Perception versus reality: Investigating the impact of talkativeness on children's credibility and reliability.

By

Helen C. Pierce (née Riley)

A thesis submitted to Victoria University of Wellington in fulfilment of the requirements of the degree of Master of Science in Psychology

Victoria University of Wellington

2019

Abstract

Does how much children say predict how credible they are as a witness? Children's talkativeness can be easily observed by jurors, but we know very little about how it affects judgements of children's credibility. The present research investigates the effect of talkativeness on juror perceptions and children's actual testimony. In Study 1 participants rated six transcripts from low/high talkative 5-, 8-, or 12-year old children. Results showed that mock jurors rated high-talkative children more favourably than low-talkative children and older children were rated more favourably than younger children. In Study 2 we analysed transcripts of memory interviews from 5-, 8-, and 12-year-old children. Talkativeness was not associated with accuracy, but child age was. Talkativeness and child age were both associated with the amount of information recalled. This research shows that talkativeness of child witnesses not only influences juror perceptions but also is an indication of the amount of information that children recall in a memory interview. It is not just what a child says, but also how they say it that matters.

Keywords: juror perception, child witness, witness credibility, talkative

Acknowledgements

"The more that you read, the more things you will know. The more that you learn, the more places you'll go." - Dr Seuss, I Can Read With My Eyes Shut

I would like to thank my supervisor Dr Deirdre Brown for her tireless support, feedback, guidance, and wisdom. I would like to especially thank Deirdre for sharing her energy and passion for research, and teaching me so many different skills along the way. As Deirdre has recently reminded me "research is a team sport" and so I would also like to thank the other Victoria University of Wellington staff, students, family, and friends that have supported me throughout this entire process.

Thank you to Dr Matt Crawford for your assistance throughout Study 1, especially for helping me think outside the square. Thank you to Dr Matt Hammond, Laina Isler, and Johannes Karl for your statistical advice and reassurance. Thank you to Dr Corrine Seals and Dr Frank Boers for your guidance and recommendations for the linguistic aspects of my research. Thank you to all the applied developmental lab members, past and present, from the Applied Developmental Lab who have listened to me and supported me throughout all stages of this research.

And last, but not at all least, thank you to my family and friends. Sean, you have been a tremendous support—I really cannot thank you enough. To my children, thank you for the weird and wonderful ways that you have supported me in this endeavour. To the rest of my family, thank you for your encouragement—especially when things got tough. To my friends, thanks for listening to me and reminding me to take breaks.

Table of Contents

List of Tables

List of Figures

List of Appendices

Introduction

Some children are just more talkative than others, but does this change how we think about them? Is talkativeness something we need to take notice of if a child is testifying about an event they have experienced or witnessed? Due to the nature of child sexual abuse, the alleged victim's testimony is often the only source of evidence. There tends to be no other witnesses, and physical evidence is often unavailable. This means that a child's testimony, and how credibly their testimony is perceived, will likely be critical to the outcome of the case. If juror perceptions of child witnesses are misguided, then miscarriages of justice are likely.

Talkativeness

Talkativeness—the number of spoken words—is a form of verbal behaviour potentially detectable by jurors and may influence judgements about witness credibility, but we do not know *how* it might influence jurors' judgements. Over 45 years ago, we saw evidence of talkativeness influencing how people evaluate others. Hayes and Meltzer (1972) found that men who spoke more, during a discussion, received more favourable evaluations than men who spoke less in the discussion. Furthermore, the men who spoke more still received more favourable evaluations even though the participants only knew the men were talking because a light would turn on to indicate the speakers turn in the conversation. So even when the content of the discussion was eliminated, talkativeness still had an effect. Would this effect (of more favourable ratings of people who speak more) transfer to children who speak more? There is certainly evidence to support this, for instance high-talkative school children were rated as more likely to be intelligent and have better academic performance than low-talkative children (Evans, 1996). If talkativeness can influence teachers' perceptions, then it seems plausible that talkativeness could also influence juror perceptions of a child witness providing testimony.

Differences in speech production emerge as early as the toddler years (Fenson et al., 1994; Pan, Rowe, Singer, & Snow, 2005). While some children may say more than others, this does not necessarily mean that they are providing more information. An example of this would be two children talking about getting their temperature taken. The first child says, "The nurse put a thermometer in my ear and it went beep" and the second child might say, "The nurse got the temperature checker thing and put it inside my ear and then the temperature checker thing made a funny noise like a loud bleep kind of noise". The first child in the above example uses 12 words and the second child uses 30 words—nearly twice as many words to describe the same event with the same fundamental information (nurse, thermometer, ear, beep). So, would both these children be considered equally credible witnesses, and would jurors rate the more talkative child more favourably than the less talkative child?

Credibility, Quality, and Quantity

Witness credibility is the extent to which a witness provides accurate and reliable testimony that is free from error. Credible witnesses are considered trustworthy, believable, and capable of portraying their account of an event with sufficient detail. Basically, child witness credibility is determined by the quality and quantity of their testimony. The quality of children's testimony is assessed by considering the accuracy of the information that children report. Quantity of child witness testimony is assessed be considering the amount of information, number of details, report length, and number of spoken words that children provide.

The impact of talkativeness on the quality and quantity of child witness testimony has not been directly researched, and so the effect of talkativeness on child witness credibility remains unknown. However, there are different perspectives of research that touch on how talkativeness might affect the quality and quantity of child witness testimony, and ultimately

child witness credibility. For instance, talkativeness is implicated in language development, where vocabulary significantly increases with age (Anglin, 1993), and maltreated children have poorer language skills compared to their peers—including expressive language skills (Lum, Powell, Timms, & Snow, 2015)—and talkativeness is inherently related to expressive language. Talkativeness is implicated in the sociocultural language theory because talkativeness is a product of both language and narrative development. Language and narrative development are fundamental aspects of the sociocultural language theory, and they, in turn, affect the development of autobiographical memory (Nelson & Fivush, 2004). Talkativeness is implicated Grice's rules of conversation, where monitoring amount and relevance (of contributions to a conversation) is considered fundamental to effective communication (Grice, 1975). Talkativeness (or lack of) has also been implicated in research investigating shyness, where shy children recalled less information than their less shy peers (Chae & Ceci, 2005) and shy children say less than their less shy peers (Crozier & Perkins, 2002). All in all, talkativeness is conceptually related to many fields of research, and since talkativeness has not been directly assessed (in terms of the effect on child witness credibility or juror perceptions) this highlights a gap in the research that the present research will address.

Children can be Reliable Witnesses

Overall, several decades of research has demonstrated that children can be reliable witnesses (Lamb, Brown, Hershkowitz, Orbach, & Esplin, 2018)—by the age of five most children can provide a coherent and chronological account about an event they have experienced or witnessed (Fivush, 2011). But do jurors agree that children can be reliable witnesses? Research shows that jurors perceive younger children as less reliable than older children (Brainerd & Reyna, 2012; Connolly, Price, Lavoie, & Gordon, 2008; Newcombe & Bransgrove, 2007; Ross, Dunning, Toglia, & Ceci, 1989; Wright, Hanoteau, Parkinson, &

Tatham, 2010). In an effort to help children provide reliable testimony there has been extensive research into the best practices for conducting forensic interviews with children—leading to the development of internationally recognised interview protocols such as the National Institute of Human Development (NICHD) protocol (Lamb & Brown, 2006; Lamb, Orbach, Hershkowitz, Esplin, & Horowitz, 2007; Lamb, Sternberg, & Esplin, 1998). By using open-ended questions these interview protocols serve to elicit testimony from children that gives children the best chance at being credible witnesses without compromising on accuracy. However, there is very little research about the verbal behaviour of children providing testimony in these interviews.

When interviewed using open-ended questions children can provide highly accurate memory reports regardless of age (Brown & Lamb, 2015; Lamb et al., 2018). When children respond to open-ended questions and are encouraged to speak freely, it opens up the potential for talkativeness to have an effect on the testimony children provide, and then talkativeness might affect juror evaluations of child witness testimony. By comparing how jurors evaluate the accuracy and informativeness of children's high-talkative/low-talkative testimony with how accurate and informative the children actually are, we can identify potentially harmful beliefs that should be addressed in court to prevent false acquittals or convictions. Therefore this study manipulates how talkative a child's description of a past event is, to determine whether jurors' ratings of the child change accordingly. We are particularly interested in whether jurors are swayed/more persuaded by children who say more, even when the amount of information included in their account is held constant, and whether those beliefs/perceptions are appropriate—so we need to understand the evidence about whether there are reliable associations between how much children report, and how reliable their testimony is.

Juror Perceptions of Child Witnesses

When assessing the credibility of a child witness, jury members may make their judgements based on factors such as child age (Ross, Dunning, Toglia, & Ceci, 1990), number of details that children recall (Henry, Ridley, Perry, & Crane, 2011), and confidence (Newcombe & Bransgrove, 2007). Jurors might also make inferences about whether a child witness is accurate or not. Worryingly, jurors (Goodman et al., 1989; Connolly et al., 2008) and various professionals (e.g., social workers, psychologists, police officers, lawyers, and judges) (Bala, Ramakrishnan, Lindsay, & Lee, 2005) are unable to reliably tell the difference between an accurate and an inaccurate account from children. Over twenty years ago Westcott, Davies, and Clifford (1991) found that jurors detection of children's accuracy was just above chance, and similar findings have been reported more recently (Gongola, Scurich, & Quas, 2017).

So, if jurors are unable to use accuracy to determine if a child witness is credible or not then they have to use other means to evaluate child witnesses. Henry et al. (2011) discovered that child witnesses who included more details in their testimony received more favourable ratings of witness credibility. Given jurors can be swayed by how much information children report when judging how reliable (or accurate) they are, what do we know about whether quality and quantity of testimony are associated? Leippe, Manion, and Romanczyk (1992) found that the way in which report length (defined as number of words from free-recall) predicted memory accuracy differed with the age of the witness. Longer report lengths were associated with better accuracy for both child (9- to 10-years) and adult witnesses. In contrast, longer report lengths were associated with decreased memory accuracy for younger children (5- to 6-years). Other speech-related characteristics that predicted accuracy were the use of "don't-know" responses, and powerfulness of speech—characterised by the lack of language features such as hedges (e.g., "kinda", "I guess") and

hesitations (e.g., “um”, “well). Subjectively rated “narrativeness” (defined by longer utterances and frequent elaborations) was not associated with memory accuracy. Thus we see that measured talkativeness was associated with memory accuracy but perceived talkativeness (“narrativeness”) was not associated with memory accuracy, and that the relationship between quantity and quality changed with age.

Other researchers have also demonstrated variability with age in these associations between quantity and accuracy—Kulkofsky, Wang, and Ceci (2008) showed a detrimental effect of lengthy reports on 3- to 5-year-old children’s accuracy, and Teoh and Chang (2018) showed the opposite with 6-to 7-year-old children. So in some cases (e.g., with older children), jurors may have it right if they evaluate more talkative children as better witnesses, whereas in others (younger children) they may overestimate how accurate the children are if they are influenced by talkativeness in the same way. But generally speaking, as children get older they provide more detail (Cashmore & Bussey, 1996; Malloy & Quas, 2009; Poole & Lamb, 1998). So, jurors would probably be justified in assuming that a child who says more is also providing more information.

To further illustrate the reasons for conducting the present research we now briefly discuss the following: the role of language, speech styles, shyness, and confidence in child witness testimony; a brief overview about general misconceptions about child witnesses; and a brief introduction to conducting juror perception research.

Language

Talkativeness is a display of expressive language and communication. Language in itself is fundamental for the development of autobiographical memory and effective communication. Expressive language is the ability to use language to communicate and express ourselves and talking is how we express ourselves through spoken language. For a child witness to provide testimony they must have sufficient expressive language and speech

skills to form and present a cohesive account of what they remember. The sociocultural theory of autobiographical memory highlights the importance that language and narrative skills play in the role of autobiographical memory development (Nelson & Fivush, 2004). While the ability to form and recall autobiographical memory emerges in the preschool years, the sociocultural theory acknowledges that autobiographical memory continues to mature throughout childhood—as do narrative skills and language skills. Nelson and Fivush (2004) illustrate three ways in which language is fundamental to the development of autobiographical memory. Firstly, language is critical to the expression and organisation of autobiographical memory. Secondly, as children interact with others they further develop skills in sharing and organising their memories. Thirdly, children then become increasingly aware that memories are reflections of the past that can be interpreted from multiple perspectives. As children get older the number of words in their repertoire (vocabulary) significantly increases (Anglin, 1993), and so does their ability to form a cohesive narrative (Miragoli, Camisasca, & Di Blasio, 2017). So would children who are more talkative—and subsequently more practised with language—have better memories, and provide more coherent accounts of these memories? Also, if talkativeness is associated with better memory and more coherent memory accounts then would jurors be able to pick up on this particular association between memory and language?

With rapid language development across childhood, and the ongoing development of autobiographical memory and memory recall, how does this transfer to conversational skills or the ability to provide testimony? In a philosophical essay, Grice (1975) posited that there are four maxims (rules) involved in having an effective conversation. These maxims are quantity, quality, relation, and manner. The quantity maxim refers to the need to provide sufficient but not excessive contributions to the conversation. The quality maxim refers to the accuracy of the content. The relation maxim refers to providing information that is relevant to

the topic. The final maxim of manner is about being clear and concise in contributions to the conversation. Talkativeness is most closely related to the quantity and the manner maxims—but all four of these maxims are important to providing reliable and credible testimony. Research that investigates children's ability to understand and adhere to these maxims is limited—however we have already discussed research that addresses the quality maxim in regards to providing testimony. What the research about children's ability to attend to the maxims does tell us is that while six-year-olds may be able to attend to Grice's maxims (particularly quality and relation), pre-schoolers may still be developing their ability to recognise the rules of good conversation (Eskritt, Whalen, & Lee, 2008; Vazquez, Delisle, & Saylor, 2013).

Whilst language skills develop over time, there is also evidence that language skills may vary as a result of whether children have been abused or not. A recent meta-analysis of 26 studies showed that maltreated children tend to have poorer language skills compared to non-maltreated children (Lum et al., 2015). These maltreated children may be required to provide testimony about being abused or neglected. So, if maltreated children are more likely to have poorer language skills then this may affect their ability to provide testimony and be seen as a credible witness by a jury.

Speech Styles

Jurors are capable of detecting differences in speech style (e.g., O'Barr & Black, 1982) and these speech styles have been found to influence how jurors evaluate child witnesses (Ruva & Bryant, 2004). Depending on the age of the witness, powerfulness of speech has different effects on juror perceptions of witness credibility. Ruva and Bryant (2004) found that speaking in a powerless style was more harmful to the credibility of an adult witness than the credibility of a child witness. Additionally, they found that child witnesses using a powerful speech style were associated with higher ratings of defendant

guilt than child witnesses using a powerless speech style (Nigro, Buckley, Hill, & Nelson, 1989; Ruva & Bryant, 2004).

Shyness

Shyness is one characteristic that is often seen as synonymous with reticence (i.e. being less talkative) and there is evidence that this is a justified assumption (Crozier & Perkins, 2002; Evans, 1996; Reynolds & Evans, 2009; Van Kleeck & Street, 1982). For instance, the reticence of shy children has been linked to shorter utterances (Crozier & Perkins, 2002); and shy children often score lower on measures of language skill than their less shy counterparts (Smith Watts et al., 2014; Spere, Evans, Hendry, & Mansell, 2009). Talkativeness is a form of expressive language, and shy children are known to score lower on measures of expressive language (Coplan & Evans, 2009; Smith Watts et al., 2014). In terms of how shyness relates to child witnesses, Chae and Ceci (2005) found that shyness was significantly correlated with the amount of information recalled—where shy children recalled fewer details than children who were rated as less shy. Shyness was also significantly correlated with age—where younger children were rated as shyer than older children (Chae & Ceci, 2005).

Confidence

Confidence can be portrayed not just by what a witness says, but also the way in which they say it. Being confident can be considered the opposite of shy, and powerful speech styles can be considered a display of confidence. Simply changing the way we speak (e.g., by speaking faster) can make us come across as more confident (Hughes, Mogilski, & Harrison, 2013) and individuals who appear more confident receive more positive evaluations from others than individuals who appear less confident (Price & Stone, 2004). Perceptions of confidence are even involved in how jurors perceive the overall credibility of a child witness, where confident child witnesses are rated in a more favourable way than child witnesses who

are less confident (Goodman, Bottoms, Herscovici, & Shaver, 1989; Leippe et al., 1992; Newcombe & Bransgrove, 2007). It is therefore important to consider confidence because of the role it plays in perceptions of credibility and because of the association is seems to have with talkativeness.

Juror Misconceptions

People tend to have negative stereotypes about child witnesses—believing them to have poor memories and be prone to misinformation and influence from others (Ross et al., 1990). But the perception of what jurors believe and the reality of what the scientific literature indicate are not always compatible—resulting in jurors having misconceptions about child witnesses. When it comes to jurors' beliefs about children's ability to provide accurate testimony jurors are generally justified in their beliefs, but misconceptions are still common (Quas, Thompson, Alison, & Stewart, 2005). For example when Quas et al. (2005) assessed jurors beliefs about children providing increasingly more details in their testimonies over time, 52% of participants were incorrect in assuming that more details was indicative of greater accuracy. This misconception about more details being associated with greater accuracy also applies to eyewitnesses in general (children and adults) (Akhtar, Justice, Knott, Kibowski, & Conway, 2018; Conway, Justice, & Morrison, 2014). Furthermore there is a general misconception that the presence of minor details in (adult) eyewitness testimony is an indicator of accuracy (Magnussen, Melinder, Stridbeck, & Raja, 2010). Even judges and magistrates have misconceptions about child witnesses, including that children under 7-years-old are unable to be competent witnesses (Cashmore & Bussey, 1996). Overall people tend to underestimate children's general ability to testify. Jurors also tend to incorrectly assume that more details in testimony means that the testimony must be more accurate.

In summary, juror perceptions can be influenced by many characteristics of child witness testimony and we have focused on a set of characteristics that are conceptually

related to talkativeness. The influence of language, speech styles, shyness, and confidence can all leave an impression on jurors and affect their perceptions of child witness credibility. Juror perception research serves as a guide for where misconceptions about child witnesses might exist, thus identifying possible risks for miscarriages of justice. We therefore conducted the present research to take a closer look at whether talkativeness plays a significant role in juror perceptions of child witnesses—and whether these perceptions are justified by the relationship between talkativeness and the quality and quantity of child witness testimony.

Juror Perception Research

Juror perception research typically uses "mock" jurors as research participants. Mock jurors tend to be university students, or in some cases a sample from the general population. University students are usually recruited as part of meeting course requirements or through university noticeboards. When participants are a sample from the general population then recruitment methods become more varied. Recruiting from the general population is challenging due to both difficulty accessing people (e.g., low uptake and difficulty advertising the research) and time constraints (participants taking time out of their day to participate). Recruitment from the general population can be leveraged by taking advantage of certain situations (e.g., recruiting jurors released from jury duty) and/or improving the accessibility of research materials (e.g., online surveys). With advances in technology, another method of recruiting participants is through crowdsourcing.

Crowdsourcing is an efficient method for gathering large numbers of people together to complete a task. In terms of crowdsourcing via the internet, this means getting individuals from the general population to come together to provide information about a particular topic—usually through a survey or computer-based task. Crowdsourcing in this way means that recruitment happens simultaneously as data collection—significantly reducing the

overall time and cost involved in collecting data. Participants recruited through crowdsourcing are typically compensated for their time via point reward programs or—as with the case for Amazon's Mechanical Turk (MTurk; www.mturk.com)—monetary compensation for time and effort expended (see Method for more details about MTurk).

The Present Research

The first study of the present research explores jurors' perceptions of talkativeness (number of spoken words) in children and how they rate different aspects of 5-, 8-, and 12-year-old children's testimony—whilst holding the content of the testimony constant. So, each child (regardless of age and level of talkativeness) talks about the same components of a health-check with a nurse and includes the same fundamental details about each component. For example, each child talks about getting their height measured with a tape measure but vary in the number of words they use—e.g., the low-talkative 5-year-old says "She said how big I was. Cos there's a measuring tape on the wall" and the high-talkative 5-year-old says "There was a yellow thing with numbers on it and I had to go stand with my heels on the wall to see how high I am and that's all". The transcripts have also been compiled from memory interviews with children and so—unlike many other studies—instead of just adding in extra words to create a different transcript, we took content from actual memory interviews with 5-, 8-, and 12-year-old children.

The second study examines transcripts from 5-, 8-, and 12-year old children, who were interviewed about a recent health-check, to determine whether talkativeness was related to the amount of information recalled, and the accuracy of children's reports.

Hypotheses.

Based on previous research demonstrating that jurors rate children more favourably who: provide more details (Henry et al., 2011), appear more confident (Goodman et al., 1989; Leippe et al., 1992; Newcombe & Bransgrove, 2007), and use a powerful style of speech

(Nigro et al., 1989; Ruva & Bryant, 2004), we expect that mock jurors in Study 1 will rate high-talkative children more favourably than low-talkative children. We believe jurors will rate high-talkative children more favourably because jurors will associate high-talkativeness with providing more details and thus as an indicator of accuracy of testimony. We also believe that jurors will associate high-talkativeness with increased confidence of the child and thus high-talkativeness will be associated with a more credible witness.

We also expect that older children will be rated more favourably than younger children, as consistently shown in previous research (Newcombe & Bransgrove, 2007; Talwar, Lee, Bala, & Lindsay, 2006; Wright et al., 2010). Interaction effects between talkativeness and child age is more difficult to predict because previous research has produced mixed findings for research conceptually related to talkativeness. For instance Ruva and Bryant (2004) did not find a relationship between speech style and 6- and 10-year-old children, but Goodman et al. (1989) did find a significant correlation between child witness confidence and child age. If we were to find an interaction effect then we expect that age effects might diminish for high-talkative children, because high-talkative children appear to provide more information or appear more confident, thus overriding the effect of child age. On the other hand we might expect no interaction effect where any effects of talkativeness are not dependant on child age because jurors consider talkativeness and child age as separate (or irrelevant) factors.

For the analysis of actual child witness memory interview transcripts from Study 2, we do not expect talkativeness to have an effect on the accuracy of children's memory for an experienced event. We base this hypothesis off research that shows how children provide highly accurate memory reports when interviewed using open-ended questions (Lamb et al., 2018) and that when a predominantly open style of questioning is used accuracy does not differ with age (Brown et al., 2013).

Thirdly, we expect that high-talkative children will provide more information than low-talkative children—as supported by research showing that shy children recall less information than less shy children (Chae & Ceci, 2005). We also expect that older children will provide more information than younger children as seen with previous research (e.g., Cashmore & Bussey, 1996; Malloy & Quas, 2009).

Method

To assess talkativeness of child witnesses and juror perceptions of child witness credibility three studies were employed. A pilot survey served as a manipulation check for Study 1 to ascertain whether mock jurors could detect a difference in amount of speech between transcripts from a high-talkative child witness and those from a low-talkative child witness. The pilot survey also tested the general administration and functionality (e.g., time to complete the survey) of the witness credibility questionnaire. Study 1 investigated the influence of talkativeness and child age on mock jurors' perceptions of child witnesses' credibility. Study 2 investigated the impact of talkativeness and child age on the accuracy and amount of information of children's recollection of an experienced event during a memory interview.

This research was approved by the Victoria University of Wellington, School of Psychology Human Ethics Committee under delegated authority to the Victoria University of Wellington Human Ethics Committee, application # 0000025460.

The current research used interview transcripts from the Memory: Age-Related Changes in Errors (MARCIE) project led by Dr Deirdre Brown (Principal Investigator). Children participating in the MARCIE Project had a health-check with a research nurse. They were then interviewed about their memory of the health-check two days later, using an adaptation of the NICHD interview protocol (e.g., Brown & Lewis, 2013). The NICHD protocol is an evidence-based protocol developed for investigative interviewing of children

and is widely used in research and situations (e.g., forensic interviews with child abuse victims) that involve interviewing children about their recollection of a previously experienced or witnessed event (Lamb et al., 2007). Composite transcripts for the pilot survey and Study 1 survey were formed from some of these interviews, and the data for Study 2 was also taken from these interviews.

Method: Pilot Survey

Recruitment

This study recruited participants using Amazon's Mechanical Turk (MTurk). MTurk is an internet based crowdsourcing platform that allows for participants (known as "workers"), who meet specific criteria, to be recruited through advertisements of Human Intelligence Tasks (HITs). MTurk has been established as a cost effective method for gathering large amounts of data over a relatively short period (Buhrmester, Kwang, & Gosling, 2011; Paolacci & Chandler, 2014). Workers see the advertised HITs and decide which task they would like to complete. Upon completion of a HIT, workers receive an agreed payment for that HIT. Filling out an online questionnaire is a common type of MTurk HIT. In academic research, TurkPrime (www.turkprime.com) is often used in conjunction with MTurk to assist with administering research studies.

