's?
CHILD	Yes.
POLICE	And he kissed you on the mouth another time, in your room?
CHILD	No, that time was in the living room. We were watching TV and he kissed me.
POLICE	Earlier, you said that you were in your room when he kissed you. Was it in your room or in the living room?
CHILD	It was in the living room, I remember now.
POLICE	Is there anything else you would like to tell us about, Alex?
CHILD	No. I told you everything. What will happen now? Can I go home?
POLICE	Yes, you can go home with Mum. Thank you for talking to us today.
POLICE	The time is now 11.12 and I am concluding the interview with Alex Franklin.

Condition 5 – Repeated, consistent, young

Interview 1

POLICE	My name is Julie and I am a Police officer. Today is the 13th of January 2019 and it is now 10.35 in the morning. I'm interviewing Alex Franklin, age 8, at Pryport Police station.
POLICE	Alex, I want to talk about why you are here today. Do you know why you are here today?
CHILD	To talk about Paul.
POLICE	Tell me everything about Paul.
CHILD	He's mum's boyfriend.
POLICE	Mum's boyfriend, okay. And why do you want to talk about Paul today?
CHILD	Because he did something bad.
POLICE	Tell me everything about what he did.
CHILD	Something dirty. I don't know how to say it.
POLICE	What do you mean by dirty?
CHILD	I mean something wrong. He touched me.
POLICE	Tell me more about Paul touching you.
CHILD	He said he just wants to give me a hug. But he touched my privates.
POLICE	You say Paul touched your privates. What do you mean by privates?
CHILD	I don't know.
POLICE	Do you know another word for privates?
CHILD	My… my front bum, I think.
POLICE	Tell me what happened from the beginning to the end, when Paul touched your privates.
CHILD	He came into my room and he said he wants to hug me, because I'm so good. He was hugging me but then he was lying on me. And he was hugging me everywhere and he touched my privates… my front bum. And then he got up and went out.
POLICE	How did he touch your privates?
CHILD	I don't understand.
POLICE	Did he touch you over your clothes or under your clothes?
CHILD	He put his hands in my trousers.
POLICE	Okay.
CHILD	But not my underwear.
POLICE	So just to make sure I understand, he touched you under your trousers but over your underwear?
CHILD	Yes.
POLICE	How did you feel when Paul touched your private?
CHILD	I was scared. It's a dirty thing.
POLICE	When did this happen?
CHILD	Last week. But I didn't tell my mum until yesterday, cause I was scared.
POLICE	Who was in the house when Paul touched you?
CHILD	Just me and Paul. Mum was at Granny's, I think Paul picked me up from school.
POLICE	Okay. Has this happened once or more than once?
CHILD	Once. But… he hugged me and kissed me another time.
POLICE	How did he kiss you?

CHILD	He kissed my mouth. He said that's how adults do it.
POLICE	Tell me everything that happened when Paul kissed you on the mouth.
CHILD	It was in my room, he was hugging me on my bed and then he kissed my mouth. And he said I should learn to do it like the adults.
POLICE	When did this happen?
CHILD	A long time ago. Before my birthday, maybe in October.
POLICE	Was anyone in the house when Paul kissed you?
CHILD	I don't remember. We were alone in my room.
POLICE	Okay, Alex, so just to summarise. Paul touched you on your privates once, in your room, when Mum was at Granny's?
CHILD	Yes.
POLICE	And he kissed you on the mouth another time, in your room?
CHILD	Yes.
POLICE	Is there anything else you would like to tell us about, Alex?
CHILD	No. I told you everything. What will happen now? Can I go home?
POLICE	Yes, you can go home with Mum. Thank you for talking to us today.
POLICE	The time is now 11.12 and I am concluding the interview with Alex Franklin.

