The Cognitive Interview: The effectiveness of cognitive load questioning when children with and without disabilities provide eyewitness reports of another's transgression

Department of Educational and Counselling Psychology

McGill University, Montreal

August 2019

A thesis submitted to the Faculty of Graduate and Postdoctoral Studies of McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in School and Applied Child Psychology

Abstract

Children, especially those with intellectual disabilities, are perceived as having lower eyewitness credibility given that they are often reluctant or unable to effectively disclose the transgressions of others. To assess methods for improving child eyewitness credibility, the following two studies evaluated the efficacy of the Cognitive Interview (CI) with typical and atypical child populations. In both studies, typically-developing children (N = 104; ages 8 to 13) and those with intellectual disabilities (N = 50; ages 6 to 18) were asked to keep the transgression of an adult a secret. Next, children were interviewed using the CI (free-recall, cognitive load and closed-ended questions) or a Standard Interview (SI: free-recall and closed-ended questions) by an unfamiliar researcher. Two raters then coded the honesty, amount of information (words, details and events), disclosure frequency and forthcomingness, temporal order accuracy and testimony consistency of each report.

In both studies, children gave testimonies with more words, transgression details and disclosures on the CI when compared to the SI, without compromising report accuracy or consistency; however, the CI did not discourage false reporting. Truth-tellers in both studies provided the most forthcoming and detailed statements; whereas, lie-tellers rarely discussed the transgression. Among the typically-developing child reports only, the free-recall and reverse-order recall questions were significantly more effective than the other cognitive load questions for encouraging detailed eyewitness responses. These results are relevant to forensic professionals who interview children as the CI was effective for increasing children's eyewitness recall, without compromising accuracy. Furthermore, the CI has the potential to be an effective lie-detection tool for distinguishing between children's true and false narratives.

Résumé

Les enfants, en particulier ceux ayant une déficience intellectuelle, sont perçus comme ayant une crédibilité de témoin inférieure étant donné qu'ils sont souvent réticents ou incapables de révéler efficacement les transgressions commises par d'autres. Pour évaluer les méthodes d'amélioration de la crédibilité de témoin chez l'enfant, les deux études suivantes ont évalué l'efficacité de l'entrevue cognitive (CI) avec des populations d'enfants typiques et atypiques. Dans les deux études, on a demandé aux enfants typiques (N = 104; âgés de 8 à 13 ans) et à ceux ayant une déficience intellectuelle (N = 50; âgés de 6 à 18 ans) de garder en secret une transgression commise par un adulte. Ensuite, les enfants ont été interrogés à l'aide du CI (questions de rappel libre, questions de charge cognitive et questions fermées) ou d'une entrevue standard (SI: questions de rappel libre et des questions fermées) par un chercheur non-connu. Deux évaluateurs ont ensuite codé pour l'honnêteté, la quantité d'informations révélée (nombre de mots, de détails et d'événements), la fréquence des révélations et la volonté de l'enfant à communiquer la révélation, la précision de l'ordre temporel et la cohérence des témoignages de chaque rapport.

Dans les deux études, les enfants ont produit plus de mots pendant leur témoignage, plus de détails de la transgression et des révélations sur le CI par rapport au SI et, ceci, sans compromettre l'exactitude ou la cohérence du rapport. Cependant, le CI n'a pas découragé les déclarations malhonnêtes. Les véristes des deux études ont fourni des déclarations plus détaillées, volontiers, et plus communicatifs plus tôt lors de leur témoignage, tandis que les menteurs discutaient rarement de la transgression. Uniquement chez les déclarations d'enfants en développement typique, les questions à rappel libre et à rappel inversé étaient significativement plus efficaces que les autres questions sur la charge cognitive pour encourager

des réponses détaillées des témoins. Ces résultats sont pertinents pour les professionnels dans les domaines de la police et de la loi qui interrogent les enfants, car le CI était efficace pour augmenter le rappel des témoins, sans compromettre la précision. En outre, le CI pourra être un outil important dans la détection de mensonges pour distinguer les récits vrais des récits faux chez les enfants.

Acknowledgements

I would like to first thank my thesis supervisor, Dr. Victoria Talwar, for her tremendous supervision, encouragement and help throughout my graduate career. She has consistently shown absolute respect to everyone, and I will always be grateful for her guidance and the countless opportunities she provided to help me develop as a person and future academic. I would like to send a big thank you to my two co-pilots on most of my research throughout my graduate career, Ida Foster and Donia Tong. We worked on a number of projects together, from initial project designs, to data collection and then to the presentation of the results. I will always appreciate your incredible contributions, hard-work and friendship during these past six years. I would also like to send a special thank you to the more than two dozen volunteers who have assisted on my various research endeavors. Moreover, a special thank you to my fellow lab mates for their friendship, encouragement and help. We always supported each other during the hard times, and we had a lot of fun too! A big thank you to the many professors and clinical supervisors who helped me develop as a clinician. I learned so much from all of you, and I will always be grateful for your kindness, patience and wisdom that was essential in helping me grow as a person and as a helping professional.