TurkPrime is an online third party platform that facilitates the collection of data from the MTurk participant pool (Litman, Robinson, & Abberbock, 2017). Using TurkPrime allows researchers to target MTurk workers based on a wide variety of criteria. For instance, HITs can be targeted to workers of a particular gender, age, level of education, locality, and experience completing previous HITs—just to name a few of the recruitment criteria options available.

Participants

Participants ("workers") were crowdsourced from the MTurk worker pool. $N = 70$ workers responded to the Human Intelligence Task (HIT) advertising the pilot survey, and 50 eligible workers completed the survey (71.4%). From here on workers who completed the survey are referred to as participants. Table 1 presents the participant demographics for the pilot survey. There was a relatively equal distribution across gender with 52.0% male and 48.0% female participants. Participants ranged in age from 20 to 51 years of age ($M = 32.62$, $SD = 8.01$). Just over half (54.0%) of the participants indicated that they had children. Participants who moved all the way through the survey (regardless of their responses to the eligibility questions) qualified for compensation of $1USD (United States Dollar).

Inclusion Criteria: HIT Requirements.

MTurk workers needed to be: at least 18 years of age; have not completed surveys from this particular research project; located in Canada or United States; have a HIT approval rating of at least 81% (have completed at least 81% of all their previous HITS); and have completed at least 100 previous HITs. Participants were recruited from within the United States and Canada as this is where the majority of MTurk workers are based. The survey was only advertised to workers who met HIT requirements.

Eligibility.

Eligibility questions—adapted from Brown and Lewis (2013)—determined whether participants were eligible for jury service and also whether participants had sufficient fluency in English to participate in the second section of the survey. The eligibility for jury duty question was derived from the relevant sections of the New Zealand Government Ministry of Justice website (New Zealand Ministry of Justice, n.d.) and United States Courts website (United States Courts, n.d.). To be eligible for jury duty participants must not: be under the age of 18 years; have an intellectual disability; work in government/congress/parliament

Table 1

Demographic Responses of Eligible Pilot Survey Participants

Demographic	*n*	%
Gender		
Male	26	52.0
Female	24	48.0
Age		
20 – 29 years	25	50.0
30 – 39 years	15	30.0
40 – 49 years	9	18.0
50 – 59 years	1	2.0
Native Language English		
Yes	49	98.0
No answer	1	2.0
English Fluency		
6 out of 7	1	2.0
7 out of 7	49	98.0
Have Children		
Yes	27	54.0
No	23	46.0

Note. Participants were also able to identify as transgender or other but no participants selected either of these options. There was one participant who did not answer question about whether English was their native language. Participants with an English Fluency rating of four or less out of seven (where seven equals "very") were excluded.

(including federal or state); work in a court of law; be employed by police/fire/ambulance departments; be a member of armed forces or defence force; or have been sentenced to imprisonment or charged with a felony.

Participants who were not eligible for jury service or were not sufficiently fluent in English were automatically directed to the Excluded Participants Debrief page (see Appendix A: Survey Participant Information). Eligible participants were automatically directed to the second section of the survey.

Exclusion Data.

A total of twenty participants were excluded from analysis—11 participants were excluded from the survey because they did not meet the jury service eligibility criteria, eight participants were excluded because they did not answer the juror perception survey questions, and one participant was excluded because they did not answer any of the questions. No participants were excluded based on their responses to the attention check-data. The data and analysis of the remaining 50 participants forms the basis of the pilot survey results.

Materials

Online research platforms.

TurkPrime facilitated the administration and recruitment of survey participants by setting and enforcing the eligibility criteria, and linking potential participants from MTurk with access to the survey via an advertisement (see Appendix A: Survey Participant Information)—that included an anonymous link—on the MTurk website.

The pilot survey was administrated using the Qualtrics research platform and utilised the survey flow tool and block randomiser functions. The survey flow tool enforced the eligibility criteria and if a participant was not eligible to participate then the survey flow tool automatically sent the respondent to the exclusion information sheet. The block randomiser ensured that participants were randomly assigned to one of the counterbalance conditions.

Pilot Survey Layout.

The pilot survey (see Appendix B: Survey Questions) had two sections. The first section contained an information sheet about the research (see Appendix A: Survey Participant Information), and a series of eligibility questions. The second section of the survey contained: an introduction to the research (with instructions on how to answer the questions), six transcripts, transcript rating questions, six attention check question, and the research debrief page (see Appendix A: Survey Participant Information).

Section One of the Pilot Survey.

Consent.

As the surveys were anonymous participants automatically implied consent by moving through the survey and answering the survey questions.

Section Two of Pilot Survey.

Eligible participants answered one set of witness credibility rating questions (for the first transcript, see Figure 1) and then five sets rating questions relating to talkativeness, conciseness, and amount of information (for the remaining five transcripts, see Figure 2). Each transcript was presented on a separate page of the survey, followed by the set of rating questions corresponding to it. At the bottom of the page was an attention check question that asked the age of the child from each transcript. This attention check served as a tool for assessing whether participants were sufficiently attending to the survey and that survey participants were authentic and genuine "mock jurors" (e.g., not robots).

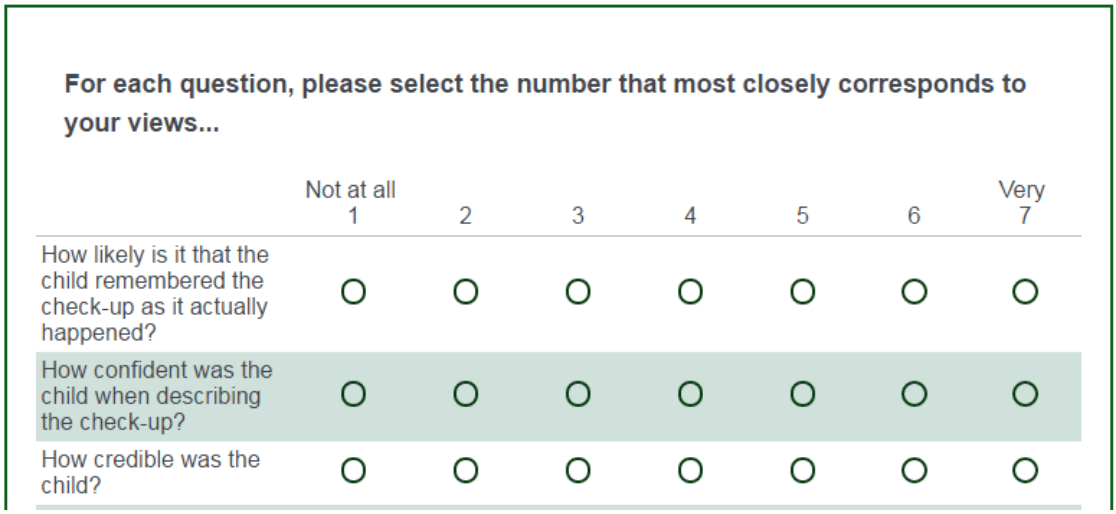

Figure 1. Screen shot from survey showing witness credibility rating questions on a seven point Likert Scale.

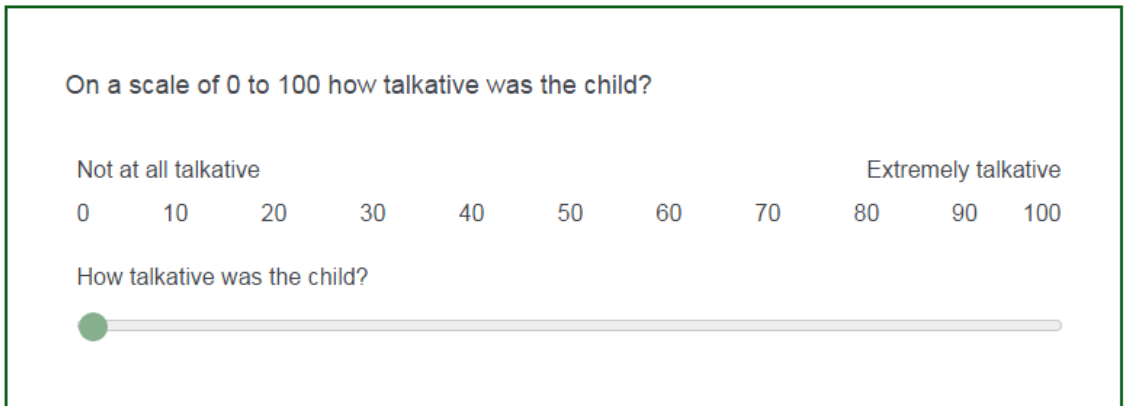

Figure 2. Screen shot showing pilot survey alternative rating question on a 100 point sliding rating scale.

Transcripts.

Transcripts were composites of interviews with children who participated in The MARCIE Project. A total of six transcripts (see Appendix C: Transcripts for Pilot Survey and Study 1 Survey) were prepared. The transcripts varied by the age of the child (5, 8, or 12 years) and the level of talkativeness of the child (low or high).

Talkativeness was ascertained by using the number of words spoken in children's interviews about the same aspect of the health-check. Interviews with the highest proportion of word counts for child utterances (for each age group) were used to create the high-talkative transcript composites. Interviews with the lowest proportion of word counts for child utterances (for each age group) were used to create the low-talkative transcript composites.

Content was controlled for by taking descriptions from the same aspects of the health-check (e.g., height measured, taking blood pressure) from age- and talkativeness-matched children to create the composite transcripts. This means that for all transcripts, the interviewer asked the same number and type of questions and that the components of the health-check recalled were the same for each transcript (in terms of what the child remembered about the

health-check). Each child talked about getting their height measured with a tape measure, getting their ears checked with a light, getting their mouth checked with a torch and stick, and getting their blood pressure taken on their arm with a strap that pumped up. Language style was controlled by reducing the number of hesitations (e.g., “um”, “er”) and hedges (e.g., “sort of, “kind of”) to just one or two per transcript.

Transcripts were presented in written format (see Figure 3 for an example of the visual presentation of a transcript and Appendix C: Transcripts for Pilot Survey and Study 1 Survey, for all transcripts versions)—previous research has found that using a written format produces similar results as using a video (Brimacombe, Quinton, Nance, & Garrioch, 1997; Ross et al., 1990). Using a written format also controls for extraneous variables such as prosody (speed of speech and use of pausing), intonation, volume of voice, and general demeanour (Areni & Sparks, 2005).

Witness credibility questions.

Witness credibility questions assessing constructs of witness credibility and talkativeness. These dimensions were memory, confidence, credibility, accuracy, honesty, amount, consistency, suggestibility, talkativeness, believability, informativeness, coherency, and reliability. Each set of witness credibility questions were always presented in the same order. Participants rated each transcript using the witness credibility questions—on a scale of one to seven, where one was labelled ‘not at all’ and seven was labelled ‘very’. Ratings for the suggestibility construct were reverse scored to align all witness credibility question ratings from least desirable child witness traits to most desirable child witness traits. For the purposes of this research, the talkativeness construct was not reverse scored as it is assumed that being more talkative is considered a desirable trait—as indicated by an emphasis on eliciting longer responses from children from open-invitation or cued invitation interviewer questions (Teoh & Lamb, 2013).

5 year old

(I = interviewer and C = child)

I I heard that a couple of days ago a nurse came to school and gave you a check-up. I wasn't there so I don't know what happened but I'd like to know all about it, so tell me everything that happened from the beginning to the end.

C She said how big I was.

I Okay, tell me some more things about the check-up

C She checked my ears

I Tell me some more things about the check-up even the little things

C She looked in my mouth and then I had a test on a blood checker.

...

I So earlier you also mentioned she said how big you were. So tell me everything you remember about when that happened.

C Cos there's a measuring tape on the wall.

I Alright, so you also talked about, checking your ears, tell me everything you remember about when that happened.

C She used a light to check them

I So you also mentioned, she looked in your mouth, so tell me everything you remember about when that happened.

C She put the stick in then up, she said open your mouth, and then, well she put the stick in and then she shined the light and then all done.

I Ok, you also talked about the test on a blood checker, so tell me everything you remember about when that happened.

C When I had that I get to pump out on my arm and it got tight a little and then got a little hand little thing where you can twist and it comes loose. And it all.

Figure 3. Example of low-transcript rated by survey participants.

The witness credibility rating questions were adapted from Brown and Lewis (2013). Some new questions were added while other questions were removed to align with the main dimensions of child witness credibility research found in a selection of other juror perception of witness credibility research (see Table 2). Questions relating to how talkative, concise, and how much information the child provided were formulated specifically for the purposes of this research and were included in the bank of witness credibility rating questions. These ratings questions were developed for the pilot survey to assess whether participants could detect a difference in the level of talkativeness, conciseness, and amount of information between transcripts.

Table 2

Credibility Factors Used in Witness Credibility Questionnaires

Credibility Constructs	Research Using Witness Credibility Questionnaires						
	Brimacombe et al. (1997)	Brown and Lewis (2013)	Cooper, Quas, and Cleveland (2014)	Henry et al. (2011)	Mueller-Johnson, Toglia, Sweeney, and Ceci (2007)	Peled, Iarocci, and Connolly (2004)	Ruva and Bryant (2004)
Accuracy	•	•	•		•	•	•
Believability			•	•		•	•
Bias							•
Capability		•		•			
Cognitive Function				•	•		
Communication		•				•	
Competence	•	•			•	•	
Confidence	•	•	•	•	•	•	•
Consistency		•	•			•	•
Convincingness				•	•		•
Credibility	•	•		•			•
Forcefulness of response							•
Honesty/Trustworthiness	•		•	•	•	•	•
Intelligence			•				•
Likeability			•				
Memory/Completeness		•	•	•	•		•
Suggestibility		•			•	•	

Note. • = presence of this construct or a very similar construct

Attention check.

Each survey included six attention checks—one for each transcript. The attention check "How old was the child?" was designed to check and assess that participants were attending to the questionnaire in an appropriate manner (Buhrmester et al., 2011; Paolacci & Chandler, 2014). Attention checks are a sufficient method for ascertaining whether participants are appropriately attending to the survey. Hauser and Schwarz (2016) found that up to 95% of MTurk workers were able to pass attention checks, even when asked to complete a relatively novel attention check.

Each attention check was asked after each set of rating questions—one attention check per transcript. Participants who did not respond to three or more of the attention checks or provided inappropriate answers (e.g., nonsense words) were excluded from the research. Failure to appropriately attend to attention checks is an indication of satisficing (only engaging in the survey enough to provide the minimal expected response and not truly considering and responding to the question at hand), or that the survey is being completed by an artificial respondent, such as a robot, also known as "bots". There has been recent debate about the prolificacy of "bots" on MTurk, where research found potential illegitimate responses to pre-screening questions (Chandler & Paolacci, 2017). With careful research design including the use of attention checks the likelihood of data quality being compromised by "bots" can be significantly mitigated. The present research mitigated potential effects of "bots" or inappropriate responding by; implementing stringent HIT criteria using TurkPrime, including attention checks throughout the survey, and executing strict exclusion criteria across all survey responses.

Procedure

Participants read each transcript, rated the child, responded to the attention check and then moved to the next transcript. This process was repeated for each of the six transcripts, however the first transcript included the witness credibility questions and subsequent transcripts only included the rating questions about talkativeness, conciseness, and amount of information. Participants then read debriefing material and obtained their unique code for claiming credit for participation.

Research Design

The pilot survey employed a 3 × 2 (Witness age: 5, 8, or 12 years × Talkativeness: Low or High) within subjects' repeated measures design, where each survey participant rated each of the six different transcripts.

Counterbalancing.

Order effects were controlled for by using incomplete counterbalancing via random allocation to one of six orders as determined by a systemic balanced Latin square procedure. The systemic balanced Latin square procedure creates six different order conditions where each transcript appears in each position in the order presentation, and that each transcript proceeds each other transcript position. This technique controls for order and carry-over effects (Freeman, 1979). The different transcript presentation orders were as follows: Group A: 1, 2, 6, 3, 5, 4; Group B: 2, 3, 1, 4, 6, 5; Group C: 3, 4, 2, 5, 1, 6; Group D: 4, 5, 3, 6, 2, 1; Group E: 5, 6, 4, 1, 3, 2; and Group F: 6, 1, 5, 2, 4, 3. The first condition order was created using the formula 1, 2, *n*, 3, *n*-1, 4 (Shuttleworth, 2009) where *n* = number of conditions. The next condition order was created by adding one to the number in the same ordinal position as the previous condition (where after six, the ordinal condition reset to one). This process was repeated until there were six order conditions. Each transcript number refers to a different

transcript combination of age and talkativeness. Counterbalance cell sizes ranged from seven to ten.

Coding of Responses

All scale questions response scores were coded within the Qualtrics survey platform. These scores were exported from Qualtrics into SPSS where the data was consolidated to show individual participants ratings of each of the six transcripts presented in the survey. The suggestibility scale question was reverse coded.

Results

Results for the Pilot Study, Study 1, and Study 2 are presented respectively. Effect sizes are reported (Cohen's *d for t*-tests, generalised eta-squared η_G^2, and partial eta squared η_p^2 for ANOVA) as recommended by Lakens (2013). Supplementary information about checking assumptions of analyses are presented in Appendix D: Assumption Checks of Data. Statistical significance threshold is set to $p < .05$ for all analyses.

Pilot Study Results

The main purpose of the pilot analysis was to determine whether there was a significant difference between participant ratings of talkativeness for low-talkative child transcripts versus high-talkative child transcripts for each child age (5-, 8-, and 12-year-olds).

Comparing low-talkative and high-talkative children.

Because we were only interested in checking whether the manipulation of the level of talkativeness was successful we decided to run three separate paired samples t-tests (comparing low- versus high- talkativeness for the three age-matched child pairs). Paired samples t-tests determine if the difference between two means (low- versus high-talkative) is statistically significantly different from zero when the data is from the same individual. To conduct the paired samples t-tests a new variable (difference score) was calculated—the difference between participant ratings of high-talkative and low-talkative children. This

variable was calculated by subtracting low-talkative child ratings from high-talkative child ratings (for each child age, for each participant). Difference scores were then used as the dependent variable and level of talkativeness as the independent variable in the paired samples t-tests analyses.

As shown in Figure 4, there was a statistically significant mean difference between participants' ratings of high- and low-talkative children—transcripts from high-talkative children were rated by participants as more talkative than transcripts from low-talkative children, for all age groups. Effect sizes (*d*) (for each child age group were large (greater than .8). See Appendix D: Assumption Checks of Data for more detail.

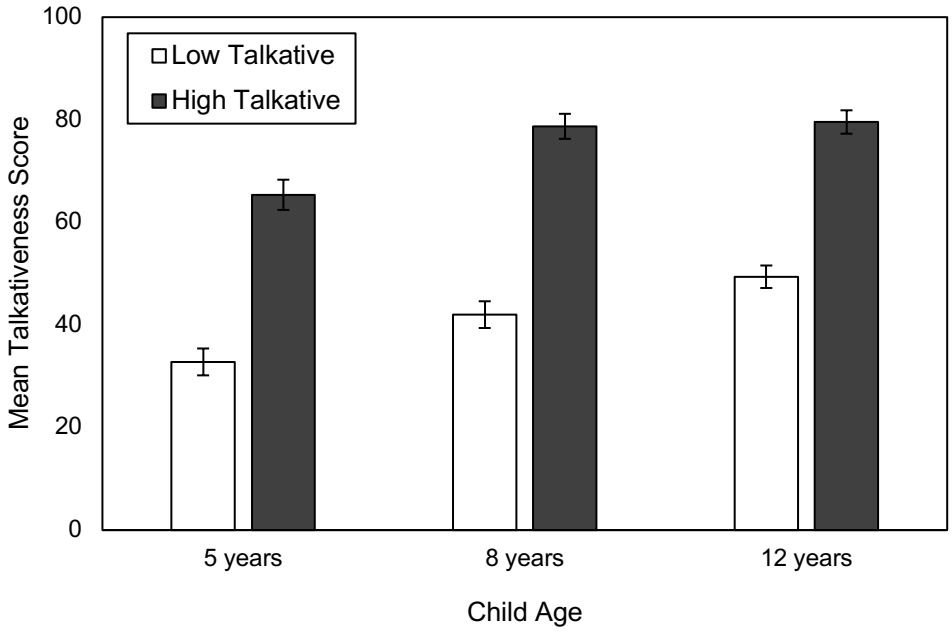

Figure 4. Mean talkativeness scores for each child with error bars representing standard error of the mean.

Discussion Pilot Study

Results of the pilot study confirmed that the manipulation of the level of talkativeness in the transcripts was detectable by participants—equally for children of different ages. Therefore these transcripts were deemed valid for use in the Study 1 survey.

Method: Study 1

The survey conducted for Study 1 was—for the most part—a replication of the pilot survey with a few adjustments and amendments. Firstly, an increased sample size was used to increase the power of analyses. An additional HIT qualification criteria was added in TurkPrime so that participants who had already responded to the pilot survey were automatically excluded from the Study 1 survey. Finally, the main difference between the pilot survey and the Study 1 survey was that the witness credibility questions (for an example

see Appendix B: Survey Questions) were presented after each transcript whereas in the pilot survey the witness credibility questions were only presented after the first transcript.

Participants

N = 425 workers responded to the HIT advertising the Study 1 survey and *n* = 263 eligible workers/participants completed the survey (61.8%). Table 3 presents the participant demographics for the Study 1 survey. There was a relatively equal distribution across gender with 45.2% of participants identifying as male and 54.0% as female. Participants ranged in age from 20 to 69 years of age (*M* = 36.40, *SD* = 11.48). Just over half (55.1%) of participants indicated that they had children.

Exclusion Data.

162 participants were excluded from analysis according to one or more of the following criteria: 105 participants were excluded from the survey because they did not meet the jury service eligibility criteria (including three participants under the age of 18 years); 18 participants were excluded because they did not provide their age; and 17 eligible participants were excluded because they answered less than 84 percent of survey questions. The decision was made to exclude five participants over the age of 72 as the veracity of their date of birth was dubious—when the distribution of participant ages was assessed these participants were identified as outliers. 19 participants were excluded on the grounds that they rated their English fluency as four or less (out of seven) or did not answer this question. The attention check data was visually assessed after all other exclusion criteria had been applied and 20 participants were then excluded due to insufficient responding to the attention checks. The data and analysis of the remaining 263 participants forms the basis of the Study 1 survey results. Counterbalance cell sizes ranged from 40 to 47.

Table 3

Demographic Responses of Eligible Study 1 Survey Participants

Demographic	*n*	%
Gender		
Male	119	45.2
Female	142	54.0
Transgender	2	0.8
Age		
20 – 29 years	89	33.8
30 – 39 years	102	38.3
40 – 49 years	34	12.9
50 – 59 years	22	8.4
60 – 69 years	16	6.1
Native Language English		
Yes	254	96.6
No	8	3.0
Did not answer	1	0.4
English Fluency		
5 out of 7	6	2.3
6 out of 7	18	6.8
7 out of 7	239	90.9
Have Children		
Yes	145	55.1
No	117	44.5
Did not answer	1	0.4

Note. Participants with an English Fluency rating of four or less (out of seven, where seven equals very) were excluded.

Materials

Survey Layout.

The Study 1 survey layout was the same as the pilot survey layout with slight changes to the information sheet, exclusion sheet, and debrief sheets (see Appendix A: Survey Participant Information). These changes were made to reflect the different purpose of the Study 1 survey research—to see whether talkativeness, and age of the child, has an effect on juror perceptions of child witnesses.

Development of Study 1 questionnaire.

Witness credibility questions.

These were mostly the same questions as those used in the pilot survey except, after reviewing the pilot survey results and finding unexpected mean ratings of the conciseness construct the additional construct of conciseness was added to the bank of witness credibility questions for the Study 1 survey to gather more data for analysis. So, the dimensions measured in the Study 1 survey were memory, confidence, credibility, accuracy, honesty, amount, consistency, suggestibility, talkativeness, believability, informativeness, coherency, reliability, and conciseness.

Study 1 Results

Study 1 tested whether talkativeness (low or high) and/or child age (5-, 8-, or 12-years) influenced participant ratings of child transcripts. Recall that we asked participants to make 14 different ratings of witness credibility constructs—on a scale of one to seven—for six different transcripts, but there were some missing data points from when participants did not provide a response. Missing data and internal consistency analyses were conducted first and these results showed that data was missing completely at random and the witness credibility scale had high levels of internal consistency. Details and statistics of the missing data and internal consistency analyses are presented in Appendix D: Assumption Checks of Data and Appendix E: Study 1 Survey Supplementary Data. Because the missing data was considered missing completely at random (and not due to any variable from the study e.g., participant gender) we created a separate dataset where missing data was imputed using the

expected maximisation (EM) method.[1] We took a conservative approach and present here analyses with the original non-imputed dataset.[2]

We ran a series of 2 (low- or high-talkative) x 3 (child age) repeated measures analyses of variance (ANOVA) to determine the effect of talkativeness and child age on participants' ratings of child transcripts. Observations, main effects, interaction effects, and simple main effects are presented below. All *p*-values were adjusted to account for multiple comparisons using the Bonferroni correction. In cases where the assumption of sphericity was violated we interpret the results using the Greenhouse-Geisser correction.