Interview 2

POLICE	My name is Julie Brown, police officer. Today is the 2nd of April 2019 and it is now 12.10. This is the second interview with Alex Franklin, age 8, at Pryport Police station.
POLICE	Alex, I would like to talk to you again about what you told me last time. Do you remember what we talked about?
CHILD	About Paul.
POLICE	Why were we talking about Paul?
CHILD	Because he touched my front bum and he kissed me.
POLICE	Can you tell me everything that happened when Paul touched your front bum?
CHILD	I told you last time. It was a long time ago now, he doesn't live with us anymore, because I told mum about what he did.
POLICE	I know we talked about this already, but could you tell me what he did again?
CHILD	He was pretending to hug me, but he was really touching my front bum.
POLICE	Tell me everything that happened when he touched your front bum from the beginning to the end.
CHILD	He picked me up from school that day, cause mum was at granny's. We went into my room and he said he was going to give me a hug, and he lay on me and started touching me. First, he was just touching my back and my legs and then he put his hands in my trousers.
POLICE	Okay, so you said he touched you under your clothes?
CHILD	Yes, under my trousers. But over my underwear.
POLICE	And what happened after that?
CHILD	He moved his hand there for a bit. I was trying to wiggle out of the hug and then he got up and left.
POLICE	What happened after that?
CHILD	Nothing. I wanted to tell mum when she came home, but I was too scared she wouldn't believe me and I was scared of him too.
POLICE	So you did not tell your mum?

CHILD	Not that day, but later.
POLICE	What did you tell your mum?
CHILD	That Paul touched my privates. She was crying and then we came here for the first time.
POLICE	And did he touch you just that one time?
CHILD	Yes.
POLICE	What else happened?
CHILD	He kissed me like adults do.
POLICE	What do you mean by ‘like adults do’?
CHILD	On my mouth.
POLICE	Tell me everything about the time when he kissed you on your mouth.
CHILD	He came into my room to talk to me. He was sitting on my bed and we were talking, I can’t remember what he said but then he hugged me and started kissing me. Then he left.
POLICE	How did you feel when this happened?
CHILD	I don’t know. He was nice to me, he hugged me all the time. But I knew he shouldn’t kiss me like that.
POLICE	Who was at home when this happened?
CHILD	Just me and him. My mum was out somewhere, I think.
POLICE	Is there anything else you would like to tell me about?
CHILD	I don’t think so.
POLICE	Do you have any questions for me?
CHILD	Will I have to come back again?
POLICE	I don’t know. If we do need to talk to you again, we will call your mum. Is that okay?
CHILD	Yes. Am I going to school now?
POLICE	You’ll have to ask your mum.
POLICE	The time is now 12.40 and I am concluding this interview.

Condition 6 – Repeated, consistent, old

Interview 1

POLICE	My name is Julie and I am a Police officer. Today is the 13th of January 2019 and it is now 10.35 in the morning. I'm interviewing Alex Franklin, age 12, at Pryport Police station.
POLICE	Alex, I want to talk about why you are here today. Do you know why you are here today?
CHILD	To talk about Paul.
POLICE	Tell me everything about Paul.
CHILD	He's mum's boyfriend.
POLICE	Mum's boyfriend, okay. And why do you want to talk about Paul today?
CHILD	Because he did something bad.
POLICE	Tell me everything about what he did.
CHILD	Something dirty. I don't know how to say it.
POLICE	What do you mean by dirty?
CHILD	I mean something wrong. He touched me.
POLICE	Tell me more about Paul touching you.
CHILD	He said he just wants to give me a hug. But he touched my privates.
POLICE	You say Paul touched your privates. What do you mean by privates?
CHILD	I don't know.
POLICE	Do you know another word for privates?
CHILD	My… my front bum, I think.
POLICE	Tell me what happened from the beginning to the end, when Paul touched your privates.
CHILD	He came into my room and he said he wants to hug me, because I'm so good. He was hugging me but then he was lying on me. And he was hugging me everywhere and he touched my privates… my front bum. And then he got up and went out.
POLICE	How did he touch your privates?
CHILD	I don't understand.
POLICE	Did he touch you over your clothes or under your clothes?
CHILD	He put his hands in my trousers.
POLICE	Okay.
CHILD	But not my underwear.
POLICE	So just to make sure I understand, he touched you under your trousers but over your underwear?
CHILD	Yes.
POLICE	How did you feel when Paul touched your private?
CHILD	I was scared. It's a dirty thing.
POLICE	When did this happen?
CHILD	Last week. But I didn't tell my mum until yesterday, cause I was scared.
POLICE	Who was in the house when Paul touched you?
CHILD	Just me and Paul. Mum was at Granny's, I think Paul picked me up from school.
POLICE	Okay. Has this happened once or more than once?
CHILD	Once. But… he hugged me and kissed me another time.
POLICE	How did he kiss you?