Last, but definitely not least, I would like to thank my family for their unconditional support and encouragement throughout my academic career. My dad is, and always will be, my role model in life. He is the most hardworking, kind and supportive person I've ever known, and I try to display the same characteristics everyday. And to my amazing wife, Hailey. You are the light of my life. I am amazed by your intelligence, strength, courage and originality, and I am so fortunate to be able spend everyday growing with you. Finally, I dedicate this work to my late mom. She always showed such incredible love towards everyone, despite her suffering. No matter what I'm going through in life, I will always feel loved and cared for because of you.

Contributions of Authors

I was the principal author on both manuscripts. However, I received tremendous assistance from the co-authors of these two manuscripts, as well as from the more than a dozen volunteers who helped with the data collection on both studies. With the assistance of my co-authors, I conceptualized the research questions and designed the experimental methodology for both studies. I also assisted with the participant testing, coding and data input. Furthermore, I wrote the full drafts of each manuscript on my own; this also included doing the statistical analyses and results interpretations by myself. My co-authors then helped with the editing of both manuscripts before they were eventually submitted for publication.

My co-authors for the first study were Donia Tong (McGill University), Ida Foster (McGill University), Dr. Angela Crossman (John Jay College of Criminal Justice) and Dr. Victoria Talwar (McGill University). For the second study, my co-authors were Dr. Christine Saykaly (Summit School) and Dr. Victoria Talwar. Dr. Victoria Talwar was my primary supervisor and she oversaw the entire research process for both studies, from the conceptualization stage until the eventual submission of each manuscript. Donia Tong and Ida Foster, both PhD students in the School/Applied Child Psychology program at McGill University, helped greatly with the data collection, coding and editing of the first manuscript. Dr. Angela Crossman also supported the research process for study 1, as well as the manuscript editing. Lastly, Dr. Saykaly helped with the research design, data collection and manuscript editing for the second study.

Table of Contents

List of Tables and Figures

Manuscript 1

Manuscript 2

Chapter 1: Introduction

While there have been notable gains in recent decades in the forensic procedures and legal policies pertaining to interviewing adult and child eyewitnesses (see Bull, Valentine, & Williamson, 2009; Faller, 2014, for historical overviews), recent forensic research continues to highlight several problems with respect to how children, especially those with intellectual deficits, are interviewed about traumatic experiences (see Lyon, 2014; and Wyman, Lavoie, & Talwar, 2018, for reviews). For example, forensic professionals often feel unprepared to interview children with intellectual disabilities (CWID), as they report having inadequate training and procedures needed to effectively interview this population (Taylor, Stalker, & Stewart, 2016). This is unsurprising given that many of the commonly used interviewing practices, such as the Cognitive Interview (CI; Fisher & Geiselman, 1992), have been tested primarily with typically-developing children (TDC; Wyman, Lavoie et al., 2018). However, the cognitive and adaptive-skill profile of CWID differs considerably when compared to their typically-developing counterparts (American Psychiatry Association, 2013), which can impact their ability to provide detailed eyewitness reports (see Wyman, Lavoie, et al., 2018). For this reason, there is a great need for more research on ways to improve the testimony performance of CWID so that they can be empowered to actively report any crimes they witness.

Apart from best-practice free-recall questioning strategies, cognitive load questions used in the Cognitive Interview (CI; Fisher & Geiselman, 1992) have been found to increase the quality and quantity of information in the testimonies of adults (see Bull, Paulo, Albuquerque, 2019; Holliday, Brainerd, Reyna, & Humphries, 2009; Memon et al., 2010, for reviews). However, the limited existing research on the benefits of these questions with children has produced inconsistent findings. In some studies, these questions increased the amount of

information recalled by TDC and CWID when compared to standard interviewing procedures (e.g., Gentle, Milne, Powell, & Sharman, 2013; Liu, Granhag, Landström, Hjelmsaeter, Strömwall, & Vrij, 2010; McCauley & Fisher, 1995). Conversely, other studies reported that these questions were too complex for children under the age of 8 years (e.g., Memon et al., 2010; Wyman, Foster, Crossman, Colwell, & Talwar, 2018). Given the lack of clarity regarding the efficacy of the CI with children, more comprehensive research is needed to assess whether cognitive load questioning can improve the eyewitness recall performance of children with varying developmental capacities.