Observations

Before interpreting the two-way repeated measures ANOVA we made some observations about the mean values for each construct including grand means, means by high-/low- talkativeness, and means by child age. Recall that participants made ratings on a seven-point Likert scale where one equals "not at all" and seven equals "very" (suggestibility scores were reversed). As illustrated in Table 4, majority of the mean rating scores are above the mid-point of the rating scale, and while the scores generally reflect a favourable rating they are still well below the top end of the rating scale.

We also ran a principal component analysis of the witness credibility constructs to see if we could reduce the number of constructs in a meaningful way but found conflicting results where anything between two and ten different factors were identified, many of which did not make sense to combine. We therefore present the findings of Study 1 using the individual witness credibility constructs.

[1] We analysed the data using the EM imputed dataset. In the main, the results from the EM imputed data were similar, with the exception that the EM imputed dataset returned an additional significant interaction effect for ratings of *coherence* which was not observed with the raw dataset (See Appendix E: Study 1 Survey Supplementary Data).

[2] Using a within subjects design means there is sufficient power to conduct analysis with the original data. Given the novelty of the research we were unable to compare our findings to other research and determine if the results were unduly influenced by the EM imputed data.

Table 4

Descriptive Statistics for all Witness Credibility Constructs of Each Transcript

		M (SD)					
		5-years		8-years		12-years	
Witness Credibility Construct	*n*	Low	High	Low	High	Low	High
Accuracy	254	4.65 (1.45)	4.74 (1.49)	5.06 (1.28)	5.10 (1.43)	5.39 (1.27)	5.53 (1.39)
Amount of information	249	3.63 (1.57)	5.02 (1.36)	4.24 (1.45)	5.38 (1.26)	4.51 (1.40)	5.58 (1.36)
Believability	242	5.10 (1.40)	5.12 (1.55)	5.23 (1.44)	5.45 (1.40)	5.61 (1.28)	5.67 (1.43)
Coherency	253	4.52 (1.48)	4.57 (1.51)	4.87 (1.38)	4.87 (1.49)	5.40 (1.36)	5.15 (1.60)
Conciseness	250	4.68 (1.60)	4.44 (1.56)	4.67 (1.51)	4.42 (1.60)	5.14 (1.45)	4.74 (1.82)
Confidence	254	4.74 (1.53)	5.05 (1.41)	4.93 (1.45)	5.00 (1.49)	5.09 (1.42)	5.31 (1.52)
Consistency	250	4.84 (1.48)	5.00 (1.48)	5.12 (1.28)	5.35 (1.29)	5.47 (1.22)	5.53 (1.37)
Credibility	252	4.90 (1.50)	4.86 (1.52)	5.14 (1.30)	5.33 (1.31)	5.56 (1.30)	5.56 (1.43)
Honesty	256	5.38 (1.45)	5.25 (1.65)	5.45 (1.29)	5.56 (1.40)	5.80 (1.29)	5.76 (1.40)
Informativeness	253	4.06 (1.57)	4.86 (1.51)	4.53 (1.52)	5.16 (1.36)	4.96 (1.36)	5.45 (1.47)
Memory	257	4.67 (1.39)	4.85 (1.44)	5.11 (1.33)	5.30 (1.33)	5.53 (1.27)	5.59 (1.38)
Reliability	252	4.75 (1.54)	4.85 (1.57)	5.16 (1.30)	5.18 (1.44)	5.39 (1.37)	5.46 (1.44)
Suggestibility	255	4.73 (1.64)	4.53 (1.69)	4.62 (1.67)	4.44 (1.79)	4.65 (1.73)	4.61 (1.86)
Talkativeness	250	3.41 (1.52)	5.08 (1.38)	3.97 (1.49)	5.54 (1.25)	4.01 (1.43)	5.78 (1.24)

Main Effects

As recommended by Interpretation of the main effects is based on the estimated marginal means and their associated standard errors. Mean differences from pairwise comparisons are significant at the $p = .05$ level and adjustments for multiple comparisons are made using the Bonferroni correction.

Main effect of talkativeness.

We identified several statistically significant main effects of talkativeness in ratings of: *conciseness* $F(1, 249) = 11.80$, $p = .001$, $\eta_G^2 = 0.01$, $\eta_p^2 = 0.05$; *confidence* $F(1, 253) = 10.17$, $p = .002$, $\eta_G^2 = 0.00$, $\eta_p^2 = 0.04$; *consistency* $F(1, 249) = 7.41$, $p = .007$, $\eta_G^2 = 0.00$, $\eta_p^2 = 0.03$; *informativeness* $F(1, 252) = 91.50$, $p < .001$, $\eta_G^2 = 0.05$, $\eta_p^2 = 0.27$; *memory* $F(1, 256) = 6.37$, $p = .012$, $\eta_G^2 = 0.00$, $\eta_p^2 = 0.02$; *suggestibility* $F(1, 254) = 8.77$, $p = .003$, $\eta_G^2 = 0.00$, $\eta_p^2 = 0.03$; and *talkativeness* $F(1, 249) = 385.97$, $p < .001$, $\eta_G^2 = 0.27$, $\eta_p^2 = 0.61$. These effects had small effect sizes, except for informativeness which had a moderate effect size, and talkativeness which had a large effect size. No statistically significant main effects of talkativeness were found for ratings of *accuracy, believability, coherency, credibility, honesty*, and *reliability*. We also found an interaction effect between talkativeness and child age for ratings of *amount of information*, which is reported in detail later.

As illustrated in Figure 5 analysis of the means revealed that the high-talkative children were rated as more talkative, confident, consistent, informative, having better memory, more suggestible, but less concise than low-talkative children.

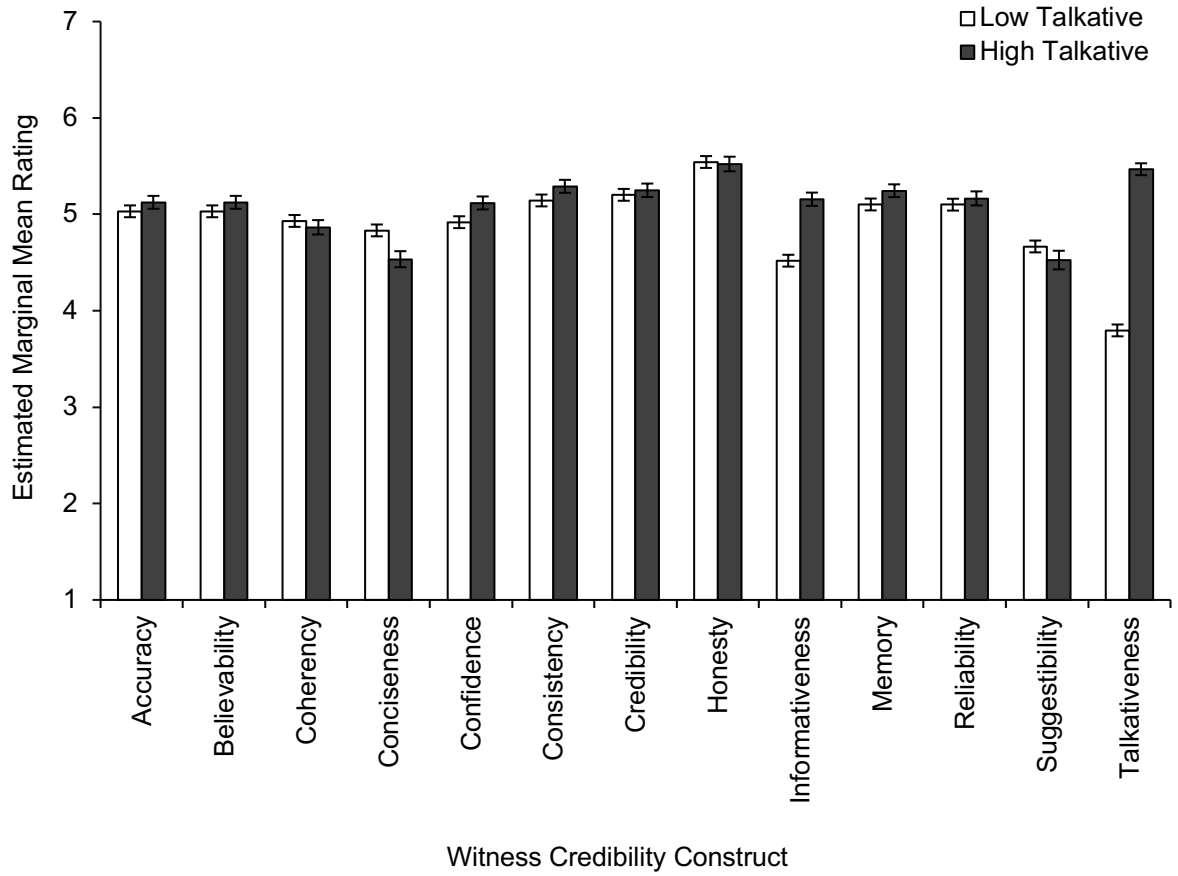

Figure 5. Estimated marginal means for low-/high- talkativeness for each construct. Error bars represent standard error of the mean. Suggestibility data presented are reverse scored.

Main effect of child age.

The main effect of child age showed statistically significant differences in ratings of: *accuracy* $F(1.92, 485.88) = 50.43$, $p < .001$, $\eta_G^2 = 0.05$, $\eta_p^2 = 0.17$; *believability* $F(2, 482) = 25.40$, $p < .001$, $\eta_G^2 = 0.02$, $\eta_p^2 = 0.10$; *coherency* $F(1.94, 490.12) = 45.37$, $p < .001$, $\eta_G^2 = 0.04$, $\eta_p^2 = 0.15$; *conciseness* $F(1.96, 486.80) = 15.78$, $p < .001$, $\eta_G^2 = 0.01$, $\eta_p^2 = 0.06$; *confidence* $F(2, 506) = 8.42$, $p < .001$, $\eta_G^2 = 0.01$, $\eta_p^2 = 0.03$; *consistency* $F(1.93,479.59) = 34.09$, $p < .001$, $\eta_G^2 = 0.03$, $\eta_p^2 = 0.12$; *credibility* $F(2, 502) = 45.97$, $p < .001$, $\eta_G^2 = 0.04$, $\eta_p^2 = 0.15$; *honesty* $F(1.89, 482.20) = 23.13$, $p < .001$, $\eta_G^2 = 0.02$, $\eta_p^2 = 0.08$; *informativeness* $F(1.92, 483.24) = 43.79$, $p < .001$, $\eta_G^2 = 0.04$, $\eta_p^2 = 0.15$; *memory* $F(2, 512) = 69.42$, $p < .001$, $\eta_G^2 = 0.06$, $\eta_p^2 = 0.21$; *reliability* $F(2, 502) = 38.91$, $p < .001$, $\eta_G^2 = 0.03$, $\eta_p^2 = 0.13$;

and *talkativeness* $F(1.93, 479.49) = 40.08$, $p < .001$, $\eta_G^2 = 0.04$, $\eta_p^2 = 0.14$. No statistically significant main effects of age were found for ratings of *suggestibility*. Most constructs returned small-medium effect sizes (believability, coherency, conciseness, honest, reliable), or medium effect sizes (accuracy, consistency, credibility, informativeness, talkativeness). A small effect was found for confidence, and a medium-large effect was found for memory.

As illustrated in Figure 6, the examination of the estimated marginal means (see Table 5) and pairwise comparisons (see Table 6) revealed that for the constructs of accuracy, believability, coherency, consistency, credibility, honesty, informativeness, memory, and reliability, ratings significantly increased between each pairing. For the constructs of conciseness, and confidence, the 12-year-olds were rated more positively than the two younger groups, who did not differ. For the construct of talkativeness, both of the two older groups of children were rated more positively than the 5-year-olds but did not differ from each other.

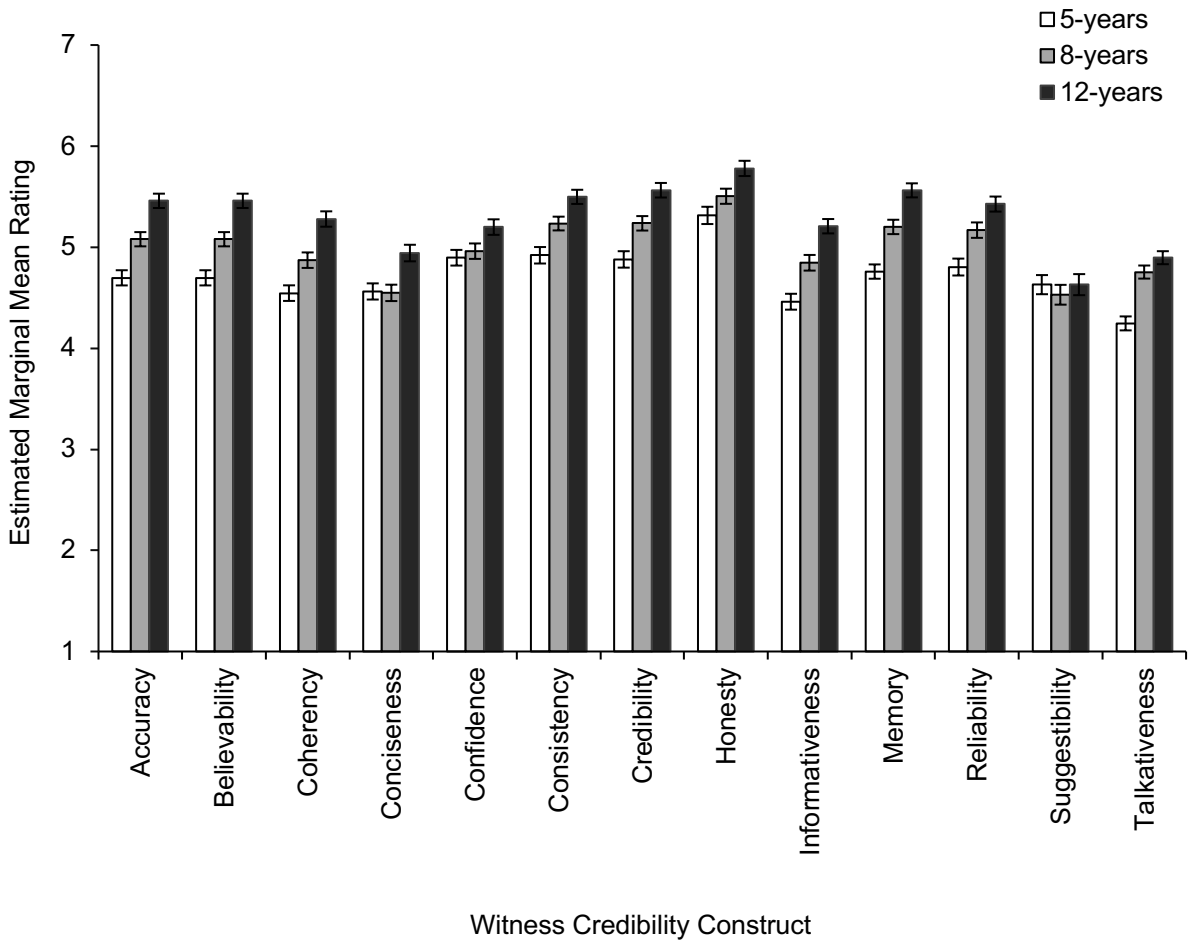

Figure 6. Estimated marginal means for 5-, 8-, and 12- year old child transcript ratings scores for each construct each construct. Error bars represent standard error of the mean. Suggestibility data presented is reverse scored.

Table 5

Estimated Marginal Means and Associated Standard Errors for Each Witness Credibility Construct by Child Age

		5-years		8-years		12-years	
Construct	*n*	*M* (*SEM*)	95% CI	*M* (*SEM*)	95% CI	*M* (*SEM*)	95% CI
Accuracy	254	5.03 (0.06)	[4.91,5.15]	5.12 (0.07)	[4.99, 5.26]	5.08 (0.06)	[4.97, 5.19]
Believability	242	5.11 (0.08)	[4.95, 5.27]	5.34 (0.08)	[5.19, 5.50]	5.64 (0.08)	[5.49, 5.78]
Coherency	253	4.55 (0.08)	[4.39, 4.70]	4.87 (0.08)	[4.72, 5.02]	5.28 (0.08)	[5.13, 5.43]
Conciseness	250	4.56 (0.08)	[4.40, 4.72]	4.55 (0.08)	[4.39, 4.71]	4.94 (0.08)	[4.78, 5.10]
Confidence	254	4.90 (0.08)	[4.74, 5.05]	4.96 (0.08)	[4.81, 5.11]	5.20 (0.08)	[5.05, 5.35]
Consistency	250	4.92 (0.08)	[4.76, 5.08]	5.23 (0.07)	[5.10, 5.37]	5.50 (0.07)	[5.36, 5.64]
Credibility	252	4.88 (0.08)	[4.72, 5.04]	5.24 (0.07)	[5.10, 5.38]	5.56 (0.07)	[5.42, 5.71]
Honesty	256	5.31 (0.09)	[5.15, 5.48]	5.50 (0.07)	[5.36, 5.65]	5.78 (0.08)	[5.63, 5.93]
Informativeness	253	4.46 (0.08)	[4.31, 4.62]	4.85 (0.08)	[4.69, 5.00]	5.21 (0.07)	[5.07, 5.35]
Memory	257	4.76 (0.07)	[4.62, 4.90]	5.20 (0.07)	[5.06, 5.34]	5.56 (0.07)	[5.43, 5.70]
Reliability	252	4.80 (0.08)	[4.64, 4.97]	5.17 (0.08)	[5.02, 5.32]	5.43 (0.07)	[5.28, 5.57]
Talkativeness	250	4.25 (0.07)	[4.11, 4.38]	4.75 (0.06)	[4.63, 4.88]	4.90 (0.06)	[4.77, 5.02]

Table 6

Mean Difference Scores and Associated Standard Errors for the Main Effect of Child Age

Construct	Comparison	M_{diff}(*SE*)	p[a]	95% *CI*
Accuracy	5-8 years	-0.38 (0.08)	< .001	[-0.56, -0.20]
	8-12 years	-0.38 (0.07)	< .001	[-0.55, -0.21]
	5-12 years	-0.76 (0.08)	< .001	[-0.96, -0.56]
Believability	5-8 years	-0.24 (0.07)	.003	[-0.41, -0.07]
	8-12 years	-0.29 (0.08)	< .001	[-0.47, -0.11]
	5-12 years	-0.53 (0.08)	< .001	[-0.72, -0.34]
Coherency	5-8 years	-0.33 (0.08)	< .001	[-0.52, -0.13]
	8-12 years	-0.41 (0.07)	< .001	[-0.58, -0.24]
	5-12 years	-0.73 (0.08)	< .001	[-0.92, -0.54]
Conciseness	5-8 years	0.01 (0.08)	1.000	[-0.18, 0.21]
	8-12 years	-0.39 (0.07)	< .001	[-0.57, -0.22]
	5-12 years	-0.38 (0.08)	< .001	[-0.58, -0.18]
Confidence	5-8 years	-0.06 (0.08)	1.000	[-0.25, 0.12]
	8-12 years	-0.24 (0.07)	.004	[-0.41, -0.06]
	5-12 years	-0.30 (0.08)	.001	[-0.50, -0.10]
Consistency	5-8 years	-0.31 (0.07)	< .001	[-0.49, -0.14]
	8-12 years	-0.26 (0.06)	< .001	[-0.42, -0.11]
	5-12 years	-0.58 (0.07)	< .001	[-0.76, -0.40]
Credibility	5-8 years	-0.36 (0.07)	< .001	[-0.53, -0.19]
	8-12 years	-0.33 (0.07)	< .001	[-0.49, -0.16]
	5-12 years	-0.69 (0.07)	< .001	[-0.86, -0.50]
Honesty	5-8 years	-0.19 (0.07)	.017	[-0.35, -0.03]
	8-12 years	-0.28 (0.06)	< .001	[-0.42, -0.13]
	5-12 years	-0.47 (0.08)	< .001	[-0.65, -0.28]
Informativeness	5-8 years	-0.39 (0.08)	< .001	[-0.59, -0.18]
	8-12 years	-0.36 (0.07)	< .001	[-0.53, -0.19]
	5-12 years	-0.75 (0.08)	< .001	[-0.95, -0.54]
Memory	5-8 years	-0.44 (0.07)	< .001	[-0.61, -0.27]
	8-12 years	-0.36 (0.06)	< .001	[-0.51, -0.21]
	5-12 years	-0.80 (0.07)	< .001	[-0.97, -0.63]
Reliability	5-8 years	-0.37 (0.07)	< .001	[-0.54, -0.19]
	8-12 years	-0.26 (0.07)	< .001	[-0.42, -0.10]
	5-12 years	-0.62 (0.07)	< .001	[-0.80, -0.44]
Talkativeness	5-8 years	-0.51 (0.08)	< .001	[-0.71, -0.31]
	8-12 years	-0.14 (0.07)	.125	[-0.31, 0.03]
	5-12 years	-0.65 (0.08)	< .001	[-0.83, -0.47]

Note. M_{diff} = mean difference of comparison; *SE* = standard error of mean difference; *CI* = confidence interval.
[a] Adjustment for multiple comparisons using Bonferroni correction

Interaction effect

Participant ratings for *amount of information* ($F(1.93, 479.54) = 3.23$, $p = .042$, $\eta_G^2 = 0.00$, $\eta_p^2 = 0.01$) produced a statistically significant interaction between talkativeness and child age with a small effect size. To unpack the interaction, we carried out two one-way repeated measures ANOVAs. Table 7 presents the means and standard deviations for low-/high-talkative and 5-, 8-, and 12-year-old children.

Table 7

Means and Standard Deviations for the Effect of Talkativeness on Amount of Information

		Low-Talkative	High-Talkative
Child Age	*n*	*M* (*SD*)	*M* (*SD*)
5-years	257	3.64 (1.56)	5.02 (1.56)
8-years	260	4.27 (1.45)	5.39 (1.25)
12-years	254	4.52 (1.41)	5.59 (1.37)

Effect of talkativeness.

In the first one-way repeated measures ANOVA we examined age effects with the low-talkative children and found a significant effect of age $F(1.93, 489.42) = 31.85$, $p < .001$, $\eta_G^2 = 0.06$, $\eta_p^2 = 0.11$—with a medium effect size. Pairwise comparisons (see Table 8) were then used to unpack the effect of age. These showed significant mean differences between all age pairings for low-talkative children.

In the second one-way repeated measures ANOVA we examined the effect of age for the high-talkative children. Again, we observed a significant effect of age $F(2, 502) = 18.17$, $p < .001$, $\eta_G^2 = 0.03$, $\eta_p^2 = 0.07$—with a small-medium effect size. Pairwise comparisons (see Table 8) were then used to unpack the effect of age. These showed significant mean differences between 5- to 8-year-olds, and 5- to 12-year-olds for high-talkative children—no significant mean difference was found between 8-year-olds and 12-year-olds.

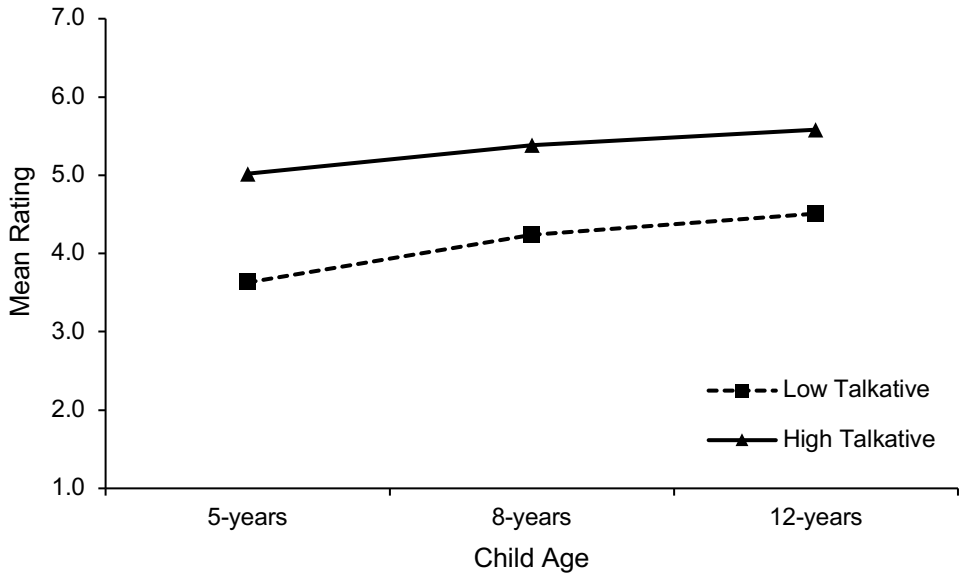

Figure 7. Interaction effect for amount of information ratings for effect of talkativeness.