CHILD	He kissed my mouth. He said that's how adults do it.
POLICE	Tell me everything that happened when Paul kissed you on the mouth.
CHILD	It was in my room, he was hugging me on my bed and then he kissed my mouth. And he said I should learn to do it like the adults.
POLICE	When did this happen?
CHILD	A long time ago. Before my birthday, maybe in October.
POLICE	Was anyone in the house when Paul kissed you?
CHILD	I don't remember. We were alone in my room.
POLICE	Okay, Alex, so just to summarise. Paul touched you on your privates once, in your room, when Mum was at Granny's?
CHILD	Yes.
POLICE	And he kissed you on the mouth another time, in your room?
CHILD	Yes.
POLICE	Is there anything else you would like to tell us about, Alex?
CHILD	No. I told you everything. What will happen now? Can I go home?
POLICE	Yes, you can go home with Mum. Thank you for talking to us today.
POLICE	The time is now 11.12 and I am concluding the interview with Alex Franklin.

Interview 2

POLICE	My name is Julie Brown, police officer. Today is the 2nd of April 2019 and it is now 12.10. This is the second interview with Alex Franklin, age 12, at Pryport Police station.
POLICE	Alex, I would like to talk to you again about what you told me last time. Do you remember what we talked about?
CHILD	About Paul.
POLICE	Why were we talking about Paul?
CHILD	Because he touched my front bum and he kissed me.
POLICE	Can you tell me everything that happened when Paul touched your front bum?
CHILD	I told you last time. It was a long time ago now, he doesn't live with us anymore, because I told mum about what he did.
POLICE	I know we talked about this already, but could you tell me what he did again?
CHILD	He was pretending to hug me, but he was really touching my front bum.
POLICE	Tell me everything that happened when he touched your front bum from the beginning to the end.
CHILD	He picked me up from school that day, cause mum was at granny's. We went into my room and he said he was going to give me a hug, and he lay on me and started touching me. First, he was just touching my back and my legs and then he put his hands in my trousers.
POLICE	Okay, so you said he touched you under your clothes?
CHILD	Yes, under my trousers. But over my underwear.
POLICE	And what happened after that?
CHILD	He moved his hand there for a bit. I was trying to wiggle out of the hug and then he got up and left.
POLICE	What happened after that?
CHILD	Nothing. I wanted to tell mum when she came home, but I was too scared she wouldn't believe me and I was scared of him too.
POLICE	So you did not tell your mum?

CHILD	Not that day, but later.
POLICE	What did you tell your mum?
CHILD	That Paul touched my privates. She was crying and then we came here for the first time.
POLICE	And did he touch you just that one time?
CHILD	Yes.
POLICE	What else happened?
CHILD	He kissed me like adults do.
POLICE	What do you mean by 'like adults do'?
CHILD	On my mouth.
POLICE	Tell me everything about the time when he kissed you on your mouth.
CHILD	He came into my room to talk to me. He was sitting on my bed and we were talking, I can't remember what he said but then he hugged me and started kissing me. Then he left.
POLICE	How did you feel when this happened?
CHILD	I don't know. He was nice to me, he hugged me all the time. But I knew he shouldn't kiss me like that.
POLICE	Who was at home when this happened?
CHILD	Just me and him. My mum was out somewhere, I think.
POLICE	Is there anything else you would like to tell me about?
CHILD	I don't think so.
POLICE	Do you have any questions for me?
CHILD	Will I have to come back again?
POLICE	I don't know. If we do need to talk to you again, we will call your mum. Is that okay?
CHILD	Yes. Am I going to school now?
POLICE	You'll have to ask your mum.
POLICE	The time is now 12.40 and I am concluding this interview.