Much is known about the narrative features of adults' true and false testimonies (see Bull, 2018; Colwell, Hiscock-Anisman, & Fede, 2013, for reviews). Specifically, adult truth-tellers produce longer and more detailed disclosures about a transgression compared to lie-tellers who falsely denied the event. It is presently unclear, however, whether these differences also exist when evaluating children's testimonies. The narrative features of a child's testimony may be influenced by factors beyond its level of veracity, such as his/her age and level of cognitive maturity (Hershkowitz, Lamb, Orbach, Katz, & Horowitz, 2012; Jack, Leov, & Zajac, 2014). Thus, child reporters with lower cognitive functioning are often perceived to have lower credibility (Henry, Ridley, Perry, & Crane, 2011). Research that specifically examines evidence-based methods for identifying indicators of truthful and false statements can better prepare forensic professionals to conduct informed credibility evaluations of child eyewitness reports.

To address these notable gaps in the literature, the following two studies comprehensively evaluated the efficacy of the CI, in comparison to a Standard Interview (SI), in cases whereby TDC (Study 1) and CWID (Study 2) were asked to falsely deny the transgression of an adult. The testimony performance of the children was then examined on a number of

evidenced-based dependent measures that were selected on the basis of prior developmental, deception and eyewitness literature. Namely, the *veracity* (i.e., attempts and final decisions to deceive), *accuracy* (i.e., temporal order accuracy and testimony consistency), *forthcomingness* (i.e., willingness to make a disclosure), and *amount of important information recalled* (e.g., transgression details and events recalled) on the CI and SI were examined to determine whether cognitive load questions can be effective tools for encouraging TDC and CWID to produce credible eyewitness reports. Lastly, the narrative characteristics of children's statements were evaluated to better understand whether their true and false reports on the CI and SI can be distinguished from each other. The results from both studies are designed to inform forensic professionals, including interviewers and those who evaluate testimony credibility, regarding the best questioning strategies for interviewing typical and atypical child eyewitnesses.

Chapter 2: Literature Review

There are more than three million investigations of crimes against children every year in the United States (U.S. Department of Health & Human Services, 2016), including physical, sexual, psychological and medical abuse, and neglect. Children are especially vulnerable to maltreatment given their dependence-on and/or loyalty towards adults who are most likely to perpetrate these crimes. Most often a child is maltreated by someone he/she knows, such as a parent, caregiver (e.g., non-kin foster care guardians), sibling, relative, family friend or teacher (see Wissink, Van Vugt, Moonen, Stams, & Hendriks, 2015, for a review). Maltreatment has severe and long-lasting consequences on life outcomes, such as: increased vulnerability to mental (e.g., addiction, depression, post-traumatic stress) and physical health (e.g., obesity, teen pregnancy) problems; low academic achievement and school dropout; socio-economical and employment challenges; and future conduct problems and risky (e.g., unsafe sexual acts) behaviours (see Gilbert et al., 2009, for a review).

CWID are especially vulnerable to being victims of maltreatment. Overall, CWID are more than four times likely to be maltreated when compared to TDC (Jones et al., 2012) because of their increased dependence on their family members, who are most often the perpetrators (Wissink et al., 2015). This is especially problematic as CWID are more vulnerable to witness coaching given that they often distrust the accuracy of their own recollections (Cederborg, Danielsson, La Rooy, & Lamb, 2009), and therefore, rely on information provided by others (Bowles & Sharman, 2014). CWID also have weaker self-protection skills and an inadequate knowledge of healthy social practices (Murphy & O'Callaghan, 2004); thus, they may struggle to understand and identify situations wherein they are being maltreated.

Despite the high number of investigations every year, many cases of abuse and neglect go unreported and/or wrongly unsubstantiated. Children often deny being maltreated, even when there is clear evidence of abuse (e.g., perpetrator confession) available to investigators (Faller, 2016; Lyon, 2007). A child may intentionally conceal the abuse because they may want to avoid any negative feelings associated with recalling the trauma (Hershkowitz, Orbach, Lamb, Sternberg, & Horowitz, 2006). Moreover, children often feel complicit in the abuse (Hershkowitz, Lanes, & Lamb, 2007), and they may feel internal or external pressure (e.g., witness coaching) to protect the perpetrator (Wissink et al., 2015).

Among the children who do report the maltreatment, legal and child welfare professionals report having low confidence in the credibility of their reports (see Wyman, Lavoie, et al., 2018, for a review). This is especially true for disclosures given by younger children and those with intellectual deficits, as legal professionals often feel that they have inadequate training and materials for meeting the developmental needs of these children (Aarons, Powell, & Browne, 2004; Taylor et al., 2016). In fact, many child welfare agencies in the United States do not have specific written policies for serving people with disabilities (Lightfoot & LaLibierte, 2006; Shannon & Agorastou, 2006). Inadequate training and interviewing procedures can lead to unreliable and suggestive eyewitness reports (Andrews, Lamb, & Lyon, 2015; Guadagno, Powell, & Wright, 2006). Given these serious concerns, this research program is designed to improve the interviewing practices available to investigators so that TDC and CWID can give credible, detailed and forthcoming reports about an adult's transgression.