Table 8

Post-hoc Pairwise Comparisons for Effect of Talkativeness on Amount of Information

	Low-Talkative			High-Talkative		
Comparisons	M_{diff} (*SE*)	p^a	95% *CI*	M_{diff} (*SE*)	p^a	95% *CI*
5-8 years	-0.61 (0.12)	< .001	[-0.89, -0.33]	-0.36 (0.10)	.001	[-0.59, -0.13]
8-12 years	-0.26 (0.10)	.029	[-0.51, -0.02]	-0.19 (0.09)	.086	[-0.41, 0.02]
5-12 years	-0.87 (0.12)	< .001	[-1.16, -0.58]	-0.56 (0.10)	< .001	[-0.79, -0.32]

Note. [a] Adjustment for multiple comparisons using Bonferroni correction

Summary of Study 1 results

Highly talkative children tended to be rated more favourably than less talkative children, although they were rated as less concise and more suggestible than less talkative children. The favourable ratings were evident in statistically significant differences in the constructs of confidence, consistency, informativeness, and memory. Older children tended to be rated more favourable than younger children in terms of accuracy, believability, coherency, consistency, credibility, honesty, informativeness, memory, and reliability. The 5-year old children were rated as less talkative than the older children. Younger children were rated as less concise and less confident than 12-year olds.

When examining the construct of amount of information, ratings depended on the talkativeness and the child age. For instance, when children were low in talkativeness, ratings of amount of information significantly increased with age. Although when the children were highly talkative then the 5-year-old children were rated less favourably than older (8- and 12-year-old) children.

Discussion Study 1

Does level of talkativeness influence juror perceptions of a child witness?

Recall that the main goal of Study 1 was to see if level of talkativeness affected witness credibility ratings of 5-, 8-, and 12-year old children. We predicted that high-talkative children would receive more favourable ratings than low-talkative children, and this was mostly the case—except participants rated high-talkative children less favourably in terms of conciseness and suggestibility than low-talkative children. If we think about talkativeness as a measure of amount of testimony then this finding is consistent with other research. Jurors are influenced by the number of details provided in testimony of adults (e.g., Bell & Loftus, 1988; Bell & Loftus, 1989; Heath, Grannemann, Sawa, & Hodge, 1997; Jones, 1997) and

reporting more details predicts more favourable child witness credibility ratings (Henry et al., 2011).

The present research is unique in that it investigates the relationship between the quantity of testimony and juror perceptions by experimenting with testimony word count, for three different age groups, whilst controlling for the testimony content (number of health-check components recalled and specific details about each health-check-component). This approach provides preliminary evidence to support the role of talkativeness (not the amount of information recalled) as a unique factor that influences juror perceptions of child witness credibility.

The influence of talkativeness on juror perceptions has implications for the criminal justice system during the forensic interview stage and when testimony is evaluated by jurors. Talkativeness first becomes implicated when child witnesses are interviewed about what they remember about an event. How much children say links into two key objectives of interviewing; building rapport and eliciting as much information as possible. Rapport building helps children understand what to expect during an interview and helps them feel more comfortable about talking and sharing information (Brown & Pipe, 2003). In other words, building rapport helps children become more talkative in preparation for the main part of the interview (Lamb et al., 2018). Thus, it is during the interview that there is an opportunity to increase talkativeness. Talkativeness then becomes implicated with how jurors perceive child witnesses in terms of credibility.

Children are generally perceived as less reliable and less competent as witnesses compared to adults (Brainerd & Reyna, 2012; Wright et al., 2010), and when the child is not very talkative then their reliability and credibility is further compromised. For instance, low-talkative children in this study reported the same aspects of the health-check as the high-talkative children yet they were still rated less favourably. Any unjust perceptions of less

talkative children is especially concerning considering that maltreated children have poorer language skills than children who have not been maltreated (Lum et al., 2015) and that shy children also tend to have poorer language skills compared to less shy children (Coplan & Evans, 2009; Smith Watts et al., 2014; Spere et al., 2009). Based on our current findings—that jurors rated low-talkative children as less confident, having poorer memory, being less informative, and providing less information than high-talkative children—we need to be mindful of children who are shy or have been maltreated because low talkativeness appears to give jurors reason to perceive children as less credible, which may not necessarily be justified. Therefore, high-talkative child witnesses may be seen as more credible because they show a greater capacity for sharing information, and less talkative child witnesses may be seen as less credible because they show a diminished capacity for sharing information.

Jurors were affected by talkativeness of children, rating high-talkative children more favourably than low-talkative children. Further research is also needed to specifically investigate any relationships between talkativeness and shyness, language skill, and communicative competence. Research also needs to see if there is a relationship between maltreatment and talkativeness of children. If children who say more are perceived as more credible, can we encourage children to be elaborative without compromising accuracy and without compromising their perceived credibility?

Age Effects

As predicted, and consistent with previous research, older children received more favourable ratings than younger children did (e.g., Newcombe & Bransgrove, 2007; Talwar et al., 2006; Wright et al., 2010). While this finding is unsurprising it is only backed up by compiling findings across many different studies that focus on many different child ages (and not all ages in one study). Therefore, future juror perception research needs to investigate age effects on juror perceptions with research that covers the entire age range of child

witnesses—from pre-schoolers right through to young adulthood. Including the entire age range of child witnesses, will allow researchers to become sensitive to any developmental changes and how jurors may or may not adjust their perceptions accordingly. Furthermore, we can ensure that juror perceptions about the age of the witness are not made at the expense of other child witness characteristics, such as ensuring that when a younger child witness provides detailed testimony they are not dismissed as less credible just because of their age.

Although we expected these age effects based on past research, in reality they should not have occurred, given that the amount of information in the transcripts was held constant and children provided the same fundamental information regardless of age.

Interaction Effect

We also found an unpredicted interaction effect for ratings of the amount of information children provided, where favourable evaluations were dependant on both the talkativeness and the age of the child. Recall that when the child was low-talkative we saw significant differences between all age pairings, but when the child was high-talkative we only saw significant age pairings between 5- and 8-year-olds, and 5- and 12-year-olds. So, when the child was low-talkative participants appeared to place more weighting on the age of the child for evaluating the amount of information. However, jurors should not have judged transcripts differently based on the amount of information because the information included in the transcripts was held constant regardless of the level of talkativeness or child age. Participants may have been sensitive to the additional, non-critical, information present in the transcripts. So, participants may have been picking up on a potential confound—that minor details were inevitably more present for the high-talkative and older children. Because the transcripts were composites from real interviews with children it was difficult to avoid this confound and have high-talkative transcripts that were not just filled with simple filler words and repetitions. An alternative explanation for the interaction effect is that jurors might have

been influenced by their personal beliefs, biases, and stereotypes when they rated the different transcripts.

Participants may also have just been picking up on the fact that high-talkative transcripts were longer (visually) than the low-talkative transcripts and then made the assumption that longer transcripts must contain more information. If participants were assuming that longer transcripts meant more information then it raises the concern that the participants may not have been attending to the data in front of them. For example, they might think that a child who said more must have remembered or known more, therefore they are providing more evidence and are more credible—when in reality this may not be true. Then when children were low-talkative they may have fallen back on their preconceived ideas about older children being more competent witnesses than younger children. This would be a problem in a courtroom because jurors might bring in their preconceived ideas and may not be fully attending to the evidence instead being influenced by their own beliefs.

However, this interaction effect returned a small effect size, and we can only tentatively consider whether these results provide sufficient evidence to apply to a judicial context. Further research about the type of information that influences juror perceptions, as well as jurors beliefs about how much information children of different ages should provide, would add an important piece to the puzzle about what affects juror perceptions, and why. Further explanation about the interaction effect and how this relates to the findings from Study 2 is presented in more detail in the general discussion.

So in summary, yes, talkativeness does have an effect on juror perception—despite the fact that we controlled the content of the testimony—*but,* it depends on the witness credibility construct of interest, with some witness credibility constructs being more influenced by talkativeness than others. Talkativeness significantly contributes to juror perceptions of child witnesses but the effect is not as strong as the effect that child age has on

juror perceptions of child witnesses. Having seen the influence of talkativeness on juror perceptions of child witness credibility, now we want to see how accurate those perceptions are by examining actual associations between child talkativeness and the quality and quantity of their testimony.

Method: Study 2

Participants

Study 2 used the data from the interviews with 5-, 8-, and 12- year old children who recalled a staged event two days later. A total of $N = 116$ participating children were included ($n = 36$ five-year-olds, $n = 39$ eight-year-olds, and $n = 41$ twelve-year-olds).

Materials

Only the free recall and cued recall sections of memory interviews were included in analysis. The beginning of the free recall section was signalled by the interviewer saying "I heard that a few days ago a nurse came to the school and gave you a check-up. Now I wasn't there but I would like to know all about it. Tell me everything you can remember from the beginning to the end". The free recall section continued with open-ended questions from the interviewer (e.g., "Tell me some more things that you can remember about the check-up") until the child indicated on at least two separate occasions that they could remember nothing else—or if the child was only offering information that was peripheral to the health-check itself (e.g., colour of the chairs) despite additional prompting with open-ended questions.

The cued recall section was signalled by the interviewer moving to asking for more information about something the child had already talked about (e.g., "Earlier you mentioned that the nurse checked your blood pressure, tell me everything you can remember about nurse checking your blood pressure"). This continued with the interviewer following up each component the child recalled from the health-check (and any false memories about the health-check). The end of the cued recall section was signalled by the interviewer saying "I

have some more questions for you now about when the nurse came to school and gave you a check-up, now you might have already told me about some of these things…".

Procedure

Free recall and cued recall sections of each memory interview with a participant were pasted into a Microsoft Excel spreadsheet to separate the interviewer utterances from the child utterances. An utterance refers to one turn in conversation. A word count formula was then used to calculate the number of words per child utterance and then calculate free recall word count, cued recall word count, and a total word count. The excel formula used to calculate word counts was: =IF(LEN(TRIM(**CELL**))=0,0,LEN(TRIM(**CELL**))-LEN(SUBSTITUTE(**CELL**," ",""))+1). This formula counts the number of words by summing the number of spaces between words, then adding one. It also returns the number zero if there are no words in the cell, and strips out any extra spaces between words that may have accidentally found their way into the text. This formula was published by Dave Bruns (n.d.) on the webpage "https://exceljet.net/formula/count-total-words-in-a-cell".

Number of spoken words, and Mean Length of Utterance words (MLUw) were calculated from the transcripts as well as accuracy scores for their memory of the health-check event. MLUw is simply the average number of words per turn of conversation. This measure is an established indicator of children's language development (Parker & Brorson, 2016). MLUw scores were calculated for free recall, cued recall, and a total MLUw score was also calculated. Accuracy scores, number of details, and number of components recalled were collated for each participant according to The MARCIE Project coding scheme (see Appendix F: The MARCIE Project Coding Scheme). Accuracy scores were calculated using the formula "correct details /total details" and expressed as a percentage.

Table 9 provides a summary of the codes included in the coding scheme for the MARCIE Project as well as their descriptions and operational examples (see Appendix F:

The MARCIE Project Coding Scheme for more details about how the transcripts were coded). Responses made by children that were off-topic, false-starts, repeated information, or unable to be verified as correct or incorrect, were not included in the calculation of accuracy scores—but were included in calculation of word counts, and mean length of utterances.

Establishing Kappa Reliability for Coding of Memory Interviews

18% of interview transcripts were double coded by independent research assistants, blind to the hypotheses of the current study. Inter-rater reliability was assessed using Cohen's kappa (κ). A score of $\kappa = .888$, $p < .001$ was obtained between coder 1 and coder 2, and a score of $\kappa = .903$, $p < .001$ was obtained between coder 2 and coder 3—indicating almost perfect reliability with each other (Viera & Garrett, 2005).

Table 9

Coding Scheme Codes and Descriptions with Examples

Code Description	Code Definition	Example
Core Correct/Action	Correctly recalled nurses action(s)	"Took my temperature"
Core Correct/Body Part	Correctly recalled part(s) of their body involved	"Put the thing in my ear"
Core Correct/Equipment	Correctly recalled equipment used	"Nurse used a thermometer"
Core Distortion/Action	Memory error relating to nurses action(s)	"Cleaned my ear with beep thing"
Core Distortion/Body Part	Memory error relating to their body parts involved	"Put thermometer under my tongue"
Core Distortion/Equipment	Memory error relating to equipment used	"Checked my temperature with the torch"
Elaborative Correct	Extra information verified as correct but not eligible for a Core Correct code	"Thermometer was blue and white"
Elaborative Incorrect	Minor memory error relating to health-check	"Thermometer was pink"
Elaborative Other	Extra information that was unable to be verified	"It tickled"
False Memory/Critical Distractor	Major memory error where child recalls the nurse listening to their chest, tummy, or back with a stethoscope	"Listened to my chest"
False Memory/Related	Major memory error where child recalls a health-check related component that did not occur—but is not a critical distractor	"Measured my head"
False Memory/Unrelated	Major memory error where child recalls a part of health-check that is not associated with a health-check	"The nurse sang a song"
False Memory Elaborative	Extra information provided about a false memory	"Stethoscope was cold on my chest"
Off Topic	Information not part of health-check event	"We were doing writing when the nurse came"
No Code	Information that is unable to be assigned a code, usually for ambiguous utterances	"We did the thing that umm"
Repeated Information	Information already provided by the child	"Took my temperature"

Note: The health-check with the nurse was scripted and included the following components: Temperature, Height, Weight, Vision, Peak-Flow, Mouth/Throat, Ears, Eyes, Neck, Blood Pressure, and Gift/Prize. Recollections about the nurse were also coded including her gender, name, and details about her appearance. Due to the purpose of the original study the components "listening to chest/back/tummy" was purposefully excluded and recollection of these components was coded as a False Memory/Critical Distractor.

Study 2 Results

Recall that study 2 examined whether there were associations between how talkative children were (measured by word count and mean length of utterance—MLUw), and how informative (number of components and number of detailed reported) and accurate (proportion correct) they were. Analyses of accuracy scores were conducted using arcsine transformed data (e.g., Brown et al., 2013), although raw data is presented in the tables and figures below. Table 10 shows the mean, standard deviations, 95% confidence intervals, and range for the measures extracted from 116 transcripts of interviews with 5-, 8-, and 12-year-old children.

Age Differences

A series of five one-way ANOVAs were run to test for age differences in word count, MLUw, accuracy scores, number of components recalled, and number of details recalled. A total of six outliers were identified and replaced using the process of winsorizing. Violations to the assumption of normality were ignored because ANOVA is considered robust to violations of normality (Field, 2013). When violations to the assumption of homogeneity were identified Welch's *F* is used to interpret ANOVA results and Games-Howell correction is used to interpret any post-hoc analyses. Post-hoc analyses are conducted using pairwise comparisons to control for the familywise error rate, where the level of significance is corrected for each test. Assessment of the one-way ANOVA assumptions are presented in Appendix D: Assumption Checks of Data.

Table 10

Descriptive Statistics of Measures of Children's Talkativeness, Details and Components Recalled, and Accuracy during an Interview about a Staged Event

Measure	*n*	*M* (*SD*)	95% *CI*	Range
Word count	116	715.44 (443.89)	[633.80, 797.08]	40–2062
5-years	36	357.92 (194.51)	[292.10, 423.73]	40–860
8-years	39	720.46 (299.10)	[623.51, 817.42]	273–1579
12-years	41	1024.59 (485.16)	[871.45, 1177.72]	285–2062
MLUw	116	19.70 (8.69)	[18.10, 21.30]	4.44–44.22
5-years	36	13.92 (6.58)	[11.69, 16.14]	4.44–29.00
8-years	39	23.65 (9.70)	[20.51, 26.79]	7.18–44.22
12-years	41	21.02 (6.53)	[18.96, 23.08]	6.20–34.00
Accuracy	116	0.91 (0.09)	[0.90, 0.93]	0.52–1.00
5-years	36	0.86 (0.13)	[0.82, 0.91]	0.52–1.00
8-years	39	0.93 (0.07)	[0.91, 0.95]	0.74–1.00
12-years	41	0.95 (0.03)	[0.94, 0.96]	0.88–1.00
Components Recalled	116	8.60 (2.32)	[8.18, 9.03]	3–13
5-years	36	6.64 (2.26)	[5.88, 7.40]	3–11
8-years	39	9.13 (1.69)	[8.58, 9.68]	5–12
12-years	41	9.83 (1.75)	[9.28, 10.38]	6–13
Details Recalled	116	40.17 (21.38)	[36.24, 44.11]	5-95
5-years	36	20.58 (11.39)	[16.73, 24.44]	5-48
8-years	39	41.82 (15.80)	[36.70, 46.94]	8-95
12-years	41	55.80 (19.11)	[49.77, 61.84]	21-94

Word count (Welch's $F(2, 70.60) = 42.89$, $p < .001$, $\eta_p^2 = 0.38$), *MLUw* (Welch's $F(2, 73.21) = 17.02$, $p < .001$, $\eta_p^2 = 0.22$), *accuracy* (Welch's $F(2, 66.98) = 4.38$, $p = .016$, $\eta_p^2 = 0.09$), *components recalled* ($F(2, 115) = 29.23$, $p < .001$, $\eta_p^2 = 0.34$), and *details recalled* (Welch's $F(2, 74.29) = 56.27$, $p < .001$, $\eta_p^2 = 0.47$) all showed statistically significant differences between the age groups—with large effect sizes for word count, MLUw, number

of components recalled, and number of details recalled, and a medium effect size for accuracy.

Post hoc analysis.

Games-Howell post hoc analysis revealed that *word counts* significantly increased with age—5-years to 8-years (a mean increase of 362.54, 95% CI [223.87, 501.22], $p < .001$); 5-years to 12-years (a mean difference of 666.67, 95% CI [468.05, 865.29], $p < .001$); and 8-years to 12-years (mean difference of 304.12, 95% CI [89.28, 518.97], $p < .001$). Games-Howell post hoc analysis of *MLUw* scores revealed that 5 year-olds had significantly lower MLUw scores than both groups of older children (with a mean difference of 9.73, 95% CI [5.18, 14.29], $p < .001$ with 8-year-olds and a mean difference of 7.10, 95% CI [3.52, 10.68], $p < .001$ with 12-year-olds). Games-Howell post hoc analysis revealed that while *accuracy* scores increased with age this difference was only significant between 5-year-olds olds and 12-year-olds olds (a mean difference of 0.17, 95% CI [0.03, 0.31], $p = .013$). Tukey post hoc analysis of *components recalled* revealed that 5 year-olds recalled significantly fewer health-check components than older children, (a mean difference of 2.49, 95% CI [1.45, 3.53], $p < .001$ with 8-year-olds and 3.19, 95% CI [2.16, 4.22], $p < .001$ with 12-year-olds), but there was no significant difference between 8-year-olds and 12-year-olds. Games-Howell post hoc analysis revealed that the *number of details* significantly increased with age and this was significant for all child age pairings—5-years to 8-years (a mean difference of 20.60, 95% CI [13.57, 27.62], $p < .001$); 8-years to 12-years (a mean difference of 14.63, 95% CI [5.703, 23.55], $p = .001$); and 5-years to 12-years (a mean difference of 35.22, 95% CI [26.74, 43.70], $p < .001$).

Prediction Models

Due to these age differences reported above—having an effect on the measures of word count, MLUw, accuracy, and number of components recalled—hierarchical multiple

regression analyses to predict accuracy and number of components recalled based on talkativeness measures also included child age in the models as an independent variable.

We ran two hierarchical multiple regressions to ascertain; 1) whether the inclusion of word count and then of MLUw improved the prediction of memory accuracy over and above child age alone, and 2) whether the inclusion of word count and then of MLUw improved the prediction of number of components of health-check recalled over and above child age alone. See Table 12 and Table 13 for full details on each regression model below. As outlined in Appendix D: Assumption Checks of Data, all assumptions were met for the first hierarchical multiple regression, and one outlier was initially identified in the hierarchical multiple regression predicting number of components recalled but was removed and the test was re-run with all assumptions being met. The number of details measure violated the assumption of multicollinearity and while it is included in the summary of correlations (see Table 11), it was excluded from the hierarchical multiple regression models to avoid ambiguity about which variables contributed to the total variance explained in the models.

Before analysing the results of the hierarchical multiple regressions, results from Pearson's product moment correlations were reviewed. Table 11 shows positive correlations between all the measures. Age was statistically significantly correlated with all the other measures. Word count, MLUw, number of components, and number of details were all statistically significantly correlated with all measures except accuracy. Strong positive correlations ($r > .7$) were found between number of details and word count, and number of details and number of components. Hierarchical multiple regressions were run without the number of details measure because correlations above .7 are indicative of a violation of the multicollinearity assumption.

Table 11

Summary of Correlations for Child Age and Measures of Quality/Quantity from Memory Interviews

Measure	1	2	3	4	5	6
1. Child Age	—					
2. Word Count	.60**	—				
3. MLUw	.29**	.45**	—			
4. Accuracy	.28**	.06	.09	—		
5. Number of Components	.54**	.60**	.45**	.03	—	
6. Number of Details	.66**	.82**	.45**	.03	.78**	—

Note. ** p < .01 two-tailed

As shown in Table 12 the full model of child age, word count, and MLUw to predict *memory accuracy* (Model 3) was statistically significant, $R^2 = .10$, $F(3, 112) = 4.06$, $p = .009$, adjusted $R^2 = .07$. The model was significant with age alone $F(1, 114) = 9.30$, $p = .003$ and the addition of neither word count $F(1, 113) = 2.26$, $p = .135$, nor MLUw $F(1, 112) = 0.57$, $p = .452$, to prediction of accuracy lead to a significant increase in R^2. For each unit of word count (one word), when other variables are held constant, we saw no significant changes in accuracy ($p = .097$).

Table 12

Hierarchical Multiple Regression Predicting Memory Accuracy from Age, Word Count, and MLUw

	Memory Accuracy					
	Model 1		Model 2		Model 3	
Variable	*B*	β	*B*	β	*B*	β
Constant	1.020**		1.010**		0.984**	
Child Age	0.00*	.275	0.003*	.376	0.003*	.373
Word Count			0.000	-.168	0.000	-.201
MLUw					0.002	.076
R^2	.075		.094		.098	
F	9.30*		5.83*		4.06*	
ΔR^2	.075		.018		.005	
ΔF	9.30*		2.26		0.57	

Note. N = 116. * $p < .05$, ** $p < .001$.

As shown in Table 13 the full model of child age, word count, and MLUw to predict *number of components recalled* (Model 3) was statistically significant, $R^2 = .50$, $F(3, 111) = 36.69$, $p < .001$, adjusted $R^2 = .48$. The model was significant with age alone $F(1, 113) = 52.37$, $p < .001$ and the addition of word count to prediction of number of components recalled led to a significant increase in R^2 of .15, $F(1, 112) = 32.24$, $p < .001$. Further addition of MLUw to predict number of components recalled led to significant increase in R^2 of .03, $F(1, 111) = 6.29$, $p = .014$. For each unit of word count (one word), when other variables are held constant, we saw a significant increase of 0.002 health-check components recalled, ($p < .001$)—in other words, for every 500 words spoken, children recall an additional component of the health-check.

Table 13

Hierarchical Multiple Regression Predicting Number of Components Recalled from Age, Word Count, and MLUw

	Testimonial Quantity					
	Model 1		Model 2		Model 3	
Variable	*B*	β	*B*	β	*B*	β
Constant	4.516**		4.767**		4.140**	
Child Age	0.038**	.563	0.019*	.274	0.018*	.270
Word Count			0.003**	.486	0.002*	.401
MLUw					0.050*	.190
R^2	.317		.469		.498	
F	52.37**		49.54**		36.69**	
ΔR^2	.317		.153		.028	
ΔF	52.37**		32.24**		6.29*	

Note. N = 115. * $p < .05$, ** $p < .001$.

Summary of Study 2 Results

Word counts, MLUw, accuracy, number of health-check components recalled, and number of details recalled all revealed significant effects of child age. Word count and number of details recalled both significantly increased as children got older. 5-year-olds had significantly lower MLUw scores than older children. Accuracy only showed a significant increase between 5-year-olds and 12-year-olds (but not between 5-year olds and 8-year-olds or between 8-year-olds and 12-year-olds). 5-year-olds also recalled fewer components of the health-check than older children.

Statistically significant correlations were found between many of the measures and all measures were positively correlated with each other. While all measures showed a positive relationship with age, the other measures also significantly increased with each other—but associations with the accuracy measure were not statistically significant.