Condition 7 – Repeated, inconsistent, young

Interview 1

POLICE	My name is Julie and I am a Police officer. Today is the 13th of January 2019 and it is now 10.35 in the morning. I'm interviewing Alex Franklin, age 8, at Pryport Police station.
POLICE	Alex, I want to talk about why you are here today. Do you know why you are here today?
CHILD	To talk about Paul.
POLICE	Tell me everything about Paul.
CHILD	He's mum's boyfriend.
POLICE	Mum's boyfriend, okay. And why do you want to talk about Paul today?
CHILD	Because he did something bad.
POLICE	Tell me everything about what he did.
CHILD	Something dirty. I don't know how to say it.
POLICE	What do you mean by dirty?
CHILD	I mean something wrong. He touched me.
POLICE	Tell me more about Paul touching you.
CHILD	He said he just wants to give me a hug. But he touched my privates.
POLICE	You say Paul touched your privates. What do you mean by privates?
CHILD	I don't know.
POLICE	Do you know another word for privates?
CHILD	My… my front bum, I think.
POLICE	Tell me what happened from the beginning to the end, when Paul touched your privates.
CHILD	He came into my room and he said he wants to hug me, because I'm so good. He was hugging me but then he was lying on me. And he was hugging me everywhere and he touched my privates… my front bum. And then he got up and went out.
POLICE	How did he touch your privates?
CHILD	I don't understand.
POLICE	Did he touch you over your clothes or under your clothes?
CHILD	Over my clothes.
POLICE	Over your clothes?
CHILD	He put his hands in my trousers.
POLICE	First you said that he touched you over your clothes but then you said he put his hands in your trousers. So, which one is it, under your clothes or over your clothes?
CHILD	Under my clothes.
POLICE	Under your clothes. How did you feel when Paul touched your private?
CHILD	I was scared. It's a dirty thing.
POLICE	When did this happen?
CHILD	Last week. But I didn't tell my mum until yesterday, cause I was scared.
POLICE	Who was in the house when Paul touched you?
CHILD	Just me and Paul. Mum was at Granny's, I think Paul picked me up from school.
POLICE	Okay. Has this happened once or more than once?
CHILD	Once. But… he hugged me and kissed me another time.

POLICE	How did he kiss you?
CHILD	He kissed my mouth. He said that's how adults do it.
POLICE	Tell me everything that happened when Paul kissed you on the mouth.
CHILD	It was in my room, he was hugging me on my bed and then he kissed my mouth. He said I should learn to do it like the adults.
POLICE	When did this happen?
CHILD	A long time ago. Before my birthday, maybe in October.
POLICE	Was anyone in the house when Paul kissed you?
CHILD	I don't remember. We were alone in my room.
POLICE	Okay, Alex, so just to summarise. Paul touched you on your privates once, in your room, when Mum was at Granny's?
CHILD	Yes.
POLICE	And he kissed you on the mouth another time, in your room?
CHILD	Yes.
POLICE	Is there anything else you would like to tell us about, Alex?
CHILD	No. I told you everything. What will happen now? Can I go home?
POLICE	Yes, you can go home with Mum. Thank you for talking to us today.
POLICE	The time is now 11.12 and I am concluding the interview with Alex Franklin.

Interview 2

POLICE	My name is Julie Brown, police officer. Today is the 2nd of April 2019 and it is now 12.10. This is the second interview with Alex Franklin, age 8, at Pryport Police station.
POLICE	Alex, I would like to talk to you again about what you told me last time. Do you remember what we talked about?
CHILD	About Paul.
POLICE	Why were we talking about Paul?
CHILD	Because he touched my front bum and he kissed me.
POLICE	Can you tell me everything that happened when Paul touched your front bum?
CHILD	I told you last time. It was a long time ago now, he doesn't live with us anymore, because I told mum about what he did.
POLICE	I know we talked about this already, but could you tell me what he did again?
CHILD	He was pretending to hug me, but he was really touching my front bum.
POLICE	Tell me everything that happened when he touched your front bum from the beginning to the end.
CHILD	He picked me up from school that day, cause mum was at granny's. We went into my room and he said he was going to give me a hug, and he lay on me and started touching me. First, he was just touching my back and my legs and then he put his hands in my trousers.
POLICE	Okay, so you said he touched you under your clothes?
CHILD	Yes, under my trousers. But over my underwear.
POLICE	And what happened after that?
CHILD	He moved his hand there for a bit. I was trying to wiggle out of the hug and then he got up and left.
POLICE	What happened after that?
CHILD	Nothing. I wanted to tell mum when she came home, but I was too scared she wouldn't believe me and I was scared of him too.