Developmental Considerations for Child Interviewing

A child's developmental capacity is an essential factor when analyzing the efficacy of interviewing strategies. In their meta-analysis of prior CI literature, Memon and colleagues

(2010) found that the cognitive load questions had smaller positive effects for children under the age of 7. Most forensic interviewing methods, such as the CI (Fisher & Geiselman, 1992) and National Institute of Child Health and Human Development (NICHD; Orbach et al., 2000) protocols, rely on the responder giving a verbal description of a previously experienced event. Unsurprisingly then, higher verbal functioning is associated with increased eyewitness recall (e.g., Jack et al., 2014; Wyman, Foster, Lavoie, Tong, & Talwar, 2019). CWID, who typically possess verbal deficits (American Psychiatric Association, 2013), therefore struggle to provide detailed (Snow & Powell, 2004) and credible (Henry et al., 2011) free-recall narratives.

Deficits in working memory and executive control reduce the amount of information that is encoded (Vicari, 2004), and limit the ability to combine several individual memories into whole reports (Tager-Flusberg, 1991). Consequently then, memory deficits limit the amount of details a child can recall about an event (Brown, Lewis, Lamb, & Stephens, 2012). In the lie-telling literature (see Talwar & Crossman, 2011, for a review), children's ability to tell consistent false reports increase with age and cognitive performance on measures of inhibitory control and theory-of-mind. Inhibitory control is utilized when suppressing interfering thoughts or behaviours when focusing on a specific goal or action (Carlson, Moses, & Breton, 2002). To tell a convincing lie, the speaker must be able to inhibit any information (truthful or false) that may contradict and/or lead to the detection of their lie (Evans, Xu, & Lee, 2011; Talwar & Lee, 2008). In contrast, theory-of-mind is the ability to understand another's perspective, and that one's thoughts or actions are neither evident nor equal to those of others (Wimmer & Perner, 1983). To be persuasive, the speaker must consider how the recipient is interpreting and processing the information they are giving (see Talwar & Crossman, 2012, for a review).

A child's gender is another potentially important, yet overlooked, factor when assessing testimony performance. Notably, girls perform better than boys on tests of verbal functioning (Hyde, 2016; Petersen, 2018) and language scholastic achievement (Voyer & Voyer, 2014) to a small degree. Socially, girls outperform boys on measures of interpersonal communication and social awareness (Hall & Matsumoto, 2004; Hampson, van Anders, & Mullin, 2006). Adult women also outperform men on measures of testimony performance (e.g., report quantity and quality) on the CI (Suckle-Nelson et al., 2010). In several lie-detection studies (Mann, Vrij, & Bull, 2004; Talwar, Lee, Bala, & Lindsay, 2006; Wyman, Foster, et al., 2018), mock-jurors perceived child and adult female eyewitnesses to be more convincing than males.

Given these potentially important developmental factors, children's age, gender and cognitive test performance will be considered when examining the efficacy of the CI and SI with TDC and CWID. In Study 1, the common predictors of lie-telling ability, namely inhibitory control, theory of mind and word productivity (Talwar & Crossman, 2011), will be considered when evaluating the efficacy of each interviewing tool with TDC. In Study 2, the CWID's Full-Scale IQ, verbal comprehension and visual reasoning will be assessed to determine if the children's general level of cognitive functioning affects their eyewitness performance.

Current Procedures for Interviewing Children

Clear, detailed and credible eyewitness reports are needed to effectively investigate, substantiate and prosecute allegations of maltreatment (Walsh, Jones, Cross, & Lippert, 2008). The forensic interview protocols that are commonly used with children, such as the CI and NICHD protocols, incorporate a range of rapport building, free-recall, prompts, forced-choice and cognitive load questions (Fisher & Geiselman, 1992; Orbach et al., 2000). Free-recall questions are open-ended and allow the responder to describe the event in question in their own

words (Lamb, 1998). Typically, these questions are asked at the beginning of the interview as the responder's free-recall narratives can be used to create non-suggestive follow-up prompts and questions (Memon et al., 2010). Children's free-recall reports are generally longer, more detailed, reliable and accurate than responses on closed-ended questions as the information obtained is presented in the speaker's own words (e.g., Lamb, Orbach, Hershkowitz, Esplin, & Horowitz, 2007; Quas, Goodman, Ghetti, & Redlich, 2000). Prompts typically include brief follow-up inquiries (e.g., "Please tell me more about that?") that encourage the responder to elaborate on their previous statements. Like free-recall questions, follow-up prompts further improve the amount of information recalled by children (Hershowitz et al., 2012).