When predicting accuracy of memory reports, age alone explained only 7.5% of the variance. The addition of neither word count nor MLUw significantly added to the percent of memory accuracy variance explained over and above child age. On the other hand, when

predicting number of components recalled about the health-check, age alone explained 31.7% of the variance and the addition of word count and MLUw measures both independently significantly added to the amount of variance explained—from 31.7% to 46.9% and 49.8% respectively, although the likely practical impact on the amount of information that children reported was minimal.

Discussion Study 2

Does talkativeness of children influence the quality of their memory interviews?

We predicted that neither talkativeness nor child age would have an effect on the quality of children's memory interviews. As expected, talkativeness was not associated with, nor did it contribute to, the quality of children's memory reports in terms of accuracy. In other words, talkativeness did not explain the differences that we saw in children's accuracy scores. This finding contradicts research that reports a quantity-accuracy trade-off (e.g., Kulkofsky et al., 2008) but it also contradicts research that suggests a positive relationship between quality and quantity (e.g., Chae, Kulkofsky, Debaran, Wang, & Hart, 2016; McGuigan & Salmon, 2006). We can speculate that differences in the types of events children recalled, the delays employed in the various studies, and the interviewing methods used in different studies may all contribute to different findings. We have much to learn about why accuracy might vary according to quantity of recall in the context of these other factors.

Contrary to our predictions, we found an effect of child age on children's memory interview accuracy scores—but only when comparing accuracy scores between 5-year-olds and the more accurate 12-year-olds (but not for any other age group comparisons). Age only explained 7.5% of the differences we saw in accuracy scores; and so clearly there are other variables that we did not identify contributing to the quality of their accounts. Overall, we found that children's memory interview accuracy scores were high—mean accuracy scores by age are all above 85%—likely reflecting the fact that we focused our analysis on

responses to open-ended questions from the first part of their recall about the health-check. We know that open-ended questions improve the quality and quantity of responses from child witnesses (Brown & Lamb, 2015; Lamb et al., 2018), but we are left wondering about what might actually explain the differences in accuracy for children's recollection of an event.

So, did talkativeness influence the quality of interviews? Because Study 2 only included children's responses to open-ended questions, we expected to see no age effects in the memory interview data. But, despite age predicting accuracy, we found that talkativeness does not account for much of the variation in accuracy scores. Therefore we conclude that talkativeness neither impaired nor improved accuracy of children's memory of the health-check.

Does talkativeness of children influence the quantity of their memory interviews?

We predicted that high-talkative children would provide more information than children who were low-talkative—and indeed, talkativeness was significantly associated with the quantity measures (number of health-check components recalled and total number of details recalled) from the memory interviews. As the word counts of children's memory interviews increased (and to a lesser extent MLUw) so did child age, total number of details recalled, and number of components recalled. Talkativeness also explained a significant amount of the variance that we saw in the number of components that children recalled about the health-check. This was supported by the finding that as memory interview word counts increased by 500 words this led to an additional component of the health-check being recalled. Children said on average 715.44 words during a memory interview, however, meaning they would need to increase their word count by about 70% to generate just one additional detail. So the practical meaningfulness of this finding is limited. While talkativeness does contribute to the number of components recalled it only provides a minimal contribution to explaining why some children recalled more components of the

health-check than others. Therefore from an applied perspective talkativeness is not a productive measure of the amount of fundamental information that children recall.

The association between talkativeness and amount of information connects with the sociocultural language theory of autobiographical memory which highlights the importance of, and exposure to, practising language and narratives in the development of autobiographical memory—particularly through sociocultural contexts (Nelson & Fivush, 2004). For instance preschool children with high-talkative parents recall more information than children from less talkative parents (Peterson, Sales, Rees, & Fivush, 2007), narrative coherence improves with age (Reese et al., 2011), and narrative elaboration training is associated with increased recall of information about an event (Saywitz & Snyder, 1996). So, if children who are exposed to more language by having role models (e.g., parents) who are high-talkative then this points to the idea that children might learn from a sociocultural perspective to be more talkative. Thus children might learn to be more talkative and therefore have more opportunities to develop enhanced language skills—subsequently improving their memory recall. This idea of exposure to language is also reflected in the relationship between shyness and language.

Coplan and Evans (2009) provide four hypotheses that may provide an explanation of the influence of talkativeness on amount of information. These hypotheses attempt to explain the relationship between shyness and language development, but here we tentatively apply them to the current research findings. Firstly, Coplan and Evans (2009) propose the hypothesis that shy (less talkative) children are less likely to use language and therefore because of fewer opportunities to practise have under developed language skills—this seems feasible but supporting this hypothesis is beyond the scope of our research. Secondly, shy children know what they want to say but just do not say it, thus giving the appearance of poorer language skills—this could potentially explain why less talkative children recalled

fewer components of the health-check. Thirdly, children who have poorer language skills leads them to becoming shy, thus the inverse of the first hypothesis—again feasible but supporting this is beyond the scope of our research. Finally, Coplan and Evans (2009) propose that being outgoing provides children with an advantage over shy children, where outgoing children have more practice, thus performing better on measures of language skill—as with the second hypothesis this could certainly explain why more talkative children recalled more health-check components than less talkative children. The fourth hypothesis, the notion of practice makes perfect, also ties into the sociocultural theory of autobiographical language development, the implications of which have already been discussed. While these hypotheses are useful they do not explain the role that memory ability might play with talkativeness. Future research needs to investigate whether there is an association between talkativeness and memory capacity/ability to recall information about an event by using memory tasks and measures of language ability in conjunction with measuring talkativeness and memory for an event.

So in summary, Study two results imply that talkativeness is an indicator of the overall quantity of information provided about an event, but that age is also a significant factor. When it comes to the effect of child age, we expected older children to provide more information in their memory interviews than the younger children—and overall this was what our findings suggested. Similar results with older children providing more information are found in studies with 3-5 year olds (Gagnon & Cyr, 2017), and 4-8 year olds (Lamb et al., 2003). We saw word counts and total number of details significantly increase with child age, but we only saw significant differences between MLUw of 5-year-olds compared to MLUw of older children. 5-year-olds also recalled less components of the health-check than older children (8- and 12-year-olds) did.

Overall the more a child says the more information they are likely to provide, and the older they are likely to be. With age comes vocabulary development (Anglin, 1993), oral language development (Honig, 2007), and cognitive development (Lamb et al., 2018). All of which likely contribute both to talkativeness, as well as to recall and reporting processes that support a more detailed memory account. The development of vocabulary, language, and cognition does not seem to occur at the expense of accuracy, and given the lack of association between talkativeness and accuracy, this suggests that children are not willing to just say anything, but are monitoring the content of their recall to keep it focussed on the event they experienced.

General Discussion

We investigated the effect of talkativeness on juror perceptions of child witness credibility as well as the effect of talkativeness on the quality and quantity of children's recall of a recent event. Below, the combined implications of Study 1 and 2 will be discussed, contrasting jurors' perceptions with evaluations of children's actual eyewitness ability, and applications to the criminal justice system are considered.

Are jurors perceptions of talkativeness in child witnesses justified?

Jurors assessed low- and high-talkative children differently on a number of witness credibility constructs. In Study 2 we were able to test the legitimacy of these perceptions with respect to two important aspects of children's testimony—how much information they reported, and how accurate they were when recounting their experience. When we consider accuracy, which is arguably the most important dimension of testimony, jurors were justified in their assessments—they did not appear to be influenced by how talkative the children were (irrespective of age) and objective measures showed that talkativeness was not associated with children's accuracy when they recounted the staged event.

When we consider the amount of information reported, jurors were again valid in their assessments—they judged highly talkative children as providing more informative than less talkative children (particularly for the 5- and 12-year-old children) and our analysis of children's testimony about a staged event showed that talkativeness did indeed predict how many components children recalled. Jurors may be justified in their belief that high-talkative children provide more information (as this is what the results from Study 2 suggest) but they made these judgements on transcripts where number of components recalled, and number of key details recalled about the event were held constant. As such, although their perceptions aligned with what we observed from our objective measures, the jurors ought not to have arrived at this conclusion (because the transcripts *did not* contain more information as a function of talkativeness). We now elaborate on two possible explanations for why jurors evaluated the transcripts differently.

First, jurors may have been influenced by the inclusion of non-critical details and/or superfluous content that was used to make the high-talkative transcripts lengthier. If this was the case, it would indicate that jurors are sensitive to very subtle cues in children's narratives, at least when they are presented in written form. Therefore future research could investigate the influence that different types of details have on juror perceptions whilst still manipulating the level of talkativeness. Examples of different types of details could include emotive details (e.g., "It made me feel sad"), opinions (e.g., "nurse was nice"), and non-critical details (e.g., "the chair had four legs").

Second, jurors may have brought pre-existing beliefs to the task, and their positive evaluations of the amount of information children were reported therefore reflected this bias. This theory could be tested by assessing juror's general beliefs about children's testimony and memory capacity prior to making the ratings, and determining the extent to which they were associated. For jurors assessing witnesses, this might mean that instead of carefully

reviewing the evidence being presented to them in the form of child witness testimony, they make their evaluations of the child witness and the child witness testimony based on preconceived ideas that they had before the court case. Fortunately the Study 2 results supports jurors' assumption that high-talkative children provide more information than low-talkative children, but this is not a particularly helpful assumption to have considering the sheer amount of additional words that were required to make a meaningful contribution to children's recall about the health-check. Therefore although talkativeness is an indicator of amount of information recalled, in reality high-talkativeness does not provide sufficient explanation for providing more fundamental information than low-talkative children—it does not follow through into actual witness behaviour in a meaningful way.

Child witnesses will not always adhere to juror expectations. The attribution theory (Kelley, 1972) provides an explanation for how violations to expectations can influence overall perceptions. Attribution theory proposes two effects that account for how perceptions can be influenced, the assimilation effect and the contrast effect. The assimilation effect explains how perceptions can adapt in line with people's stereotypes and beliefs. The contrast effect explains how when behaviour is unexpected it will have a larger influence on people's beliefs and perceptions of others. When attribution theory is used as an explanation for the influence of language on persuasiveness of messages it is known the expectancy theory (Burgoon, 1990). For instance, a low-talkative 12-year-old might not meet juror expectations and thus give the impression that they are not a credible witness (contrast effect), which may not be a fair evaluation. Therefore, attribution/expectancy theory could provide a possible explanation for the Study 1 results—particularly the interaction effect that talkativeness and child age had on juror perceptions.

Study 1 results showed that for many of the witness credibility constructs the average ratings were around five out of seven, which suggests that juror perceptions of the child

witnesses were only mildly favourable. So when we see significant differences between ratings of low-talkative and high-talkative children the ratings are still grounded in the mildly favourable part of the rating scale. Unfortunately, our results suggest that less talkative children will be at a disadvantage in terms of how their testimony is evaluated by a jury. Even when low-talkative children provided the same amount of fundamental information as high-talkative children they were rated as less informative (although no less accurate). When we consider the many other dimensions of testimony that jurors also rated less talkative children negatively on in Study 1 (confidence, consistency, memory) there is clearly a danger of jurors unfairly evaluating less talkative child witnesses more generally. On the flip side our research suggests that if we can increase how talkative children are in forensic interviews then this is likely to have a positive effect on how they are perceived by jurors. Fortunately, there are already evidence based practises available that have been shown to increase the amount children say and in turn the amount of information that children provide. Evidence based practises include rapport-building and the use of open-ended questions (Lamb et al., 2018), which are part of evidence-based interview protocols, such as the NICHD protocol.

The results of this research provide another reason for the importance of preparing and supporting children to provide elaborative descriptions of their experiences. Any additional information that children provide will not only benefit the investigation, but encouraging children to be more talkative may also enhance juror perceptions of them in the courtroom.

Age Effects

Jurors thought that both groups of older children were significantly more accurate than the youngest children were—even though all the children in the transcripts accurately recalled the same components of the check-up. These perceptions are only somewhat justified by the findings from Study 2. Memory interview data showed that 12-year-olds were more

accurate than 5-year-olds were, but the 8-year-olds did not differ from either of these two groups. Age was significantly correlated with accuracy, but only explained 7.5% of the variation in children's accuracy scores from the memory interviews. As mentioned earlier, accuracy scores were relatively high, regardless of child age. So from this standpoint there is not sufficient evidence from our research to suggest that jurors can realistically rate older or younger children as more or less accurate, because the accuracy scores were relatively high across the full age range used in the research. As we proposed as an explanation about the interaction effect, jurors might be judging the transcripts based on their personal beliefs about child witnesses without fully attending to the actual content of the transcripts—and we may be seeing the effects of a contrast/assimilation effect.

Also, the findings from eyewitness studies regarding age and accuracy are inconsistent, where the full spectrum of associations have been reported (e.g., positive associations, negative associations, and no associations). The lack of consistency in the accuracy research implies that there are many different factors that do and do not have an age effect on accuracy of children's recollections. So perhaps jurors have expectations about children that are at odds with reality. Because of the variation in the scientific literature perhaps jurors should not rely on child age as a way to gauge the accuracy of their testimony.

Jurors also thought that older children were more informative/provided more information than younger children, despite this being held constant across the transcripts. The evidence from Study 2 suggests that juror perceptions of older children providing more information did appear to be justified. However, when children were low-talkative, age seemed to carry more weight for evaluating the children—there were significant differences for all age group comparisons. In contrast, when children were more talkative, only the more extreme comparisons came through—5-year-olds compared to older children. This interaction between talkativeness and age was reflected by the Study 2 findings, where 5-year

olds did not recall as many health-check components as the older children. Ornstein et al. (2006) reports similar results where older children recalled more components of an examination with a doctor than younger children (age range for the study was 4- to 7-years). Nevertheless, as children got older they recalled more details about the health-check, and age alone explained 31.7% of the variance when predicting the number of components children recalled in memory interviews—thus supporting juror perceptions of older children providing more information than younger children.

So were jurors perceptions justified? The answer is "yes, but…". In the grand scheme of things, juror perceptions about talkativeness in children were justified when we look at Study 2, but they were not justified in terms of the information available to them to make their evaluations. What the current research does not tell us is whether these juror perceptions from ratings of child witnesses align with pre-existing beliefs about child witnesses because we did not include any belief based questions in our survey (e.g., "Children are capable of accurately recalling an instance of abuse"). We do not know whether jurors based their evaluations on the content of the testimony, the appearance of the testimony (e.g., visual length), their prior beliefs, or a combination of these things. This raises a general concern about how jurors evaluate the testimony of child witnesses, and what they base their evaluations on.

Limitations

Each juror read and rated six different transcripts thus creating the potential for jurors to compare transcripts against previously rated transcripts and influence their perceptions. Comparing witnesses against each other has produced significant effects on juror perceptions in previous research (Newcombe & Bransgrove, 2007; Pickel, 1993). However, the counterbalancing of order used in Study 1 should have mitigated any effect of jurors comparing witnesses in our analysis of the data. Random allocation to the different

counterbalance conditions also controlled for any other order or fatigue effects from reading and rating multiple transcripts. Future research could, however, use between-subjects design to reduce the length and any effects of jurors comparing one transcript against another.

Finally, instead of comparing findings from the memory interviews in Study 2 with juror ratings of these same memory interviews, we compared Study 2 findings with juror ratings of amalgamated transcripts. Getting jurors to evaluate the amalgamated transcripts meant that we compromised the ecological validity of the research. This was necessary for us to control for the accuracy and content of the transcripts. We therefore recommend replicating the research but with jurors rating actual memory interview transcripts that have been coded for accuracy and amount of information.

Looking Ahead

The current research provides a basis for future research to consider the role of talkativeness from a more behavioural perspective. Is there a point when children become too talkative and it becomes detrimental to how they are perceived by jurors? Is there a point where children become too talkative and it becomes detrimental to the accuracy of their testimony?

This also means moving beyond word counts and towards some of the other characteristics associated with verbal behaviour and variations in speech. For instance manipulating verbal behaviours such as speed of speech, intonation, and pitch. Study 1 could be replicated, but with the memory interviews presented as audio recordings rather than written transcripts. Using audio recordings is more ecologically valid and if larger effect sizes are observed then this would indicate that there is more to talkativeness than just the number of words child witnesses say.

Research needs to investigate the mechanisms that might contribute to how talkative a child is. For instance, do talkative children have better memories from an encoding

perspective, or is their recall enhanced? Are talkative children better communicators than untalkative children? What role do other characteristics (e.g., confidence, shyness) play in relation to talkativeness?

Finally, juror perception research need to untangle the mechanisms behind how jurors evaluate child witnesses and child witness testimony. To what extent (and under what circumstances) do jurors attend to the testimony, and how much do jurors rely on their personal beliefs?

Conclusions

The talkativeness of a child witness plays a significant role when jurors assess child witness testimony—where children that are more talkative tend to be rated more favourably on many witness credibility constructs, compared with less talkative children. While talkativeness does not explain variations in accuracy scores (quality) it certainly contributes towards explaining variations of the amount of information (quantity) that children provide in memory interviews. When it comes to child witnesses providing testimony, it is not just what the child says that is important, but how they say it.

References

Akhtar, S., Justice, L. V., Knott, L., Kibowski, F., & Conway, M. A. (2018). The 'common sense' memory belief system and its implications. *The International Journal of Evidence and Proof, 22*, 289–304. https://doi.org/10.1177/1365712718784045

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of The Society for Research in Child Development, 58*. https://doi.org/10.2307/1166112

Areni, C. S., & Sparks, J. R. (2005). Language power and persuasion. *Psychology and Marketing, 22*, 507–525. https://doi.org/10.1002/mar.20071

Bala, N., Ramakrishnan, K., Lindsay, R., & Lee, K. (2005). Judicial assessment of the credibility of child witnesses. *Alberta Law Review, 42*, 995–1017. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4640896/

Bell, B. E., & Loftus, E. F. (1988). Degree of detail of eyewitness testimony and mock juror judgments. *Journal of Applied Social Psychology, 18*, 1171–1192. https://doi.org/10.1111/j.1559-1816.1988.tb01200.x

Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology: Attitudes and Social Cognition, 56*, 669–679. https://doi.org/10.1037/0022-3514.56.5.669

Brainerd, C., & Reyna, V. (2012). Reliability of children's testimony in the era of developmental reversals. *Developmental Review, 32*, 224–267. https://doi.org/10.1016/j.dr.2012.06.008

Brimacombe, E., C.A. , Quinton, N., Nance, N., & Garrioch, L. (1997). Is age irrelevant? Perceptions of young and old adult eyewitnesses. *Law and Human Behavior, 21*, 619–634. https://doi.org/10.1023/A:1024808730667

Brown, D., & Lamb, M. (2015). Can children be useful witnesses? It depends how they are questioned. *Child Development Perspectives, 9*, 250–255. https://doi.org/10.1111/cdep.12142

Brown, D., Lamb, M., Lewis, C., Pipe, M.-E., Orbach, Y., & Wolfman, M. (2013). The NICHD investigative interview protocol: an analogue study. *Journal of Experimental Psychology: Applied, 19*, 367–382. https://doi.org/10.1037/a0035143

Brown, D., & Lewis, C. (2013). Competence is in the eye of the beholder: Perceptions of intellectually disabled child witnesses. *International Journal of Disability, Development and Education, 60*, 3–17. https://doi.org/10.1080/1034912X.2013.757132

Brown, D., & Pipe, M.-E. (2003). Individual differences in children's event memory reports and the narrative elaboration technique. *Journal of Applied Psychology, 88*, 195–206. https://doi.org/10.1037/0021-9010.88.2.195

Bruns, D. (n.d.). *Count total words in a cell.* Retrieved from https://exceljet.net/formula/count-total-words-in-a-cell

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5. https://doi.org/10.1177/1745691610393980

Burgoon, M. (1990). Language and social influence. In H. Giles & W. P. Robinson (Eds.), *Handbook of Language and Social Psychology.* (pp. 51–72). Oxford, England: John Wiley & Sons.

Cashmore, J., & Bussey, K. (1996). Judicial perceptions of child witness competence. *Law and Human Behavior, 20*, 313–334. https://doi.org/10.1007/BF01499026

Chae, Y., & Ceci, S. (2005). Individual differences in children's recall and suggestibility: the effect of intelligence, temperament, and self-perceptions. *Applied Cognitive Psychology, 19*, 383–407. https://doi.org/10.1002/acp.1094

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 and the Law, 34*, 55–73. https://doi.org/10.1002/bsl.2242

Chandler, J. J., & Paolacci, G. (2017). Lie for a dime: When most prescreening responses are honest but most study participants are impostors. *Social Psychological and Personality Science,* 8, 500–508. https://doi.org/10.1177/1948550617698203

Connolly, D. A., Price, H. L., Lavoie, J. 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, 92–112. https://doi.org/10.1007/sl0979-006-9083-3

Conway, M. A., Justice, L. V., & Morrison, C. M. (2014). Beliefs about autobiographical memory and why they matter. *Psychologist, 27*, 502–505. Retrieved from http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=0&sid=712e99ce-82ba-492b-96d3-07261d73ab91%40pdc-v-sessmgr01

Cooper, A., Quas, J. A., & Cleveland, K. C. (2014). The emotional child witness: Effects on juror decision-making. *Behavioral Sciences and the Law, 32*, 813–828. https://doi.org/10.1002/bsl.2153

Coplan, R. J., & Evans, M. A. (2009). At a loss for words? Introduction to the special issue on shyness and language in childhood. *Infant and Child Development, 18*, 211–215. https://doi.org/10.1002/icd.620

Crozier, W. R., & Perkins, P. (2002). Shyness as a factor when assessing children. *Educational Psychology in Practice, 18*, 239–244. https://doi.org/10.1080/0266736022000010267

Eskritt, M., Whalen, J., & Lee, K. (2008). Preschoolers can recognize violations of the Gricean maxims. *British Journal of Developmental Psychology, 26*, 435–443. https://doi.org/10.1348/026151007X253260

Evans, M. A. (1996). Reticent primary grade children and their more talkative peers: Verbal, nonverbal, and self-concept characteristics. *Journal of Educational Psychology, 88*, 739–749. https://doi.org/10.1037/0022-0663.88.4.739

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., . . . Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i–185.

Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics: and sex and drugs and rock 'n' roll* (4th ed.). Los Angeles: Sage.

Fivush, R. (2011). The development of autobiographical memory. *Annual Review of Psychology, 62*, 559–582. https://doi.org/10.1146/annurev.psych.121208.131702

Freeman, G. H. (1979). Complete latin squares and related experimental designs. *Journal of the Royal Statistical Society. Series B (Methodological), 41*, 253–262. https://doi.org/10.1111/j.2517-6161.1979.tb01080.x

Gagnon, K., & Cyr, M. (2017). Sexual abuse and preschoolers: Forensic details in regard of question types. *Child Abuse and Neglect, 67*, 109–118. https://doi.org/10.1016/j.chiabu.2017.02.022

Gongola, J., Scurich, N., & Quas, J. A. (2017). Detecting deception in children: A meta-analysis. *Law and Human Behavior, 41*, 44–54.

Goodman, G.S., Bottoms, B.L., Herscovici, B.B., Shaver, P. (1989) Determinants of the Child Victim's Perceived Credibility. In Ceci S.J., Ross D.F., Toglia M.P. (Eds.), *Perspectives on children's testimony* (pp. 1–22). New York, NY: Springer. https://doi.org/10.1007/978-1-4613-8832-6_1

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics: Vol. 3, speech acts* (pp. 41–58). Mahwah, NJ: Seminar Press.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48*, 400–407. https://doi.org/10.3758/s13428-015-0578-z

Hayes, D. P., & Meltzer, L. (1972). Interpersonal judgments based on talkativeness: I fact or artifact? *Sociometry*, 538–561. https://doi.org/10.2307/2786532

Heath, W. P., Grannemann, B. D., Sawa, S. E., & Hodge, K. M. (1997). Effects of detail in eyewitness testimony on decisions by mock jurors. *Journal of Offender Rehabilitation, 25*, 51–71. https://doi.org/10.1300/J076v25n03_04

Henry, L., Ridley, A., Perry, J., & Crane, L. (2011). Perceived credibility and eyewitness testimony of children with intellectual disabilities. *Journal of Intellectual Disability Research, 55*, 385–391. https://doi.org/10.1111/j.1365-2788.2011.01383.x

Honig, A. S. (2007). Oral language development. *Early Child Development and Care, 177*, 581–613. https://doi.org/10.1080/03004430701377482

Hughes, S. M., Mogilski, J. K., & Harrison, M. A. (2013). The perception and parameters of intentional voice manipulation. *Journal of Nonverbal Behavior, 38*, 107–127. https://doi.org/10.1007/s10919-013-0163-z

Jones, M. (1997). Preventing the application of stereotypic biases in the courtroom: The role of detailed testimony. *Journal of Applied Social Psychology, 27*, 1767–1784.

Kelley, H. H. (1972). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior.* (pp. 1–26). Morristown, NJ: General Learning Press.