POLICE	So you did not tell your mum?
CHILD	Not that day, but later.
POLICE	What did you tell your mum?
CHILD	That Paul touched my privates. She was crying and then we came here for the first time.
POLICE	And did he touch you just that one time?
CHILD	Yes.
POLICE	What else happened?
CHILD	He kissed me like adults do.
POLICE	What do you mean by ‘like adults do’?
CHILD	On my mouth.
POLICE	Tell me everything about the time when he kissed you on your mouth.
CHILD	It was in the living room. We were watching TV and he kissed me. I can’t remember what he said but then he hugged me and started kissing me. Then he left.
POLICE	Last time you said that he kissed you in your room. But now you’re saying it happened in the living room, so which one was it, your room or the living room?
CHILD	It was in the living room, I remember now.
POLICE	Okay. How did you feel when this happened?
CHILD	I don’t know. He was nice to me, he hugged me all the time. But I knew he shouldn’t kiss me like that.
POLICE	Who was at home when this happened?
CHILD	Just me and him. My mum was out somewhere, I think.
POLICE	Is there anything else you would like to tell me about?
CHILD	I don’t think so.
POLICE	Do you have any questions for me?
CHILD	Will I have to come back again?
POLICE	I don’t know. If we do need to talk to you again, we will call your mum. Is that okay?
CHILD	Yes. Am I going to school now?
POLICE	You’ll have to ask your mum.
POLICE	The time is now 12.40 and I am concluding this interview.

Condition 8 – Repeated, inconsistent, old

Interview 1

POLICE	My name is Julie and I am a Police officer. Today is the 13th of January 2019 and it is now 10.35 in the morning. I'm interviewing Alex Franklin, age 12, at Pryport Police station.
POLICE	Alex, I want to talk about why you are here today. Do you know why you are here today?
CHILD	To talk about Paul.
POLICE	Tell me everything about Paul.
CHILD	He's mum's boyfriend.
POLICE	Mum's boyfriend, okay. And why do you want to talk about Paul today?
CHILD	Because he did something bad.
POLICE	Tell me everything about what he did.
CHILD	Something dirty. I don't know how to say it.
POLICE	What do you mean by dirty?
CHILD	I mean something wrong. He touched me.
POLICE	Tell me more about Paul touching you.
CHILD	He said he just wants to give me a hug. But he touched my privates.
POLICE	You say Paul touched your privates. What do you mean by privates?
CHILD	I don't know.
POLICE	Do you know another word for privates?
CHILD	My… my front bum, I think.
POLICE	Tell me what happened from the beginning to the end, when Paul touched your privates.
CHILD	He came into my room and he said he wants to hug me, because I'm so good. He was hugging me but then he was lying on me. And he was hugging me everywhere and he touched my privates… my front bum. And then he got up and went out.
POLICE	How did he touch your privates?
CHILD	I don't understand.
POLICE	Did he touch you over your clothes or under your clothes?
CHILD	Over my clothes.
POLICE	Over your clothes?
CHILD	He put his hands in my trousers.
POLICE	First you said that he touched you over your clothes but then you said he put his hands in your trousers. So, which one is it, under your clothes or over your clothes?
CHILD	Under my clothes.
POLICE	Under your clothes. How did you feel when Paul touched your private?
CHILD	I was scared. It's a dirty thing.
POLICE	When did this happen?
CHILD	Last week. But I didn't tell my mum until yesterday, cause I was scared.
POLICE	Who was in the house when Paul touched you?
CHILD	Just me and Paul. Mum was at Granny's, I think Paul picked me up from school.
POLICE	Okay. Has this happened once or more than once?
CHILD	Once. But… he hugged me and kissed me another time.