Forced-choice questions are closed-ended and require the interviewee to provide a response based on a set of options, including multiple-choice and yes/no questions (Peterson & Biggs, 1997; Peterson & Grant, 2001). Particularly with young children, forced-choice questions have been criticized for being suggestive (Peterson & Briggs, 1997; Peterson & Grant, 2001) as responders can be intentionally or accidentally manipulated into choosing a specific answer. Moreover, by not encouraging the responder to elaborate on their prior responses, important information pertaining to the event may be omitted as children may simply opt to give brief and uninformative yes/no responses to such questions (Okanda & Itakura, 2010; Peterson & Grant, 2001). On the other hand, forced-choice questions have been found to be particularly beneficial in cases where the responder has low recall or language skills (Lyon, 1999), as younger children may not have the requisite cognitive skills needed to openly describe an event.

Despite the clear benefits of free narrative questions for eliciting accurate and detailed eyewitness reports (e.g., Quas et al., 2000; Wyman et al., 2019), prior research also highlights some clear limitations with these questions. Lyon (2014) argued that asking children to broadly

describe everything they remembered during a specified time period can result in broad and uninformative responses. Notably, younger children provide less information in their free-recall responses, as the ability to produce longer and more detailed free narratives increases with age (e.g., Jack et al., 2014; Wyman et al., 2019). Further, children who are motivated to conceal information from the interviewer are more likely to do so in their free-recall narratives (Pipe & Wilson, 1994). At the same time, closed-ended recognition questions can be very problematic since children are more vulnerable to providing suggestive and uninformative responses to these questions (Okanda & Itakura, 2010; Quas et al., 2007). Given these limitations with current interviewing practices, alternative, open-ended interviewing strategies have been developed to enhance the quality, completeness and accuracy of children's eyewitness reports.

The Cognitive Interview

Cognitive load questions used in the CI have shown to be effective mnemonic tools for increasing the amount of important information recalled by adult eyewitnesses (see Bull et al., 2019; Holliday et al., 2009; Memon et al., 2010, for reviews). The CI, originally developed by Geiselman (1984), is based on memory and cognitive processing research on remembering and retrieval of information. A revised version of the CI was later created (Fisher and Geiselman, 1992), which was adapted in the two current studies, that placed more emphasis on the social and communicative dynamics between the interviewer and respondent (e.g., rapport building, transfer of control and nonverbal responding). Cognitive load questioning is based on two main principles of memory: (1) encoding specificity and (2) varied retrieval (McCauley & Fisher, 1995). The encoding specificity principle (Tulving & Thomson, 1973) argues that memory recall is enhanced when information that was available at the initial encoding (e.g., sensory cues) is also present during the retrieval stage. Whereas, varied retrieval (Tulving, 1974) posits that a

person's recollection accuracy is highly influenced by the specific retrieval cues they use when searching through their memories. While a person typically encodes an experience only once, they can use multiple retrieval cues (e.g., recalling different senses, perspectives and temporal orders) to enhance their recollection of this initial encoded information. Based on these memory principles, three cognitive load strategies were developed for the CI to encourage truthful eyewitnesses to recall more information about an event, without decreasing recall accuracy.

First, a *mental context reinstatement* question asks respondents to reconstruct the event in question according to their sensory experiences (e.g., visual, olfactory, auditory and tactile cues) and thoughts at the time of the event. While standard free-recall questions ask respondents to recall only visual (e.g., environmental facts) and auditory (e.g., conversations) details, memory research also highlights the benefits of inquiring about other senses as well. Recalling sensory information activates different areas of the brain (Manning, Polyn, Baltuch, Litt, & Kahana, 2011). For instance, olfactory information gets processed by the olfactory bulb and then gets relayed to the amygdala and hippocampus in the limbic system (Sullivan, Wilson, Ravel, & Mouly, 2015); these brain systems are associated with memory processes, as well as instincts, basic emotions and mood. In contrast, visual information is processed through the posterior parietal cortex of the parietal lobe (Berryhill & Olson, 2008), which is responsible for the spatial, attentional and imagery components of memory. In one notable study (Hershkowitz, Orbach, Lamb, Sternberg, & Horowitz, 2001), victims of sexual abuse (ages 4 to 13) who were interviewed using mental context reinstatement techniques demonstrated improved memory recall in their eyewitness reports.