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*, 21–38. https://doi.org/10.1002/acp.1326

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4, 863*. https://doi.org/10.3389/fpsyg.2013.00863

Lamb, M., & Brown, D. (2006). Conversational apprentices: Helping children become competent informants about their own experiences. *British Journal of Developmental Psychology, 24*, 215–234. https://doi.org/10.1348/026151005X57657

Lamb, M. E., Brown, D., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2018). *Tell me what happened: Structured investigative interviews of child victims and witnesses*. Chichester, UK and Hoboken, NJ: Wiley.

Lamb, M., Orbach, Y., Hershkowitz, I., Esplin, P., & Horowitz, D. (2007). A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: A review of research using the NICHD Investigative Interview Protocol. *Child Abuse and Neglect*, 31, 1201–1231. https://doi.org/10.1016/j.chiabu.2007.03.021

Lamb, M., Sternberg, K., & Esplin, P. (1998). Conducting investigative interviews of alleged sexual abuse victims. *Child Abuse and Neglect, 22*, 813–823. https://doi.org/10.1016/S0145-2134(98)00056-8

Lamb, M., Sternberg, K., Orbach, Y., Esplin, P., 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*, 926–934.

Leippe, M. R., Manion, A. P., & Romanczyk, A. (1992). Eyewitness persuasion: How and how well do fact finders judge the accuracy of adults' and children's memory reports? *Journal of Personality and Social Psychology: Attitudes and Social Cognition, 63*, 181–197. https://doi.org/10.1037/0022-3514.63.2.181

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*, 433–442. https://doi.org/10.3758/s13428-016-0727-z

Lum, J. A. G., Powell, M., Timms, L., & Snow, P. (2015). A meta-analysis of cross sectional studies investigating language in maltreated children. *Journal of Speech, Language, and Hearing Research, 58*, 961–976. https://doi.org/10.1044/2015_JSLHR-L-14-0056

Magnussen, S., Melinder, A., Stridbeck, U., & Raja, A. Q. (2010). Beliefs about factors affecting the reliability of eyewitness testimony: A comparison of judges, jurors and the general public. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 24*, 122–133. https://doi.org/10.1002/acp.1550

Malloy, L. C., & Quas, J. A. (2009). Children's suggestibility: Areas of consensus and controversy. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 267-297). Hoboken, NJ, US: John Wiley & Sons Inc.

McGuigan, F., & Salmon, K. (2006). The influence of talking on showing and telling: adult-child talk and children's verbal and nonverbal event recall. *Applied Cognitive Psychology, 20*, 365–381. https://doi.org/10.1002/acp.1183

Miragoli, S., Camisasca, E., & Di Blasio, P. (2017). Narrative fragmentation in child sexual abuse: The role of age and post-traumatic stress disorder. *Child Abuse and Neglect, 73*, 106–114. https://doi.org/10.1016/j.chiabu.2017.09.028

Mueller-Johnson, K., Toglia, M. P., Sweeney, C. D., & Ceci, S. J. (2007). The perceived credibility of older adults as witnesses and its relation to ageism. *Behavioral Sciences and the Law, 25*, 355–375. https://doi.org/10.1002/bsl.765

Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: a social cultural developmental theory. *Psychological review, 111*, 486–511.

New Zealand Ministry of Justice. (n.d.). *Who can't be a juror.* Retrieved from https://www.justice.govt.nz/courts/jury-service/who-cant-be/

Newcombe, P. A., & Bransgrove, J. (2007). Perceptions of witness credibility: Variations across age. *Journal of Applied Developmental Psychology, 28*, 318–331. Retrieved from https://doi.org/10.1016/j.appdev.2007.04.003

Nigro, G. N., Buckley, M. A., Hill, D. E., & Nelson, J. (1989). When juries "hear" children testify: The effects of eyewitness age and speech style on jurors' perceptions of testimony. In Ceci S.J., Ross D.F., Toglia M.P. (Eds.), *Perspectives on children's testimony* (pp. 57–70). New York, NY: Springer. https://doi.org/10.1007/978-1-4613-8832-6_4

O'Barr, W. M., & Black, D. (1982). Speech styles in the courtroom. In D. Black (Ed.) *Linguistic Evidence: Language, Power, and Strategy in the Courtroom* (pp. 61–91). New York, NY: Academic Press.

Ornstein, P., Baker-Ward, L., Gordon, B., Pelphrey, K., Tyler, C., & Gramzow, E. (2006). The influence of prior knowledge and repeated questioning on children's long-term retention of the details of a pediatric examination. *Developmental Psychology, 42*, 332–344. https://doi.org/10.1037/0012-1649.42.2.332

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development, 76*, 763–782. https://doi.org/10.1111/1467-8624.00498-i1

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*, 184–188. https://doi.org/10.1177/0963721414531598

Parker, M. D., & Brorson, K. (2016). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language, 25*, 365–376. https://doi.org/10.1177/0142723705059114

Peled, M., Iarocci, G., & Connolly, D. (2004). Eyewitness testimony and perceived credibility of youth with mild intellectual disability. *Journal of Intellectual Disability Research, 48*, 699–703. https://doi.org/10.1111/j.1365-2788.2003.00559.x

Peterson, C., Sales, J. M., Rees, M., & Fivush, R. (2007). Parent–child talk and children's memory for stressful events. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 21*, 1057–1075. https://doi.org/10.1002/acp.1314

Pickel, K. L. (1993). Evaluation and integration of eyewitness reports. *Law and Human Behavior, 17*, 569–595. https://doi.org/10.1007/BF01045074

Poole, D. A., & Lamb, M. E. (1998). *Investigative interviews of children: A guide for helping professionals.* Washington, DC, US: American Psychological Association. https://dx.doi.org/10.1037/10301-000

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: evidence for a confidence heuristic. *Journal of Behavioral Decision Making, 17*, 39–57. https://doi.org/10.1002/bdm.460

Quas, J. A., Thompson, W. C., Alison, K., & Stewart, C. (2005). Do jurors "know" what isn't so about child witnesses? *Law and Human Behavior, 29*, 425–456. https://doi.org/10.1007/s10979-005-5523-8

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*, 424–462. https://doi.org/10.1080/15248372.2011.587854

Reynolds, K. P., & Evans, M. A. (2009). Narrative performance and parental scaffolding of shy and nonshy children. *Applied Psycholinguistics, 30*, 363–384. https://doi.org/10.1017/s0142716409090158

Ross, D. F., Dunning, D., Toglia, M. P., & Ceci, S. J. (1989). Age stereotypes, communication modality, and mock jurors' perceptions of the child witness. In Ceci S.J., Ross D.F., Toglia M.P. (Eds.), *Perspectives on children's testimony* (pp. 37–56). New York, NY: Springer. https://doi.org/10.1007/978-1-4613-8832-6_3

Ross, D. F., Dunning, D., Toglia, M. P., & Ceci, S. J. (1990). The child in the eyes of the jury: Assessing mock jurors' perceptions of the child witness. *Law and Human Behavior*, 14, 5–23. https://doi.org/10.1007/BF01055786

Ruva, C. L., & Bryant, J. B. (2004). The impact of age, speech style, and question form on perceptions of witness credibility and trial outcome. *Journal of Applied Social Psychology, 34*, 1919–1944. https://doi.org/10.1111/j.1559-1816.2004.tb02593.x

Saywitz, K. J., & Snyder, L. (1996). Narrative elaboration: Test of a new procedure for interviewing children. *Journal of Consulting and Clinical Psychology, 64*, 1347–1357. https://doi.org/10.1037/0022-006X.64.6.1347

Shuttleworth, M. (2009, May 8, 2009). *Counterbalanced measures design.* Retrieved from https://explorable.com/counterbalanced-measures-design

Smith Watts, A. K., Patel, D., Corley, R. P., Friedman, N. P., Hewitt, J. K., Robinson, J. L., & Rhee, S. H. (2014). Testing alternative hypotheses regarding the association between behavioral inhibition and language development in toddlerhood. *Child development, 85*, 1569–1585.

Spere, K. A., Evans, M. A., Hendry, C. A., & Mansell, J. (2009). Language skills in shy and non-shy preschoolers and the effects of assessment context. *Journal of Child Language, 36*, 53–71. https://doi.org/10.1017/S0305000908008842

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. (2006). Adults' judgments of children's coached reports. *Law and Human Behavior, 30*, 561–570. https://doi.org/10.1007/s10979-006-9038-8

Teoh, Y.-S., & Chang, T.-F. (2018). Comparing the effects of drawing and verbal recall techniques on children's memory accounts. *Scandinavian Journal of Psychology, 59*, 631–633. https://doi.org/10.1111/sjop.12496

Teoh, Y.-S., & Lamb, M. (2013). Interviewer demeanor in forensic interviews of children. *Psychology, Crime and Law, 19*, 145–159. https://doi.org/10.1080/1068316X.2011.614610

United States Courts. (n.d.). *Juror qualifications.* Retrieved from https://www.uscourts.gov/services-forms/jury-service/juror-qualifications

Van Kleeck, A., & Street, R. (1982). Does reticence mean just talking less? Qualitative differences in the language of talkative and reticent preschoolers. *Journal of Psycholinguistic Research, 11*, 609–629. https://doi.org/10.1007/BF01067615

Vazquez, M. D., Delisle, S. S., & Saylor, M. M. (2013). Four- and six-year-olds use pragmatic competence to guide word learning. *Journal of Child Language, 40*, 291–306. https://doi.org/10.1017/S0305000911000420

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Family Medicine, 37*, 360–363.

Westcott, H. L., Davies, G. M., & Clifford, B. R. (1991). Adults' perceptions of children's videotaped truthful and deceptive statements. *Children and Society, 5*, 123–135. https://doi.org/10.1111/j.1099-0860.1991.tb00378.x

Wright, D. B., Hanoteau, F., Parkinson, C., & Tatham, A. (2010). Perceptions about memory reliability and honesty for children of 3 to 18 years old. *Legal and Criminological Psychology, 15*, 195207. https://doi.org/10.1348/135532508X400347

Appendix A: Survey Participant Information

Invitation to Participate: Mturk Advertisement

Pilot Survey.

We are interested in finding out how well people describe events they have experienced. You must be at least 18 years old to participate in this survey. This survey will take approximately 20 minutes to complete.

Study 1.

We are interested in finding out how well people describe events they have experienced. You must be at least 18 years old to participate in this survey. This survey will take approximately 25 minutes to complete.

Pilot Survey Information Sheet

Children's memory for an event they experienced

INFORMATION FOR PARTICIPANTS

Thank you for your interest in this project. Please read this information before deciding whether or not to take part. If you decide to participate, thank you. If you decide not to take part, thank you for considering my request.

Who am I?

My name is Helen Pierce and I am a Masters student at Victoria University of Wellington, New Zealand. This research project is work towards my thesis.

What is the aim of the project?

This project aims to find out about what people think about how well children can describe a recent event.

This research has been approved by the Victoria University of Wellington School of Psychology Human Ethics Committee application # 0000025460.

How can you help?

If you agree to take part you will read some short passages from interviews with children about a recent event and then answer some questions about each passage. The survey will take you approximately 20 minutes to complete.

You will receive payment as a thank you for your participation. Upon completion of the survey you will receive a 'special key' which you will need to complete the HIT.

What will happen to the information you give?

This research is anonymous. This means that nobody, including the researchers will be aware of your identity. By answering it, you are giving consent for us to use your responses in this research. Your answers will remain completely anonymous and unidentifiable. Once you submit the survey, it will be impossible to retract your answer. Please do not include any personal identifiable information in your responses.

The final report will be available from the applied developmental lab website: www.applieddevelopmentallab.com

What will the project produce?

The information from my research will be used in my Master's Thesis as well as academic publications and conferences.

If you have any questions or problems, who can you contact?

If you have any questions, either now or in the future, please feel free to contact either myself or my supervisor:

Student:
Name: Helen Pierce
piercehele@myvuw.ac.nz

Supervisor:
Name: Dr Deirdre Brown
Role: Supervisor
School: School of Psychology
deirdre.brown@vuw.ac.nz

Human Ethics Committee information

If you have any concerns about the ethical conduct of the research you may contact the Victoria University HEC Convenor: Associate Professor Susan Corbett. Email susan.corbett@vuw.ac.nz or telephone +64-4-463 5480.

If you would like to take part: please continue through the survey by clicking the ">>" button below.

If you would not like to take part you can close your browser window and exit the survey.

Pilot Survey Excluded Participant Debrief

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

Children's memory for an event they experienced

Thank you for your interest in this research project and taking the time to participate. You have completed the screening questions and unfortunately you are not eligible to participate in this particular research project.

If you have any questions or problems, who can you contact?

If you have any questions, either now or in the future, please feel free to contact either myself or my supervisor:

Student:
Name: Helen Pierce
piercehele@myvuw.ac.nz

Supervisor:
Name: Dr Deirdre Brown
Role: Supervisor
School: School of Psychology
deirdre.brown@vuw.ac.nz

Human Ethics Committee information

If you have any concerns about the ethical conduct of the research you may contact the Victoria University HEC Convenor: Associate Professor Susan Corbett. Email susan.corbett@vuw.ac.nz or telephone +64-4-463 5480.

Pilot Survey Debrief Information

Children's memory for an event they experienced

DEBRIEFING INFORMATION FOR PARTICIPANTS

Thank you for participating in this research project. Now that you have completed the questionnaire here is some more information about what we are examining in this research.

You were asked to read and rate a number of different passages from memory interviews with children. The passages were taken from memory interviews with three different age groups (5 year olds, 8 year olds, and 12 year olds). For each age group one of the passages you rated was from a less talkative child and one was from a more talkative child.

This research is to investigate whether there is a detectable difference in the amount of information that children say—their talkativeness. We did not share this information with you until now because we wanted to know your thoughts about each excerpt without influencing you in any way.

Knowing more about any potential effects of talkativeness of child witnesses may affect instructions that are given to jurors about child witnesses, how forensic interviews are conducted, and how child witnesses are prepared for testifying in court. Being able to determine if there is a detectable difference in the amount children say will help us to conduct future research about the effect of talkativeness of child witnesses.

Please do not share this information with other potential participants unless they have also completed the questionnaire.

If you have any questions or problems, who can you contact?
If you have any questions, either now or in the future, please feel free to contact either myself or my supervisor:

Student:
Name: Helen Pierce
piercehele@myvuw.ac.nz

Supervisor:
Name: Dr Deirdre Brown
Role: Supervisor
School: School of Psychology
deirdre.brown@vuw.ac.nz

Human Ethics Committee information

If you have any concerns about the ethical conduct of the research you may contact the Victoria University HEC Convenor: Associate Professor Susan Corbett. Email susan.corbett@vuw.ac.nz or telephone +64-4-463 5480.

Study 1 Survey Information Sheet

Children's memory for an event they experienced

INFORMATION FOR PARTICIPANTS

Thank you for your interest in this project. Please read this information before deciding whether or not to take part. If you decide to participate, thank you. If you decide not to take part, thank you for considering my request.

Who am I?

My name is Helen Pierce and I am a Masters student at Victoria University of Wellington, New Zealand. This research project is work towards my thesis.

What is the aim of the project?

This project aims to find out about what people think about how well children can describe a recent event.

This research has been approved by the Victoria University of Wellington School of Psychology Human Ethics Committee application # 0000025460.

How can you help?

If you agree to take part you will read some short passages from interviews with children about a recent event and then answer some questions about each passage. The survey will take you 25 minutes to complete.

You will receive payment as a thank you for your participation. Upon completion of the survey you will receive a 'special key' which you will need to complete the HIT.

What will happen to the information you give?

This research is anonymous. This means that nobody, including the researchers will be aware of your identity. By answering it, you are giving consent for us to use your responses in this research. Your answers will remain completely anonymous and unidentifiable. Once you submit the survey, it will be impossible to retract your answer. Please do not include any personal identifiable information in your responses.

The final report will be available from the applied developmental lab website: www.applieddevelopmentallab.com

What will the project produce?

The information from my research will be used in my Master's Thesis as well as academic publications and conferences.

If you have any questions or problems, who can you contact?

If you have any questions, either now or in the future, please feel free to contact either myself or my supervisor:

Student:
Name: Helen Pierce
piercehele@myvuw.ac.nz

Supervisor:
Name: Dr Deirdre Brown
Role: Supervisor
School: School of Psychology
deirdre.brown@vuw.ac.nz

Human Ethics Committee information

If you have any concerns about the ethical conduct of the research you may contact the Victoria University HEC Convenor: Associate Professor Susan Corbett. Email susan.corbett@vuw.ac.nz or telephone +64-4-463 5480.

If you would like to take part: please continue through the survey by clicking the ">>" button below.

If you would not like to take part you can close your browser window and exit the survey.

Study 1 Survey Excluded Participant Debrief

Children's memory for an event they experienced

Thank you for your interest in this research project and taking the time to participate. You have completed the screening questions and unfortunately you are not eligible to participate in this particular research project.

If you have any questions or problems, who can you contact?

If you have any questions, either now or in the future, please feel free to contact either myself or my supervisor:

Student:
Name: Helen Pierce
helen.pierce@myvuw.ac.nz

Supervisor:
Name: Dr Deirdre Brown
Role: Supervisor
School: School of Psychology
deirdre.brown@vuw.ac.nz

Human Ethics Committee information

If you have any concerns about the ethical conduct of the research you may contact the Victoria University HEC Convenor: Associate Professor Susan Corbett. Email susan.corbett@vuw.ac.nz or telephone +64-4-463 5480.

Study 1 Survey Debrief Information

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

Children's memory for an event they experienced

DEBRIEFING INFORMATION FOR PARTICIPANTS

Thank you for participating in this research project. Now that you have completed the questionnaire here is some more information about what we are examining in this research.

You were asked to read and rate a number of different passages from interviews with children. The passages were taken from interviews with three different age groups (5 year olds, 8 year olds, and 12 year olds). For each age group one of the passages you rated was from a less talkative child and one was from a more talkative child.

This research is investigating people's perceptions about children's memory for an event that they have experienced. The research specifically focuses on whether the amount children say—their talkativeness—has an effect on how credible they are seen as witnesses. Another focus of the research is about the effect that witness age might have on perceptions of witness credibility and whether the influence of talkativeness varies according to the age of the child. We did not share this information with you until now because we wanted to know your thoughts about each excerpt without influencing you in any way.

Knowing more about any potential effects of talkativeness of child witnesses may affect instructions that are given to jurors about child witnesses, how forensic interviews are conducted, and how child witnesses are prepared for testifying in court.

Please do not share this information with other potential participants unless they have also completed the questionnaire.

If you have any questions or problems, who can you contact?
If you have any questions, either now or in the future, please feel free to contact either myself or my supervisor:

Student:
Name: Helen Pierce
piercehele@myvuw.ac.nz

Supervisor:
Name: Dr Deirdre Brown
Role: Supervisor
School: School of Psychology
deirdre.brown@vuw.ac.nz

Human Ethics Committee information

If you have any concerns about the ethical conduct of the research you may contact the Victoria University HEC Convenor: Associate Professor Susan Corbett. Email susan.corbett@vuw.ac.nz or telephone +64-4-463 5480.

Appendix B: Survey Questions

Pilot: Pilot Study Questionnaire

Eligibility and Demographic Questions

Date of birth
Month
Day
Year

Gender
- ○ Female
- ○ Male
- ○ Transgender
- ○ Other _____

Is English your native language?
- ○ Yes
- ○ No
  (please specify your native language) _____

How fluent are you in English? [If '4' or less exclude]

1	2	3	4	5	6	7
not at all fluent						very

What is your main occupation? _____

Do you have children?
- ○ Yes
- ○ No

Do any of the following apply to you?
- Have an intellectual disability
- Work in government/congress/parliament (including federal or state)
- Work in a court of law
- Employed by police/fire/ambulance departments
- Member of armed forces or defence force
- Have been sentenced to imprisonment or charged with a felony

○ Yes (if yes>EXCLUDE)
○ No (continue)

----PAGE BREAK----

How much time do you spend with children of the following ages?

Please rate how much contact (past or present) you have with children of the following ages? *e.g.,, your own children, nieces, nephews, siblings or relatives, through personal contact, through work, through volunteer activities*

	Daily	Weekly	Monthly	Yearly	Hardly ever	None at all
Babies (less than 2 years)						
Toddlers (2-4 years)						
5-6						
7-8						
9-10						
11-12						
13-14						
15-16						
17-18						

Please indicate the context(s) in which you have contact with them (and briefly describe)

- ☐ Home _____
- ☐ Work _____
- ☐ Family _____
- ☐ Volunteer activities _____
- ☐ Other _____
- ☐ No contact with children

Have you or a family member ever been involved in a jury-tried court case?

- ☐ As a defendant
- ☐ As a witness
- ☐ As the complainant
- ☐ As a jury member
- ☐ None of the above

Have you or a family member ever given testimony in court?

- ☐ As a defendant
- ☐ As a witness
- ☐ As the complainant
- ☐ As a jury member
- ☐ None of the above

Have you or a family member ever been involved in a court case involving child witnesses?

- ☐ As a defendant

- ☐ As a witness
- ☐ As the complainant
- ☐ As a jury member
- ☐ None of the above

----PAGE BREAK----

Instructions

Thank you for taking part in our study. The information you provide will be of relevance to our research, and on completion we will credit you for your time.

You will be given a series of passages to read from interviews with a child about a recent event and you will be asked to rate them as a witness. Please respond to all questions, even if you are unsure of your responses. If you have any questions at the conclusion of the study, you can ask Helen Pierce piercehele@myvuw.ac.nz. You are free to withdraw from the study at any time.

You will now read a series of passages from interviews with children about a recent event and you will then be asked to answer some questions about each passage.

The interviews were conducted as part of a separate study where children had a health-check (check-up) with a nurse at their school and two days were interviewed by a trained interviewer to elicit their recall of the event.

----PAGE BREAK----

Please read the following passage then move onto the next page.

[First Excerpt – followed by main bank of questions]

----PAGE BREAK----

For each question, please select the number that most closely corresponds to your views...

1 2 3 4 5 6 7
not at all very

1	How likely is it that the child remembered the check-up as it actually happened? [Score: 1 – 7]	**Memory**
2	How confident was the child when describing the check-up? [Score: 1 – 7]	**Confident**
3	How credible was the child? [Score: 1 – 7]	**Credible**

4	How accurate do you think the child's description of the check-up was? [Score: 1 – 7]	**Accurate**
5	How likely is it that the child provided honest testimony? [Score: 1 – 7]	**Honest**
6	How much did the child say about the check-up? [Score: 1 – 7]	**Amount**
7	How consistent was the information that the child reported? [Score: 1 – 7]	**Consistent**
8	How swayed by interviewer suggestion do you think the child was? [Reverse Score: 7 – 1]	**Suggestible**
9	How talkative was the child? [Score: 1 – 7]	**Talkative**
10	How believable was the child's testimony? [Score: 1 – 7]	**Believable**
11	How informative was the child's description of the check-up? [Score: 1 – 7]	**Informative**
12	How coherent was their testimony? [Score: 1 – 7]	**Coherent**
13	How reliable do you think the information was that the child reported? [Score: 1 – 7]	**Reliable**

How old was the child? _____

----PAGE BREAK----

[Remaining five passages – followed by quick ratings]

----PAGE BREAK----

Please read the following passage then move onto the next page.

----PAGE BREAK----

Quick ratings

On a scale of 0 to 100 how talkative was the child?
(0=not at all talkative, 100 = extremely talkative) _____

On a scale of 0 to 100 how concise was the child?
(0=not at all concise, 100 = extremely concise) _____

On a scale of 0 to 100 how much information did the child provide?
(0=no information, 100 = too much information) _____

How old was the child? _____

Study 1: Juror Perception Questionnaire

Information Sheet

----PAGE BREAK----

Eligibility questions

Date of birth
Month
Day
Year

Gender
- ○ Female
- ○ Male
- ○ Transgender
- ○ Other _____

Is English your native language?
- ○ Yes
- ○ No
  (please specify your native language) _____

How fluent are you in English? [If '4' or less exclude]

1	2	3	4	5	6	7
not at all fluent						very

What is your main occupation? _____

Do you have children?
- ○ Yes
- ○ No

Do any of the following apply to you?
- Have an intellectual disability
- Work in government/congress/parliament (including federal or state)
- Work in a court of law
- Employed by police/fire/ambulance departments
- Member of armed forces or defence force
- Have been sentenced to imprisonment or charged with a felony

○ Yes (if yes>EXCLUDE)
○ No (continue)

----PAGE BREAK----

How much time do you spend with children of the following ages?

Please rate how much contact (past or present) you have with children of the following ages? *e.g.,, your own children, nieces, nephews, siblings or relatives, through personal contact, through work, through volunteer activities*

	Daily	Weekly	Monthly	Yearly	Hardly ever	None at all
Babies (less than 2 years)						
Toddlers (2-4 years)						
5-6						
7-8						
9-10						
11-12						
13-14						
15-16						
17-18						

Please indicate the context(s) in which you have contact with them (and briefly describe)

- ☐ Home _____
- ☐ Work _____
- ☐ Family _____
- ☐ Volunteer activities _____
- ☐ Other _____

Have you or a family member ever been involved in a jury-tried court case?