POLICE	How did he kiss you?
CHILD	He kissed my mouth. He said that's how adults do it.
POLICE	Tell me everything that happened when Paul kissed you on the mouth.
CHILD	It was in my room, he was hugging me on my bed and then he kissed my mouth. He said I should learn to do it like the adults.
POLICE	When did this happen?
CHILD	A long time ago. Before my birthday, maybe in October.
POLICE	Was anyone in the house when Paul kissed you?
CHILD	I don't remember. We were alone in my room.
POLICE	Okay, Alex, so just to summarise. Paul touched you on your privates once, in your room, when Mum was at Granny's?
CHILD	Yes.
POLICE	And he kissed you on the mouth another time, in your room?
CHILD	Yes.
POLICE	Is there anything else you would like to tell us about, Alex?
CHILD	No. I told you everything. What will happen now? Can I go home?
POLICE	Yes, you can go home with Mum. Thank you for talking to us today.
POLICE	The time is now 11.12 and I am concluding the interview with Alex Franklin.

Interview 2

POLICE	My name is Julie Brown, police officer. Today is the 2nd of April 2019 and it is now 12.10. This is the second interview with Alex Franklin, age 12, at Pryport Police station.
POLICE	Alex, I would like to talk to you again about what you told me last time. Do you remember what we talked about?
CHILD	About Paul.
POLICE	Why were we talking about Paul?
CHILD	Because he touched my front bum and he kissed me.
POLICE	Can you tell me everything that happened when Paul touched your front bum?
CHILD	I told you last time. It was a long time ago now, he doesn't live with us anymore, because I told mum about what he did.
POLICE	I know we talked about this already, but could you tell me what he did again?
CHILD	He was pretending to hug me, but he was really touching my front bum.
POLICE	Tell me everything that happened when he touched your front bum from the beginning to the end.
CHILD	He picked me up from school that day, cause mum was at granny's. We went into my room and he said he was going to give me a hug, and he lay on me and started touching me. First, he was just touching my back and my legs and then he put his hands in my trousers.
POLICE	Okay, so you said he touched you under your clothes?
CHILD	Yes, under my trousers. But over my underwear.
POLICE	And what happened after that?
CHILD	He moved his hand there for a bit. I was trying to wiggle out of the hug and then he got up and left.
POLICE	What happened after that?
CHILD	Nothing. I wanted to tell mum when she came home, but I was too scared she wouldn't believe me and I was scared of him too.

POLICE	So you did not tell your mum?
CHILD	Not that day, but later.
POLICE	What did you tell your mum?
CHILD	That Paul touched my privates. She was crying and then we came here for the first time.
POLICE	And did he touch you just that one time?
CHILD	Yes.
POLICE	What else happened?
CHILD	He kissed me like adults do.
POLICE	What do you mean by ‘like adults do’?
CHILD	On my mouth.
POLICE	Tell me everything about the time when he kissed you on your mouth.
CHILD	It was in the living room. We were watching TV and he kissed me. I can’t remember what he said but then he hugged me and started kissing me. Then he left.
POLICE	Last time you said that he kissed you in your room. But now you’re saying it happened in the living room, so which one was it, your room or the living room?
CHILD	It was in the living room, I remember now.
POLICE	Okay. How did you feel when this happened?
CHILD	I don’t know. He was nice to me, he hugged me all the time. But I knew he shouldn’t kiss me like that.
POLICE	Who was at home when this happened?
CHILD	Just me and him. My mum was out somewhere, I think.
POLICE	Is there anything else you would like to tell me about?
CHILD	I don’t think so.
POLICE	Do you have any questions for me?
CHILD	Will I have to come back again?
POLICE	I don’t know. If we do need to talk to you again, we will call your mum. Is that okay?
CHILD	Yes. Am I going to school now?
POLICE	You’ll have to ask your mum.
POLICE	The time is now 12.40 and I am concluding this interview.

Appendix I

Instructions – Single interview

Please read the following text carefully. You will read a fictional interview scenario developed on the basis of real forensic investigations with children alleging physical abuse. After reading the interview, you will be asked to answer a series of questions regarding your impressions about the case. You will be able to return to the interview if you need to re-read it when answering the questions.