Second, respondents are asked to recall an event from *multiple perspectives* whereby they discuss everything they remember from their own perspective, and that of someone else (e.g.,

parent). On standard free-recall questions, people are typically asked to recall memories from a first-person, or autobiographical, perspective whereby they retrieve and express past experiences from their own point of view. On the other hand, recalling events from *another perspective* (third-person recall) involves imagining past experiences from the external point of view of someone else (Paulo, Albuquerque, Saraiva, & Bull, 2015). Third person memory recall can affect how someone thinks and feels about a memory by increasing the emotional intensity (Berntsen & Rubin, 2006) and vividness (Rice & Rubin, 2009) of the experience in question.

Third, respondents are asked to recall events in *different temporal orders*, including chronological (i.e., forward) and reverse-order. Forward recall consists of schema-based retrieval, whereby the respondent simply recalls the events in the order in which they occurred; this results in stronger recall of consistent information that is more easily recalled (Geiselman & Callot, 1990). Conversely, reverse-order recall incorporates more data-driven (or non-schema-based) retrieval as the respondent is recalling the events backwards, starting with the last event and ending with the first event. Data-driven recall can lead to better recall of incidental information that is not as readily accessible in memory (Geiselman & Callot, 1990), such as information that the respondent was not directly asked to discuss by the interviewer.

Adults typically give longer (i.e., more words) and more detailed eyewitness reports in response to cognitive load questions, without significantly compromising accuracy (Bull et al., 2019; Memon et al., 2010). The cognitive load questions significantly increased the number of correct details recalled, such as relevant information about the people, places, objects and actions at the time of the event. Despite this increased recall, there were only minimal increases in recollection errors (Memon et al., 2010). In addition to increased memory recall performance for honest speakers, these questions have also shown to discourage lie-telling in adults by

increasing the cognitive effort associated with maintaining a false story (Vrij, Granhag, Mann, & Leal, 2011; Vrij, Leal, Mann, & Fisher, 2012). Whereas truth-tellers simply recall events from memory, lie-telling requires more cognitive effort as liars have to simultaneously manage their verbal and non-verbal behaviours, and assess the mental state of the lie-recipient to convince them of the false information (DePaulo et al., 2003). Thus, lie-tellers are more likely to produce inconsistent statements in response to the stressful cognitive load questions (Colwell et al., 2013; Vrij et al., 2012), which can discourage them from continuing with their initial lie-attempt.

Evaluating Testimony Performance

Both studies in this research program will provide one of the most comprehensive evaluations of the CI with typical and atypical child populations. The majority of CI studies with children (e.g., Geiselman & Padilla, 1988; McCauley & Fisher, 1995; Memon, Cronin, Eaves, & Bull, 1993), particularly those with disabilities (Gentle et al., 2013; Milne & Bull, 1996; Robinson & McGuire, 2006), have evaluated testimony performance according to the number of words and details recalled (e.g., environmental information), along with the accuracy (e.g., distortions and confabulations) of this information. While these measures are important, there are other impactful indicators of testimony performance that have been overlooked.

Report Honesty. Children's overall *Report Honesty* will be evaluated in both studies to assess whether the CI and SI increase children's willingness to truthfully report the transgression of an adult. Much is known about children's unintentional false reports, including false memories (e.g., Howe, Toth, & Cicchetti, 2011) and suggestive responses to unreliable interviewing strategies (e.g., Lewy, Cyr, & Dion, 2015). Nevertheless, children often intentionally fail to disclose the crimes of others (Smith et al, 2000), which can have serious implications for the victim and future child eyewitnesses. There is considerable literature on the

development of children's lie-telling (see Talwar & Crossman, 2011, 2012, for reviews), but there is far less information available to forensic professionals regarding how interviewers can discourage dishonest reporting. Finally, considerable research with adults suggest that truth-tellers and lie-tellers respond differently to cognitive load questions (see Colwell et al., 2013, for a review) due to the increased cognitive effort associated with deceiving another (Vrij et al., 2011, 2012). The two studies in this research program will provide forensic professionals with new information about whether these questions can be used to deter and detect false reports.

Response Length. Children's *Response Length*, or the number of words used, on each question and throughout the interview will be examined in both studies. Response Length is used in the Statement Validity Analysis (SVA) literature, an evidence-based tool used in some North American and European courts for evaluating eyewitness credibility (Vrij, 2005), as a measurement of the amount of information a witness is able to recall (Colwell et al., 2007, 2013). However, a limitation with this measure is that it does not evaluate the specific content being discussed by the witness. For example, Wyman and colleagues (2019) found that child truth-tellers and lie-tellers had similar Response Lengths when interviewed using the CI. Nevertheless, there were notable differences in the actual content being discussed in their reports (Wyman et al., 2019). For this reason, the third and fourth dependent measures in both studies will assess for interview (CI vs. SI statements) and veracity group (honest vs. dishonest reports) differences in the actual content of children's statements.