- ☐ As a defendant
- ☐ As a witness
- ☐ As the complainant
- ☐ As a jury member
- ☐ None of the above

Have you or a family member ever given testimony in court?

- ☐ As a defendant
- ☐ As a witness
- ☐ As the complainant
- ☐ As a jury member
- ☐ None of the above

Have you or a family member ever been involved in a court case involving child witnesses?

- ☐ As a defendant

- ☐ As a witness
- ☐ As the complainant
- ☐ As a jury member
- ☐ None of the above

----PAGE BREAK----

Instructions

Thank you for taking part in our study. The information you provide will be of relevance to our research, and on completion we will credit you for your time.

You will be given a series of passages to read from interviews with a child about a recent event and you will be asked to rate them as a witness. Please respond to all questions, even if you are unsure of your responses. If you have any questions at the conclusion of the study, you can ask Helen Pierce helen.pierce@vuw.ac.nz. You are free to withdraw from the study at any time.

You will now read a series of passages from interviews with children about a recent event and you will then be asked to answer some questions about each passage.

The interviews were conducted as part of a separate study where children had a health-check (check-up) with a nurse at their school and two days were interviewed by a trained interviewer to elicit their recall of the event.

----PAGE BREAK----

Please read the following passage then move onto the next page.

[Excerpt – followed by main bank of questions]

----PAGE BREAK----

For each question, please select the number that most closely corresponds to your views…

1 2 3 4 5 6 7
not at all very

1	How likely is it that the child remembered the check-up as it actually happened? [Score: 1 – 7]	Memory
2	How confident was the child when describing the check-up? [Score: 1 – 7]	Confident
3	How credible was the child? [Score: 1 – 7]	Credible
4	How accurate do you think the child's description of the check-up was? [Score: 1 – 7]	Accurate

5	How likely is it that the child provided honest testimony? [Score: 1 – 7]	**Honest**
6	How much did the child say about the check-up? [Score: 1 – 7]	**Amount**
7	How consistent was the information that the child reported? [Score: 1 – 7]	**Consistent**
8	How swayed by interviewer suggestion do you think the child was? [Reverse Score: 7 – 1]	**Suggestible**
9	How talkative was the child? [Score: 1 – 7]	**Talkative**
10	How believable was the child's testimony? [Score: 1 – 7]	**Believable**
11	How informative was the child's description of the check-up? [Score: 1 – 7]	**Informative**
12	How coherent was their testimony? [Score: 1 – 7]	**Coherent**
13	How reliable do you think the information was that the child reported? [Score: 1 – 7]	**Reliable**
14	How concise was the child? [Score: 1 – 7]	**Concise**

How old was the child? _____

Participants then read another excerpt and answer the rating of memory interview questions before moving onto the next excerpt—until all six excerpts have been read and rated.

Appendix C: Transcripts for Pilot Survey and Study 1 Survey

Low talkative five-year-old

5 year old

(I = interviewer and C = child)

I I heard that a couple of days ago a nurse came to school and gave you a check-up. I wasn't there so I don't know what happened but I'd like to know all about it, so tell me everything that happened from the beginning to the end.

C She said how big I was.

I Okay, tell me some more things about the check-up

C She checked my ears

I Tell me some more things about the check-up even the little things

C She looked in my mouth and then I had a test on a blood checker.

…

I So earlier you also mentioned she said how big you were. So tell me everything you remember about when that happened.

C Cos there's a measuring tape on the wall.

I Alright, so you also talked about, checking your ears, tell me everything you remember about when that happened.

C She used a light to check them

I So you also mentioned, she looked in your mouth, so tell me everything you remember about when that happened.

C She put the stick in then up, she said open your mouth, and then, well she put the stick in and then she shined the light and then all done.

I Ok, you also talked about the test on a blood checker, so tell me everything you remember about when that happened.

C When I had that I get to pump out on my arm and it got tight a little and then got a little hand little thing where you can twist and it comes loose. And it all.

High talkative five-year-old

5 year old

(I = interviewer and C = child)

I I heard that a couple of days ago a nurse came to school and gave you a check-up. I wasn't there so I don't know what happened but I'd like to know all about it, so tell me everything that happened from the beginning to the end.

C There was a yellow thing with numbers on it and I had to go stand with my heels on the wall to see how high I am and that's all

I Okay, tell me some more things about the check-up

C Then she checked my ears and then she put the thing where you use the ice block in my mouth, then she used the light thing what she used for my ears and put it in my mouth

I Tell me some more things about the check-up even the little things

C I thought there was a real needle but it was just a pumping thing on my arm and they put, what you use to listen to your heart and they can, on my arm and I could heard my blood going up and down like this. And she also pumped my arm up a little bit and she put that little thingy that you use to listen to your heart, and it can feel my and that's all I remember

…

I So earlier you also mentioned seeing how high you are. So tell me everything you remember about when that happened.

C I can't remember the number but I was super big, I think it was 250 I think it was 250 a hundred and I said when I go on my tippie toes I go to a big number

I Great, so you also talked about, checking your ears, tell me everything you remember about when that happened.

C She said they were clean. And they weren't dirty. She checked my ear then she checked my other ear and then- then my other ear was clean

I So you also mentioned, the ice block in your mouth, so tell me everything you remember about when that happened.

C She told me to open my mouth and just put my tongue out. And then she saw my teeth.

I Alright, you also talked about the pumping thing on your arm, so tell me everything you remember about when that happened.

C And I also thought that, that that was a real super needle but it wasn't it was a pumping thingy on my arm. And it was really cold. When she pumped my arm up she put what you use to listen to your heart inside on my arm and I could feel my blood going ah pump pump like your heart

Low talkative eight-year-old

8 year old

(I = interviewer and C = child)

I I heard that a couple of days ago a nurse came to school and gave you a check-up. I wasn't there so I don't know what happened but I'd like to know all about it, so tell me everything that happened from the beginning to the end.

C She did how tall I was

I Okay, tell me some more things about the check-up

C Then she checked my mouth**,** then my ears

I Tell me some more things about the check-up even the little things

C And she put, yeah and then, she put a little strap around my arm then just test my arm thing and pumped it up.

…

I So earlier you also mentioned she did how tall you are. So tell me everything you remember about when that happened.

C So there was like this wall with this tape measure stuck onto it and, I stood up there and, she looked at it and, looked at how tall I was

I Alright, so you also talked about, checking your mouth, tell me everything you remember about when that happened.

C All she did was like put this popsicle thing in my mouth and tell me to go ahh and got this torch thing and all that

I So you also mentioned, checking your ears, so tell me everything you remember about when that happened.

C All she did was check my left ear with the same torch thing and then checked my right ok

I Cool, you also talked about putting a little strap around your arm, so tell me everything you remember about when that happened.

C So I think she was seeing how, good my arm was like how much blood it had on it or how much air it had or something like that so she strapped a little cardboard thing which had a bump in it and she said if it hurts just wave your hand quite, it felt quite nice. yeah she pumped it.

High talkative eight-year-old

8 year old

(I = interviewer and C = child)

I I heard that a couple of days ago a nurse came to school and gave you a check-up. I wasn't there so I don't know what happened but I'd like to know all about it, so tell me everything that happened from the beginning to the end.

C I think last time she checked my height. How high I was, like I had to take my shoes off so yeah

I Okay, tell me some more things about the check-up

C I think she looked at my like, ears, The ear check-up, she looked in my ear, she, got something that was like a light, something to look at your ears**.** And then after that she let, she looked in my left ear. She even also looked at my mouth as well, to, to check if there's any bugs or something or anything.

I Tell me some more things about the check-up even the little things

C And… I did that pumpy thingy where you put the thing on your arm and you push it.

…

I So earlier you also mentioned she checked your height. So tell me everything you remember about when that happened.

C She checked my height because the tape measure was accidentally upside down but she, I think she could still see my, like turning my head, it'll be a bit hard for doing that cause maybe your head like goes funny. Yeah and, before that I, had to take my shoes off cause she told me to. Then, she looked at, my height. And I couldn't remember how high I was, so yeah

I Ok, so you also talked about, looking at your ears, tell me everything you remember about when that happened.

C When she looked at my ears, it tickled a little bit, I was, I was just facing at facing at the window over there that I saw, and, I can't remember… When she, when she used it she put it in my ear and then she went to look of what was inside my ears and, like near the hole so there wasn't, like, kind of bad things inside of my like ears so, yeah

I So you also mentioned, looking at your mouth, so tell me everything you remember about when that happened.

C She used something like a, popsicle stick but it was bigger. And then she told me to go like aah and then I did it and the, she kind of shut her eyes and then she just checked with the light if there was anything wrong with my mouth but there was nothing wrong so yeah.

I ok, you also talked about the pumpy thing, so tell me everything you remember about when that happened.

C So it's like a strap you put on your arm and there will be a long thingy, a little squishy thingy where you squish it. Yep and after that she started squishing it and after that it

started getting like pulling into it- my arm. And after that she used stethoscope thingy's and after that she put it where the pump thingy is and after that she started pumping up and after that she heared it so that's why she put the stethoscope on it.

Low talkative twelve-year-old

12 year old

(I = interviewer and C = child)

I I heard that a couple of days ago a nurse came to school and gave you a check-up. I wasn't there so I don't know what happened but I'd like to know all about it, so tell me everything that happened from the beginning to the end.

C She measured my height with the measuring tape, I did take my shoes off to take my height

I Okay, tell me some more things about the check-up

C She looked in ears and my mouth

I Tell me some more things about the check-up even the little things

C Then she put that thing around your arm, those things you put around your arm and you pump it up

…

I So earlier you also mentioned measuring your height. So tell me everything you remember about when that happened.

C I think I was, 1 metre and 58 centimetres high and, yeah I can't remember anything else about that

I Alright, so you also talked about, looking in your ears, tell me everything you remember about when that happened.

C She had a light as well and she put that in and checked my ear. It really tickled and she had to put it right in so she could I think kinda see whether it was ok

I So you also mentioned, she looked in your mouth, so tell me everything you remember about when that happened.

C She used a like, a fat popsicle stick to hold my tongue down and she used the same torch which she did for my ears

I ok, you also talked about that thing around your arm, so tell me everything you remember about when that happened.

C I think she was taking my blood pressure I'm not too sure, yeah she just pump- she just put it on, pumped it up a few times, let it go down, pumped it up again, and then took it off that's all I can remember.

High talkative twelve-year-old

12 year old

(I = interviewer and C = child)

I I heard that a couple of days ago a nurse came to school and gave you a check-up. I wasn't there so I don't know what happened but I'd like to know all about it, so tell me everything that happened from the beginning to the end.

C We stood up and checked my height, took my shoes off, did the height. There was like the height, the measuring tape

I Okay, tell me some more things about the check-up

C Then sat back down and tested this stick on my tongue. You know the tongue and the stick. Like it was on my mouth and there was a light and it was black, with this like stick little stick. Then she checked my ears, she checked from the left ear to the right ear.

I Tell me some more things about the check-up even the little things

C Then we put, she put this, arm, arm thing on and it inflates oh yeah and then when we were doing the pump thing on your arm, we did that and I took my jersey off to put it on my arm yeah and then we, yeah. Yeah that's all I remember from that, that day and then sat back down, took my jersey off to put the pump thing and then the pump.

…

I So earlier you also mentioned checking your height. So tell me everything you remember about when that happened.

C I stood against the wall, it was like a thin, a long, yellow, measuring tape, stood against the wall and then she just like, lined my head up to which number it was, not sure which, what my height is. Yeah. And she said oh you have to stand like right up back against it so, like perfect.

I Alright, so you also talked about, testing the stick on your tongue, tell me everything you remember about when that happened.

C It's like one of those ice block sticks but it was a bit softer, it wasn't as rough and it was a bit wider longer, and then she had this black and silver, with a light on it. It was like a yellowy, like a warm light, and she puts the stick on my tongue and then looked into my mouth then, oh no she looked into my ears and my mouth and, yeah. She said like open your mouth as wide as you can, put the stick on then had a little look.

I So you also mentioned, she checked your ears, so tell me everything you remember about when that happened.

C That one had a little light coming out so there's like, a weird, like, thing coming out like it's like, it's like a shape. Yeah so like a bit like a triangle kind of, and then the top is flat with a hole and there's the light. And then you can see, and I think there's a camera behind the light. You can see properly yeah

I ok, you also talked about the pump thing, so tell me everything you remember about when that happened.

C The pump was like grey with the numbers on, and there was a, there was like dark blue that had like Velcro on it, you put it around and it has like connected to this grey tube and then it has this like pump thing you kind of squeeze it like that. And then it pumps it up, inflates it. Yeah and then, I just sat there breathing normally and, yeah, that was all that happened.

Appendix D: Assumption Checks of Data

Pilot Study

Missing Data.

Due to the counterbalance and overall design of the pilot survey each participant only rated five out of six transcripts using the talkativeness, conciseness, and amount of information scales. Consequently, for each participant there was a set of missing data for one of the transcripts. Little's Missing Completely at Random (MCAR) test was run to assess for patterns of missingness. This returned a non-significant result, χ^2 (118, N = 50) = 113.87, p = .591, meaning the data was found to be missing completely at random. The Expected Maximisation (EM) method was then used to impute missing data using SPSS software. The decision was made to conduct a paired samples t-test on the dataset with missing data and also the dataset with EM missing data imputation. The same pattern of results was found and so the dataset with EM imputation was used and the following pilot results are derived from this dataset.

Paired samples t-test.

There was one outlier—more than 1.5 box lengths from the edge of the box in a boxplot—was identified in the set of difference scores for the talkativeness ratings of five-year-old children, no outliers were identified in the other sets of difference scores. Closer inspection of the outlier found it was not extreme and it was kept in further analyses. A Shaprio-Wilk's test was conducted and the assumption of normality was not violated for any of the child age groups (five-year olds p = .856, eight-year olds p = .431, twelve-year-olds p = .609). Results from the paired samples t-test are presented in Table 14.

Table 14

Mean Difference between High-Talkative and Low-Talkative Children Matched by Age of Child

Child Age	M_{diff} (SD)	95% CI	*t*(49)	*p*	Cohen's *d*	1 - β
5-years	32.58 (24.29)	[25.68, 39.48]	9.49	<.001[a]	1.34	1.00
8-years	36.71 (21.22)	[30.68, 42.74]	12.24	<.001[a]	1.73	1.00
12-years	30.17 (23.11)	[23.60, 36.74]	9.23	<.001[a]	1.31	1.00

Note. $N = 50$. M_{diff} (SD) = Mean difference score (Standard Deviation). CI = confidence interval.
[a] These results are considered statistically significant after a Bonferroni Correction where the *p* value is adjusted.

Study 1

Witness credibility question analysis.

Before analysis of the effect of talkativeness and child age was conducted an analysis of the witness credibility questions was carried out to determine the internal consistency of the question items. This included an analysis of missing data to ascertain whether data was missing at random, completely missing at random, or not missing at random. Little's Missing Completely At Random (MCAR) analysis revealed that data was MCAR χ^2 (4226, N = 263) = 3970.22, $p = .998$. The witness credibility scale had a high level of internal consistency for each transcript, as determined by Cronbach's alpha ranging from .911 to .943.

Assumption checks for two-way repeated measures ANOVA.

Analysis of the studentized residuals revealed that the assumption of normality was violated but ignored as ANOVA is considered robust against violations of normality. A total of 67 outliers were identified—the studentized residual furthest from the mean had a value of -3.73 standard deviations. These outliers were reviewed and the decision was made to ignore them as they are likely to be genuinely unusual values and not due to data entry or measurement error.

Main effects assumptions.

Assumption of sphericity is not required for the main effect of talkativeness as there are only two levels (low and high). The assumption of sphericity was met for main effect of child age ratings of *believability* χ^2 (2) = 3.72, p = .155; *conciseness* χ^2 (2) = 5.77, p = .056; *confidence* χ^2 (2) = 5.46, p = .065; *credibility* χ^2 (2) = 2.32, p = .313; *memory* χ^2 (2) = 5.80, p = .055; and *reliability* χ^2 (2) = 3.61, p = .165. The assumption of sphericity was violated for main effect of child age ratings of *accuracy* χ^2 (2) = 10.66, p = .005, ε = .96; *coherence* χ^2 (2) = 7.21, p = .027, ε = .97; *consistency* χ^2 (2) = 9.71, p = .008, ε = 0.96; *honesty* χ^2 (2) = 15.08, p = .001, ε = .95; *informativeness* χ^2 (2) = 11.02, p = .004, ε = 0.96; *suggestibility* χ^2 (2) = 8.39, p = .015, ε = 0.97 and *talkative* χ^2 (2) = 9.76, p = .008, ε = 0.96.

Interaction effects assumptions.

Mauchly's test of sphericity indicated that, the assumption of sphericity was violated for the *amount of information* two-way interaction, χ^2 (2) = 8.63, p = .013, ε = 0.97.

Simple main effect of child age: one-way repeated measures ANOVA assumptions.

The assumption of sphericity was met for high-talkative ratings for *amount of information* χ^2 (2) = 2.94, p = .230. The assumption of sphericity was violated for low-talkative ratings for *amount of information* χ^2 (2) = 9.79, p = .007, ε = 0.96—Data are mean ± standard deviation, unless otherwise stated. Post-hoc analyses were all run with Bonferroni adjustments.

Main effects assumptions.

Assumption of sphericity is not required for the main effect of talkativeness as there are only two levels (low-talkative and high-talkative). The assumption of sphericity was met for main effect of child age ratings of *believability* χ^2 (2) = 5.57, p = .062; *credibility* χ^2 (2) = 2.49, p = .289; and *reliability* χ^2 (2) = 4.75, p = .093.

The assumption of sphericity was violated for ratings of *accuracy* χ^2 (2) = 11.64, p = .003, ε = 0.96; *conciseness* χ^2 (2) = 7.24, p = .027, ε = .97; *confidence* χ^2 (2) = 6.27, p = .044, ε = .98; *consistency* χ^2 (2) = 11.90, p = .003, ε = .96; *honesty* χ^2 (2) = 15.96, p < .001, ε = 0.94; *informative* χ^2 (2) = 13.55, p = .001, ε = 0.95; *memory* χ^2 (2) = 7.42, p = .024, ε = 0.97; *suggestible* χ^2 (2) = 8.66, p = .013, ε = 0.97; and *talkativeness* χ^2 (2) = 9.45, p = .009, ε = 0.97.

Interaction effects.

Mauchly's test of sphericity indicated that, the assumption of sphericity was met for the *coherence* two-way interaction χ^2 (2) = 0.51, p = .774; but the assumption of sphericity was violated for the *amount of information* two-way interaction, χ^2 (2) = 9.24, p = .010, ε = 0.97.

Study 2

Assumptions of one-way ANOVA.

Outliers. A total of six outliers were identified and replaced using the process of winsorization—where outlier values were replaced with the next highest/lowest value ± 1 unit. Dealing with the outliers in this way maintains the values rank in the dataset but moves it to within a reasonable range of the rest of the data. Analyses were re-run after outliers were adjusted.

Normality. The assumption of normality was assessed by Shapiro-Wilk test. Normality was met for *number of components* measure for 12-year olds p = .147 but violated for 5-year olds p = .029 and 8-year-olds p = .030. Normality was violated for transformed *accuracy* measure regardless of child age (5-years p = .009; 8-years p = .003; and 12-years p = .001). Normality was met for *MLUw* 5-year olds (p = .080) 8-year olds (p = .124) and 12 year-olds (p = .129). Normality was met for *word counts* measures of all ages (5-years p = .080; 8-years p = .124; and 12-years p = .129). Normality was met for *number of details* measure for 8-year-olds p = .703, and 12-year-olds p = .632, but violated for 5-year-olds p =

.034. The decision was made to ignore the violations to the assumption of normality because ANOVA is considered robust to these kinds of violations.

Homogeneity. Assumption of homogeneity was assessed by Levene's test for equality of variances based on means. Homogeneity was met for *number of components* recalled measure $p = .067$. Homogeneity was violated for transformed *accuracy* ($p < .001$), *MLUw* ($p = .029$), *word count* ($p = .029$), and *number of details* ($p = .003$)—ANOVA results for these measures are interpreted using the Welch ANOVA instead.

Hierarchical multiple regression assumptions.

DV = Memory accuracy; IV's = Age, Word Count, and MLUw.

The assumption of linearity was met, as assessed using visual inspection of partial regression plots and a plot of studentized residuals against the predicted values. Independence of residuals was assumed due to independent observations of each participant. Assumption of homoscedasticity was met, as assessed by visual inspection of the studentized residuals against the unstandardized predicted values. Assumption of multicollinearity was met, as assessed by tolerance values greater 0.1 and no correlations greater than 0.7. There were no significant outliers (studentized deleted residuals greater than ±3 standard deviations). There were no high leverage points (leverage values greater than 0.2). There were no highly influential points (Cook's distance above 1). Assumption of normality was by, as assessed by visual inspection of Q-Q Plot and the evaluation of regression standardized residual mean (approximate to zero) and standard deviation (approximately one).

DV = Number of components recalled; IV's = Age, Word Count, and MLUw.

As above all assumptions were met as assessed by the criteria above. There was one outlier initially identified with a studentized deleted residual value of -3.096—this case was removed, and the test was re-run.

Appendix E: Study 1 Survey Supplementary Data

Study 1 Participant Information

Table 15 shows the count and percentage of Study 1 participants and their frequency of contact with children and Table 16 shows the context of that contact. Participant contact with children was evenly distributed across the full range of child ages as well as frequency of contact. Table 17 shows that most participants had no court experience. Table 17 shows that most participants had no court experience.

Table 15

Study 1 Participant Responses for Frequency of Contact with Children

	Frequency of Contact											
	Daily		Weekly		Monthly		Yearly		Hardly ever		None at all	
Age of Children	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Babies (< 2 years)	59	22.4	33	12.5	33	12.5	28	10.6	49	18.6	53	20.2
Toddlers (2-4 years)	79	30.0	35	13.3	35	13.3	28	10.6	36	13.7	43	16.3
5-6 years	63	24.0	37	14.1	31	11.8	32	12.2	40	15.2	52	19.8
7-8 years	42	16.0	40	15.2	42	16.0	23	8.7	42	16.0	60	22.8
9-10 years	48	18.3	32	12.2	45	17.1	29	11.0	41	15.6	62	23.6
11-12 years	47	17.9	33	12.5	42	16.0	30	11.4	39	14.8	63	24.0
13-14 years	40	15.2	30	11.4	34	12.9	31	11.8	49	18.6	67	25.5
15-16 years	42	16.0	33	12.5	44	16.7	28	10.6	42	15.6	64	24.3
17-18 years	36	13.7	40	15.2	41	15.6	20	7.6	47	17.9	65	24.7

Note. Not all participants answered this question.

Table 16

Study 1 Survey Participants Context of the Contact they have with Children

Context of Contact	*n*	%
Home	175	66.5
Work	77	29.3
Family	168	63.9
Volunteer Activities	54	20.5
Other	34	12.9

Note. Participants were able to select multiple contexts in which they had contact with children.

Table 17

Study 1 Survey Participants Court Experience for a Family Member or for Themselves

Court Experience	*n*	%
Involvement in Jury-Tried Court Case		
As a Defendant	9	3.4
As a Witness	10	3.8
As Complainant	9	3.4
As a Jury Member	39	14.8
None of the Above	202	76.8
Given Testimony in Court		
As a Defendant	12	4.6
As a Witness	26	9.9
As Complainant	13	4.9
As a Jury Member	6	2.3
None of the Above	215	81.7
Involved in Court Case Involving Child Witnesses		
As a Defendant	8	3.0
As a Witness	10	3.8
As Complainant	8	3.0
As a Jury Member	5	1.9
None of the Above	236	89.7

Note. Participants were able to select multiple answers.

Study 1 EM Imputed Results

Using the EM imputed dataset, we ran a series of 2 (low- or high-talkative) x 3 (child age) repeated measures analyses of variance (ANOVA) to compare with results from the original dataset. Main effects, interaction effects, and simple main effects are presented below. All *p*-values were adjusted to account for multiple comparisons using the Bonferroni correction. In cases where the assumption of sphericity was violated we interpret the results using the Greenhouse-Geisser correction. Data are mean ± standard deviation, unless otherwise stated. Post-hoc analyses were all run with Bonferroni adjustments.

Assumption checks.

Analysis of the studentized residuals revealed that the assumption of normality was violated but ignored as ANOVA is considered robust against violations of normality. A total

of 61 outliers were identified—the studentized residual furthest from the mean had a value of -3.74 standard deviations. As with the original dataset these outliers were ignored.