Questions

1. On the basis of the interview only, without any knowledge of any other case details, how reliable does this witness appear to you?

(sliding scale)

2. Rate the following factors in terms of relevance when judging the reliability of the witness.

A. Child's verbal competency
B. Amount of detail in the child's statement
C. Consistency of child's statements
D. Interviewing style
E. Logical presentation of story elements from start to finish
F. Child's age
G. Plausibility of the situation described by the child

3. On the basis of the interview only, without knowledge of any other case details, how competent does this witness appear to you?

(sliding scale)

4. Rate the following factors in terms of relevance when judging the competence of the witness.

A. Child's verbal competency
B. Amount of detail in the child's statement
C. Consistency of child's statements
D. Interviewing style
E. Logical presentation of story elements from start to finish
F. Child's age
G. Plausibility of the situation described by the child

5. On the basis of the interview only, without knowledge of any other case details, how believable does the statement of this witness appear to you?

(sliding scale)

6. Rate the following factors in terms of relevance when judging the believability of the witness statement.

A. Child's verbal competency
B. Amount of detail in the child's statement
C. Consistency of child's statements
D. Interviewing style
E. Logical presentation of story elements from start to finish
F. Child's age
G. Plausibility of the situation described by the child

7. Imagine you are part of a jury situation where you have to determine the defendant's guilt. The defendant denies the allegation of sexual abuse. No medical or physical evidence is available. On the basis of the witness' statement, what verdict would you choose?

Convict / acquit

8. How confident are you in your choice of verdict?

(sliding scale)

9. Rate the following factors in terms of relevance when determining the defendant's guilt.

A. Child's verbal competency
B. Amount of detail in the child's statement
C. Consistency of child's statements
D. Interviewing style
E. Logical presentation of story elements from start to finish
F. Child's age
G. Plausibility of the situation described by the child
H. Defendant's statement
I. Lack of medical evidence

Instructions –multiple interviews

Please read the following text carefully. You will read a fictional scenario involving two interviews developed on the basis of real forensic investigations with children alleging physical abuse. After reading the interviews, you will be asked to answer a series of questions regarding your impressions about the case. You will be able to return to the interviews if you need to re-read them when answering the questions.

Questions

1. On the basis of the interviews only, without any knowledge of any other case details, how reliable does this witness appear to you?

(sliding scale)

2. Rate the following factors in terms of relevance when judging the reliability of the witness.

A. Child's verbal competency
B. Amount of detail in the child's statement
C. Consistency of child's statements
D. Interviewing style
E. Logical presentation of story elements from start to finish
F. Child's age
G. Plausibility of the situation described by the child

3. On the basis of the interview only, without knowledge of any other case details, how competent does this witness appear to you?

(sliding scale)

4. Rank the following factors in terms of relevance when judging the competence of the witness.

A. Child's verbal competency
B. Amount of detail in the child's statement
C. Consistency of child's statements
D. Interviewing style
E. Logical presentation of story elements from start to finish
F. Child's age
G. Plausibility of the situation described by the child

5. On the basis of the interview only, without knowledge of any other case details, how believable do the statements of this witness appear to you?

(sliding scale)

6. Rate the following factors in order of relevance when judging the believability of the statements of the witness.

A. Child's verbal competency
B. Amount of detail in the child's statement
C. Consistency of child's statements
D. Interviewing style
E. Logical presentation of story elements from start to finish
F. Child's age
G. Plausibility of the situation described by the child

7. Imagine you are part of a jury situation where you have to determine the defendant's guilt. The defendant denies the allegation of sexual abuse. No medical or physical evidence is available. On the basis of the witness' statement, what verdict would you choose?

Convict/Acquit

8. How confident are you in your choice of verdict?

(sliding scale)

9. Rate the following factors in terms of relevance when determining the defendant's guilt.

A. Child's verbal competency
B. Amount of detail in the child's statement
C. Consistency of child's statements

D. Interviewing style

E. Logical presentation of story elements from start to finish

F. Child's age

G. Plausibility of the situation described by the child

H. Defendant's statement

I. Lack of medical evidence