Transgression Details. The specific *Transgression Details* discussed by the TDC and CWID will analyzed. SVA coding procedures track every new (or unique) relevant detail about an event disclosed by the witness (Colwell et al., 2013). Based on the SVA literature, truth-tellers add new relevant information about an event (e.g., the specific people present during a

crime, along with their actions) because they are only tasked with recalling information from memory. Conversely, lie-tellers rarely add new relevant information about the event because they are trying to maintain a consistent lie-script. Instead, they discuss irrelevant experiences and avoid discussing the actual event in question (see Colwell et al., 2007, 2013, for reviews). Thus, this dependent measure will provide insight into whether the CI encourages increased discussion and memory recall by TDC and CWID of details about the transgression of another.

Events Recalled. Both studies will also track the number of events children recall about their respective experiences with the transgressor. This measure differs from Transgression Details because it assesses how many different events a child discusses (e.g., each magic trick), rather than just the individual details about the transgression. This measure is particularly relevant to CWID, who often struggle to provide specific details due to limitations with their vocabulary and expressive language skills (Wyman, Lavoie et al., 2018). Thus, CWID may be stronger at recalling broad information about each event, instead of highly specific details.

Many forensic studies using SVA coding typically only calculate new information that the witness discloses (Colwell et al., 2007; 2013; Wyman et al., 2019). However, it is also important to code for information that is *repeated* by the witness. Events or specific details that are repeated can indicate that this information is important to the witness given that they are discussing it several times. For these reasons, both studies will calculate the *unique* and *repeated* information (Transgression Details and Events Recalled) disclosed by each child. A *total score* will then be calculated that combines the unique and repeated scores for both these measures.

Temporal Order Accuracy. Prior eyewitness research has evaluated the accuracy of the specific details children and adults disclosed (e.g., Geiselman & Padilla, 1988; McCauley & Fisher, 1995). While this provides important insight into the accuracy of children's memory

recall, this method can also be problematic because it can be challenging to reliably clarify which details are correct or incorrect. For example, memory recall can be influenced by the perspective of the witness, such as their moral evaluations of the event that is informed by their prior learning and experiences; thus, recollections of the same event can be drastically different between people based on how it is interpreted (i.e., The Rashomon Effect; Heider, 1988).

Both the current studies assessed *Temporal Order Accuracy*, or serial order memory accuracy, of the events the TDC and CWID recalled to provide a more reliable assessment of children's recall accuracy. This measure has been used in memory studies (e.g., Jack, Friedman, Reese, & Zajac, 2016; Mirandola & Toffalini, 2018) to assess several important functions of memory. As discussed in Brown and Smith-Petersen (2014), this measure provides insight into a person's memory performance with respect to: (1) *duration* (i.e., the time extent of an experience), (2) *order* (i.e., sequencing correctness of a series of events), (3) *successiveness and simultaneity* (i.e., the temporal relation between two or more events), and (4) *chang*e (i.e., time shifting of the events). Within both studies, it was possible to score Temporal Order Accuracy of the participants' recalled events since they engaged in the same events in identical order.

Disclosure Forthcomingness. Some recent forensic studies have analyzed how quickly, or early in the interview, children make a clear disclosure (Malloy & Mugno, 2016; Wyman et al., 2019). Children frequently delay their disclosure of another's transgression for an extended period of time (Hershkowitz et al., 2006; Smith et al., 2000). Thus, this dependent measure provides insight into whether a specific questioning strategy encourages children to make a clear true (e.g., accuse the transgressor) or false (i.e., falsely deny a person's crime) disclosure earlier in the interviewing process. Interviewing strategies that elicit more forthcoming disclosures

have the potential to improve the speed with which investigators respond to potential maltreatment, as well as provide the victim with more immediate safety, support and care.

Disclosure Frequency. It is also important to examine the number of times a child makes a disclosure on each question-type. In Wyman and colleagues (2019) for example, only a child's first (or new) disclosure was coded. Nevertheless, like the repeated Transgression Details and Events Recalled measures, recording every unique and repeated disclosure allows interviewers to better understand which questioning strategies encourage increased discussion about a transgression a child witnessed.

Testimony Consistency. Prior deception literature has examined the consistency of a child's closed-ended responses as a measure of how well they can maintain a true or false report (see Talwar & Crossman, 2011, for a review). The common method of measuring children's disclosure maintenance abilities has incorporated the use of up to three follow-up questions to evaluate whether a child's initial truth or lie was properly maintained throughout the questioning (e.g., Talwar, Gordon, & Lee, 2007; Talwar & Lee, 2002a; Talwar & Lee, 2008). However, these findings may not fully generalize to real-life settings wherein child witnesses are interviewed. In the commonly used forensic interviewing procedures, children answer a large range of open-ended and closed ended questions; thus, a child must maintain their true or false report across several follow-up questions. The use of lengthier interviews in both studies, including eight Testimony Consistency questions, will permit for a more thorough and generalizable evaluation of children's abilities to produce consistent true and false reports.