Assumption of sphericity is not required for the main effect of talkativeness as there are only two levels (low and high). The assumption of sphericity was met for main effect of child age ratings of *believability* χ^2 (2) = 5.57, p = .062; *credibility* χ^2 (2) = 2.49, p = .289; and *reliability* χ^2 (2) = 4.75, p = .093. The assumption of sphericity was violated for main effect of child age ratings of *accuracy* χ^2 (2) = 11.64, p = .003, ε = .96; *conciseness* χ^2 (2) = 7.24, p = .027, ε = 0.97; *confidence* χ^2 (2) = 6.27, p = .044, ε = 0.98; *consistency* χ^2 (2) = 11.90, p = .003, ε = 0.96; *honesty* χ^2 (2) = 15.96, p < .001, ε = .94; *informativeness* χ^2 (2) = 13.55, p = .001, ε = 0.95; *memory* χ^2 (2) = 7.42, p = .024, ε = 0.97; *suggestibility* χ^2 (2) = 8.66, p = .013, ε = 0.97 and *talkative* χ^2 (2) = 9.45, p = .009, ε = 0.97. The assumption of sphericity was met for the interaction effect of *coherence* χ^2 (2) = 0.51, p = .774—but violated for interaction effect of *amount of information* χ^2 (2) = 9.24, p = .010, ε = .97.

Main effects.

Interpretation of main effects are based on the estimated marginal means and their associated standard errors. Mean differences from pairwise comparisons are significant at the p = .05 level and adjustments for multiple comparisons are made using the Bonferroni correction.

Table 18

Descriptive Statistics for all Witness Credibility Constructs of Each Transcript

Witness Credibility Construct	*M* (*SD*)					
	5-years		8-years		12-years	
	Low	High	Low	High	Low	High
Accuracy	4.68 (1.45)	4.77 (1.49)	5.06 (1.29)	5.13 (1.43)	5.41 (1.27)	5.53 (1.41)
Amount of information	3.64 (1.55)	5.01 (1.36)	4.27 (1.45)	5.38 (1.25)	4.54 (1.41)	5.58 (1.36)
Believability	5.07 (1.40)	5.09 (1.54)	5.19 (1.44)	5.43 (1.39)	5.59 (1.28)	5.63 (1.41)
Coherency	4.56 (1.48)	4.60 (1.51)	4.88 (1.37)	4.88 (1.48)	5.42 (1.35)	5.15 (1.60)
Conciseness	4.68 (1.58)	4.43 (1.55)	4.66 (1.50)	4.44 (1.58)	5.16 (1.44)	4.74 (1.79)
Confidence	4.76 (1.53)	5.06 (1.40)	4.93 (1.45)	5.02 (1.48)	5.11 (1.42)	5.32 (1.51)
Consistency	4.86 (1.45)	5.00 (1.49)	5.12 (1.28)	5.35 (1.29)	5.49 (1.22)	5.54 (1.35)
Credibility	4.91 (1.49)	4.86 (1.51)	5.13 (1.30)	5.32 (1.30)	5.57 (1.28)	5.57 (1.43)
Honesty	5.38 (1.44)	5.26 (1.64)	5.44 (1.29)	5.55 (1.40)	5.79 (1.29)	5.75 (1.40)
Informativeness	4.08 (1.58)	4.84 (1.50)	4.51 (1.52)	5.16 (1.36)	4.95 (1.36)	5.45 (1.45)
Memory	4.69 (1.39)	4.86 (1.43)	5.10 (1.33)	5.31 (1.32)	5.54 (1.27)	5.59 (1.37)
Reliability	4.77 (1.53)	4.85 (1.57)	5.14 (1.30)	5.17 (1.45)	5.38 (1.37)	5.46 (1.43)
Suggestibility	4.74 (1.62)	4.53 (1.67)	4.64 (1.65)	4.44 (1.78)	4.63 (1.73)	4.61 (1.84)
Talkativeness	3.42 (1.54)	5.06 (1.37)	3.94 (1.48)	5.51 (1.25)	4.02 (1.43)	5.78 (1.26)

Note. $n = 263$

Main effect of talkativeness.

The main effect of talkativeness showed statistically significant differences in ratings of: *conciseness* $F(1, 262) = 12.70$, $p < .001$, $\eta_G^2 = 0.02$, $\eta_p^2 = 0.05$; *confidence* $F(1, 262) = 10.75$, $p = .001$, $\eta_G^2 = 0.01$, $\eta_p^2 = 0.04$; *consistency* $F(1, 262) = 7.41$, $p = .007$, $\eta_G^2 = 0.01$, $\eta_p^2 = 0.03$; *informativeness* $F(1, 262) = 95.68$, $p < .001$, $\eta_G^2 = 0.07$, $\eta_p^2 = 0.27$; *memory* $F(1, 262) = 6.71$, $p = .010$, $\eta_G^2 = 0.01$, $\eta_p^2 = 0.02$; *suggestibility* $F(1, 262) = 9.25$, $p = .003$, $\eta_G^2 = 0.01$, $\eta_p^2 = 0.03$; and *talkativeness* $F(1, 262) = 399.68$, $p < .001$, $\eta_G^2 = 0.34$, $\eta_p^2 = 0.60$. No statistically significant main effects of talkativeness were found for ratings of *accuracy, believability, credibility, honesty*, and *reliability*. We found an interaction effect for talkativeness versus child age for ratings of *amount of information* and *coherence* and this is reported in detail later.

Main effect of child age.

The main effect of child age showed statistically significant differences in ratings of: *accuracy* $F(1.92, 502.10) = 49.92$, $p < .001$, $\eta_G^2 = 0.08$, $\eta_p^2 = 0.16$; *believability* $F(2, 524) = 28.72$, $p < .001$, $\eta_G^2 = 0.05$, $\eta_p^2 = 0.10$; *conciseness* $F(1.95, 510.05) = 17.92$, $p < .001$, $\eta_G^2 = 0.02$, $\eta_p^2 = 0.06$; *confidence* $F(1.95, 511.86) = 8.91$, $p < .001$, $\eta_G^2 = 0.01$, $\eta_p^2 = 0.03$; *consistency* $F(1.91, 501.65) = 37.02$, $p < .001$, $\eta_G^2 = 0.06$, $\eta_p^2 = 0.12$; *credibility* $F(2, 524) = 47.66$, $p < .001$, $\eta_G^2 = 0.07$, $\eta_p^2 = 0.15$; *honesty* $F(1.89, 494.66) = 22.92$, $p < .001$, $\eta_G^2 = 0.04$, $\eta_p^2 = 0.08$; *informativeness* $F(1.90, 498.77) = 44.53$, $p < .001$, $\eta_G^2 = 0.07$, $\eta_p^2 = 0.15$; *memory* $F(1.95, 509.71) = 68.13$, $p < .001$, $\eta_G^2 = 0.10$, $\eta_p^2 = 0.21$; *reliability* $F(2, 524) = 38.63$, $p < .001$, $\eta_G^2 = 0.06$, $\eta_p^2 = 0.13$; and *talkativeness* $F(1.93, 506.00) = 41.09$, $p < .001$, $\eta_G^2 = 0.05$, $\eta_p^2 = 0.14$. No statistically significant main effects of age were found for ratings of *suggestibility*.

Table 19

Estimated Marginal Means and Associated Standard Errors for Each Witness Credibility Construct by Child Age

Construct	5-years		8-years		12-years	
	M (*SEM*)	95% CI	*M* (*SEM*)	95% CI	*M* (*SEM*)	95% CI
Accuracy	4.73 (0.07)	[4.58, 4.87]	5.10 (0.07)	[4.96, 5.24]	5.47 (0.07)	[5.33, 5.61]
Believability	5.08 (0.08)	[4.93, 5.23]	5.31 (0.08)	[5.16, 5.46]	5.61 (0.07)	[5.47, 5.76]
Conciseness	4.55 (0.08)	[4.40, 4.71]	4.55 (0.08)	[4.40, 4.70]	4.95 (0.08)	[4.80, 5.11]
Confidence	4.91 (0.08)	[4.76, 5.06]	4.97 (0.07)	[4.83, 5.12]	5.21 (0.08)	[5.07, 5.36]
Consistency	4.93 (0.08)	[4.77, 5.09]	5.24 (0.07)	[5.11, 5.37]	5.51 (0.07)	[5.38, 5.65]
Credibility	4.89 (0.08)	[4.73, 5.04]	5.22 (0.07)	[5.09, 5.36]	5.57 (0.07)	[5.43, 5.71]
Honesty	5.32 (0.08)	[5.16, 5.49]	5.50 (0.07)	[5.35, 5.64]	5.77 (0.07)	[5.63, 5.92]
Informativeness	4.46 (0.08)	[4.31, 4.61]	4.35 (0.08)	[4.68, 4.98]	5.20 (0.07)	[5.06, 5.34]
Memory	4.77 (0.07)	[4.64, 4.91]	5.20 (0.07)	[5.07, 5.34]	5.56 (0.07)	[5.43, 5.70]
Reliability	4.81 (0.08)	[4.65, 4.97]	5.15 (0.08)	[5.00, 5.30]	5.42 (0.07)	[5.28, 5.56]
Talkativeness	4.24 (0.07)	[4.11, 4.38]	4.73 (0.06)	[4.60, 4.85]	4.90 (0.06)	[4.77, 5.02]

Note. $n = 263$

Table 20

Post-hoc Mean Difference Scores and Associated Standard Errors for Main Effect of Child Age

Construct	Comparison	M_{diff} (*SE*)	*p*[a]	95% *CI*
Accuracy	5-8 years	-0.37 (0.07)	< .001	[-0.55, -0.20]
	8-12 years	-0.37 (0.07)	< .001	[-0.53, -0.21]
	5-12 years	-0.74 (0.08)	< .001	[-0.94, -0.55]
Believability	5-8 years	-0.23 (0.07)	.002	[-0.39, -0.07]
	8-12 years	-0.31 (0.07)	< .001	[-0.48, -0.14]
	5-12 years	-0.53 (0.07)	< .001	[-0.71, -0.35]
Conciseness	5-8 years	0.00 (0.08)	1.000	[-0.18, 0.19]
	8-12 years	-0.40 (0.07)	< .001	[-0.57, -0.23]
	5-12 years	-0.10 (0.08)	< .001	[-0.59, -0.20]
Confidence	5-8 years	-0.06 (0.08)	1.000	[-0.24, 0.12]
	8-12 years	-0.24 (0.07)	.002	[-0.41, -0.07]
	5-12 years	-0.30 (0.08)	.001	[-0.50, -0.11]
Consistency	5-8 years	-0.31 (0.07)	< .001	[-0.48, -0.14]
	8-12 years	-0.28 (0.06)	< .001	[-0.42, -0.13]
	5-12 years	-0.58 (0.07)	< .001	[-0.76, -0.41]
Credibility	5-8 years	-0.34 (0.07)	< .001	[-0.50, -0.17]
	8-12 years	-0.34 (0.07)	< .001	[-0.51, -0.18]
	5-12 years	-0.68 (0.07)	< .001	[-0.86, -0.51]
Honesty	5-8 years	-0.18 (0.07)	.024	[-0.34, -0.02]
	8-12 years	-0.27 (0.06)	< .001	[-0.42, -0.13]
	5-12 years	-0.45 (0.07)	< .001	[-0.63, -0.27]
Informativeness	5-8 years	-0.37 (0.08)	< .001	[-0.57, -0.17]
	8-12 years	-0.37 (0.07)	< .001	[-0.53, -0.20]
	5-12 years	-0.74 (0.08)	< .001	[-0.94, -0.54]
Memory	5-8 years	-0.43 (0.07)	< .001	[-0.60, -0.26]
	8-12 years	-0.36 (0.06)	< .001	[-0.51, -0.21]
	5-12 years	-0.79 (0.07)	< .001	[-0.96, -0.62]
Reliability	5-8 years	-0.34 (0.07)	< .001	[-0.51, -0.17]
	8-12 years	-0.27 (0.07)	< .001	[-0.42, -0.11]
	5-12 years	-0.61 (0.07)	< .001	[-0.78, -0.43]
Talkativeness	5-8 years	-0.49 (0.08)	< .001	[-0.68, -0.29]
	8-12 years	-0.17 (0.07)	.040	[-0.33, -0.01]
	5-12 years	-0.66 (0.08)	< .001	[-0.84, -0.47]

Note. M_{diff} = mean difference of comparison; *SE* = standard error of mean difference; *CI* = confidence interval.
[a] Adjustment for multiple comparisons using Bonferroni correction

Interaction effects.

Participant ratings for *amount of information* $F(1.93, 506.39) = 3.77$, $p = .025$, $\eta_G^2 = 0.00$, $\eta_p^2 = 0.01$ and *coherency* $F(2, 524) = 3.43$, $p = .033$, $\eta_G^2 = 0.00$, $\eta_p^2 = 0.01$; produced a

statistically significant interaction between talkativeness and child age with small effect sizes. To unpack these interactions we carried out one-way repeated measures ANOVAs to further investigate the simple main effect of talkativeness and then the simple main effect of child age for each interaction. Table 21 presents the means and standard deviations for low/high talkative and 5-, 8-, and 12-year-old children.

Table 21

Means and Standard Deviations for Simple Main Effect of Talkativeness on Amount of Information

	Amount of information		Coherence	
	Low-Talkative	High-Talkative	Low-Talkative	High-Talkative
Child Age	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
5-years	3.64 (1.55)	5.01 (1.36)	4.56 (1.48)	4.60 (1.51)
8-years	4.27 (1.45)	5.38 (1.25)	4.88 (1.37)	4.88 (1.48)
12-years	4.54 (1.41)	5.58 (1.36)	5.42 (1.35)	5.15 (1.60)

Note. $n = 263$

Simple Main effects.

Post-hoc one-way repeated measures ANOVA were carried out to unpack the simple main effects of talkativeness and child age for the *amount if information* and *coherence* constructs. The assumption of sphericity was violated for low-talkative ratings for *amount of information* $\chi^2(2) = 10.45$, $p = .005$, $\varepsilon = 0.96$. The assumption of sphericity was met for high-talkative ratings for *amount of information* $\chi^2(2) = 3.91$, $p = .142$. The assumption of sphericity was also met for low- $\chi^2(2) = 3.61$, $p = .165$ and high- $\chi^2(2) = 0.83$, $p = .659$ talkative ratings for *coherence*. The assumption of sphericity is not relevant for unpacking the simple main effect of child age.

Amount of information.

There were statistically significant differences in *amount of information* ratings for all age group comparisons—*5-year-olds* $F(1, 262) = 157.49$, $p < .001$, $\eta_G^2 = 0.24$, $\eta_p^2 = 0.38$, *8-year-olds* $F(1, 262) = 133.39$, $p < .001$, $\eta_G^2 = 0.20$, $\eta_p^2 = 0.34$, and *12-year-olds* $F(1, 262) = 91.68$, $p < .001$, $\eta_G^2 = 0.17$, $\eta_p^2 = 0.26$. Additional post-hoc one-way repeated measures ANOVAs showed statistically significant differences in *amount of information* ratings for *low-talkative* $F(1.92, 504.21) = 35.60$, $p < .001$, $\eta_G^2 = 0.06$, $\eta_p^2 = 0.12$, and *high-talkative*

$F(2, 524) = 19.75$, $p < .001$, $\eta_G^2 = 0.03$, $\eta_p^2 = 0.07$. Pairwise comparisons are presented in Table 22.

Coherency.

There was only a statistically significant difference in *coherence* ratings between low- and high-talkative ratings for *12-year-olds* $F(1, 262) = 6.69$, $p = .010$, $\eta_G^2 = 0.01$, $\eta_p^2 = 0.02$, with a small effect size. No significant differences were found between low- and high-talkative ratings for *5-year-olds* $F(1, 262) = 0.19$, $p = .660$ or *8-year-olds* $F(1, 262) < 0.01$, $p = .946$.

Additional post-hoc one-way repeated measures ANOVAs showed statistically significant differences in *coherence* ratings for *low-talkative* $F(2, 524) = 40.01$, $p < .001$, $\eta_G^2 = 0.06$, $\eta_p^2 = 0.13$, and *high-talkative* $F(2, 524) = 13.87$, $p < .001$, $\eta_G^2 = 0.02$, $\eta_p^2 = 0.05$. Pairwise comparisons are presented in Table 22.

Table 22

Post-hoc Pairwise Comparisons for Simple Main Effect of Child Age

	Low-Talkative			High-Talkative		
Comparisons	M_{diff} (*SE*)	*p*[a]	95% *CI*	M_{diff} (*SE*)	*p*[a]	95% *CI*
Amount of information						
5-8 years	-0.63 (0.11)	< .001	[-0.90, -0.35]	-0.37 (0.09)	< .001	[-0.60, -0.14]
8-12 years	-0.27 (0.10)	.017	[-0.51, -0.04]	-0.20 (0.09)	.066	[-0.40, 0.01]
5-12 years	-0.90 (0.12)	< .001	[-1.18, -0.62]	-0.57 (0.09)	< .001	[-0.80, -0.34]
Coherency						
5-8 years	-0.32 (0.10)	.005	[-0.57, -0.08]	-0.28 (0.11)	.024	[-0.54, -0.03]
8-12 years	-0.54 (0.09)	< .001	[-0.76, -0.32]	-0.26 (0.10)	.030	[-0.51, -0.02]
5-12 years	-0.87 (0.10)	< .001	[-1.11, -0.62]	-0.55 (0.10)	< .001	[-0.79, -0.30]

Appendix F: The MARCIE Project Coding Scheme

Definitions

Component: The different parts that make up the content of the health-check.

CORE: Aspects of each component that are central to recalling that component of the health-check. Each component is broken down into action, body part, and equipment.

What happened? The *action* that took place and the goal or function of this action.

Where on the body? The body part(s) involved in the component of the health-check.

Equipment – What equipment was used? Descriptions of the equipment used in the particular component of the health-check.

Correct (verbatim) [CC]: A point is scored in this category for each accurate piece of recalled information associated with a component. This includes a correct description of what happened, correct identification of where on the body it took place, and correct description or naming of the equipment that was used. Each component has a maximum possible correct score.

Errors: May reflect gist driven recall or entirely false memory (intrusion) as attributed below…

Gist: A point is scored in this category if the recollection is not 100% accurate due to either a distortion or an omission—one point is given for each distortion or omission.

Core Distortion [CD]: When a component that actually occurred is incorrectly recalled. This includes an incorrect description of what happened, incorrect identification of where on the body it took place, or incorrect description or naming of the equipment that was used.

Core Omission [CO]: When a component that actually occurred is not recalled. This includes no description of what happened, no identification of

where on the body it took place, or no description or naming of the equipment that was used.

Intrusion: When false information is presented during the interview. This is considered a false memory as the event did not actually occur—one point is given for each detail of reported intrusion. Intrusions are coded based on actions/what happened.

Gist/Related Intrusion [IR]: This is an intrusion that is related to the health-check and includes our critical components/distractors as well as the related distractors from the recognition test.

Unrelated Intrusion [IU]: This is an intrusion that is completely unrelated to the health-check and includes unrelated distractors from the recognition test.

Off topic information [OT]: Points are given for information that is unrelated to the event.

Elaborative/additional information: Points are given for information that is additional to the CORE information or particularly detailed descriptions of a component, this includes opinions of the participant. Elaborative information may also include talking about aspects of the health-check but not a specific component. Elaborative information is scored as true/correct **[EC],** false/incorrect **[EI]** or other **[EO]**.

No Code [NC]: Information that is not given a code

Repeated [R] information is not scored

Query [?] information could be assigned more than one code and needs to be reviewed

Coding Rules

A. Each component of the health-check should be assigned a code, if the component is not mentioned at all then it is considered an omission and should be coded as such
B. Information provided about the video, or walking to/from the health-check should be coded as Off Topic [OT]
C. Repeated information about the health-check should be coded with the prefix X/R- (e.g., X/R-TCC/A)
D. The different coding levels are as follows:
    i. Level 1:
        - Component or Key Action/Body Area
        - Error: False Memory/Intrusion
            - Related (Gist) [FM/R]
            - Unrelated [FM/U]
        - Off topic (Unrelated information) [X/OT]
        - No Code [X/NC]
        - Repeated information [X/R-]
        - Query [?]
    ii. Level 2:
        - Component Correct [CC]
            - Action [A]
            - Body Part [B]
            - Equipment [E]
        - Error
            - Component Distortion [CD]
                - Action [A]
                - Body Part [B]
                - Equipment [E]
            - Component Omission [CO]
                - Action [A]
                - Body Part [B]
                - Equipment [E]
        - Elaboration (Related information)
            - Correct [EC]
            - Incorrect [EI]
            - Opinion [EO]
E. The hierarchy for coding each components is as follows; Action [A], Body Part [B], Equipment [E].
    i. Give a default point for first level for each component if any lower level points have been given
F. CORE codes (e.g., XXX/A) should only be given if the child has given a clear description or named the particular component – if the description or naming of the component is not clear then an ELABORATE code should be given instead (e.g., She put a thing around my arm = BEC)

i. If the action is described or named then the Action code can be assigned, (e.g., She checked my temperature = TCC/A)
ii. If the body part is described or named then the Body code can be assigned (e.g., She looked in my mouth = MCC/B[1])
iii. If the equipment if described or named then the Equipment code can be assigned (e.g., She put a strap around my arm =BCC/E[1]"

G. Use "?" as a prefix followed by the possible coding options for when information could be given more than one code, or as a suffix if only a body part is named and it is unclear what component it is associated with.
H. Elaborative information that is unable to be coded as correct or incorrect should be coded as other e.g., "the nurse rammed it in really hard to my ear"
I. Points are not given for left/right body parts (this detail should be coded as elaborative).
J. Assign a point for each alternative aspect of a component e.g., "I covered my left eye (1 point) with my left hand (1 point) and then my right eye (1 point) with my right hand (1 point)" [total of 4 points]
K. She checked my eyes should be given the codes [ICC/A & ICC/B]
L. She checked my ear should be given code [RCC/A & RCC/B] unless there is sufficient information to give an alternative code (e.g., Temperature TCC/A)
M. Components can only be coded as a distortion if: their account of that component is the ONLY version of that particular component. E.g., if they talk about the nurse taking their temperature under their tongue then this is a distortion *unless* they have already talked about the nurse taking their temperature in their ear.
N. Additional components (that are distortions of a health-check component) are coded as False Memory Related [FM/R] e.g., a second temperature check. They should not be coded as a distortion of the component unless it is the only time they mention that particular component.
O. If assigning multiple points where some are correct and some are incorrect follow the coding hierarchy e.g., assign points to higher levels first and then remaining information can be coded as elaborative correct/incorrect.
P. If child says something like 'She checked my blood" this should be coded as [A-B] where only a default point for blood pressure is given.
Q. To get the point for covering their eye [VCC/A] children need to either state verbally that they covered their eye or they need to clearly gesture covering their eye. Pointing to the eye is not sufficient.
R. If describing the function of equipment then points are assigned to the equipment code, if they are describing the appearance of equipment then points are assigned to the elaboration code.
S. When talking about location no points assigned
T. If they mention the nurse just talked to me it only gets an OEC point.

Verifiable information:

- Descriptions of equipment
- Appearance of nurse (as shown in photos)

- Results from health-check (as recorded on health-check form)

Reliability

Calculating reliability must be done by simultaneously comparing code for code using the reliability scoring sheet. For codes that are the same this is marked as an agreement, for codes that are not the same this is marked as a disagreement. The line number and conflicting codes should be recorded for each disagreement point along with a reason for the conflict. Disagreements should then be discussed and any new rules should be recorded. If disagreements are unable to be resolved this should also be recorded.

Definitions of CORE Components

Table 23

Definition of Codes and their Associated Category

Code	Category	Definition
_CC/A	CORE Action	The action that the nurse makes;
	CORE Goal	The goal/purpose of the health-check component;
	CORE Function	The function/outcome of the health-check component
_CC/B	CORE Body part	The body part(s) directly involved in the health-check component
_CC/E	CORE Equipment	A description of the equipment used as part of the health-check component
_CO/A	Core Distortion/Action-Goal-Function	A distortion of the action, goal, or function of the health-check component
_CO/B	Core Distortion/Body part	A distortion of the body part(s) directly involved in the health-check component
_CO/E	Core Distortion/Equipment	A distortion of the description of the equipment used as part of the health-check component
_CD/A	Core Omission/Action-Goal-Function	No mention of the action, goal, or function of the health-check component
_CD/B	Core Omission/ Body part	No mention of the body part(s) directly involved in the health-check component
_CD/E	Core Omission/Equipment	No mention of any of the equipment used as part of the health-check component
_EC	Elaborative Correct	Additional information that can be verified as TRUE about the health-check component but isn't already covered by the CORE components
_EI	Elaborative Incorrect	Additional information that can be verified as FALSE about the health-check component but isn't already covered by the CORE components
_EO	Elaborative Opinion	Additional information that can be can't be verified because it is a personal opinion or feeling about the health-check component