Present Investigation

The first study in this research program examines whether the CI, when compared to the SI, increases the performance of TDC on eight dependent measures of testimony performance:

(1) Report Honesty; (2) Response Length; (3) number of Transgression Details; (4) number of Events Recalled; (5) Temporal Order Accuracy; (6) Disclosure Frequency (7) Disclosure Forthcomingness; and (8) Testimony Consistency. Similarly, the second study assesses the efficacy of the CI across these eight dependent measures, but in a sample of CWID. Both studies then examine whether the CI can be used to distinguish between children's true and false reports when SVA coding procedures are used.

The prospective findings of these two studies are designed to benefit professionals who investigate crimes against children. By analyzing interview strategies beyond the commonly used free-recall question, interviewers will acquire valuable information about additional tools that can be used to increase the quality and quantity of children's eyewitness reports. At the same time, this research paradigm will generalize to real-life situations wherein children witness or experience crimes; thus, while these two studies are experimental in nature, the topics children are discussing in their eyewitness reports correspond to actual forensic settings. Interviewers frequently encounter cases whereby children intentionally omit or falsely deny the transgressions of others (e.g., Smith et al., 2000). While, cognitive load questioning makes lie-telling more challenging for adults (Vrij et al., 2011, 2012), it is unclear if these questions have the same effects with children. By using a research paradigm whereby children feel pressure to falsely deny the transgression of another, forensic investigators can better understand whether the CI can also serve as an effective tool for discouraging false reports. Furthermore, parents and other professionals who work with children (e.g., teachers, guidance counselors, social workers and psychologists) can also benefit from these findings. Namely, both studies can potentially inform them of ways to discourage everyday lying in children, as well as the best methods for asking children nonsuggestive questions about difficult topics.

Chapter 3: Manuscript #1

The Cognitive Interview: The effectiveness of cognitive load questioning when children provide true and false eyewitness reports of another's transgression

Joshua Wyman[1], Donia Tong[1], Ida Foster[1], Angela Crossman[2], & Victoria Talwar[1]

[1]Department of Educational and Counselling Psychology, McGill University

[2]Department of Psychology, John Jay College of Criminal Justice, City University of New York

Author Note: This manuscript is currently under review at *Legal and Criminological Psychology*.

Abstract

The current study was designed to provide forensic professionals with much needed information about the efficacy of a range of interviewing strategies in cases where children are coached to deceive an interviewer. After being asked to falsely deny a theft they witnessed, children (ages 8 to 13; $N = 104$) were interviewed by a second researcher using either the Cognitive Interview (CI; $n = 52$) or a Standard Interview (SI; $n = 52$). Two raters then coded each child's: (1) report honesty; (2) response length; (3) number of transgression-specific details disclosed; (4) number of events recalled; (5) temporal order accuracy; (6) disclosure frequency; (7) disclosure forthcomingness; and (8) the overall testimony consistency. When compared to the SI, children produced longer testimonies with more transgression details, disclosures and events on the CI, without compromising temporal order accuracy or testimony consistency. Within the CI, free-recall reports (initial and re-tell) had the most words (response length), along with more unique transgression details, disclosures and events recalled. At the same time, statements on the reverse-order cognitive load question had significantly more repeated transgression details and events than the other questions. Significant narrative differences were also found between the honest and dishonest speakers. Truth-tellers in the CI group gave more forthcoming and detailed eyewitness reports when compared to dishonest speakers. Finally, girls in the CI group had significantly higher scores on the response length, transgression details, disclosure frequency and events recalled dependent measures when compared to the boys.

The Cognitive Interview: The effectiveness of cognitive load questioning when children provide true and false eyewitness reports of another's transgression

Beyond free-recall questioning, which is widely encouraged by those involved in the investigation of crimes against children (e.g., Guadagno et al., 2006; Orbach et al., 2000), it is unclear whether other questioning strategies can be used to further elicit detailed and accurate disclosures from this population. Cognitive load questions, which are a key component CI (Fisher & Geiselman, 1992), have shown to be effective tools for increasing the quality and quantity of information in adults' testimonies (see Bull et al., 2019; Memon et al., 2010, for reviews). Nevertheless, the existing literature on the benefits of cognitive load questioning with children has produced both positive (e.g., Gentle et al., 2013; McCauley & Fisher, 1995) and inconclusive results (e.g., Saykaly, Crossman, Morris, & Talwar, 2016; Wyman et al., 2019).

The limited cognitive load questioning research with children to date typically has had the eyewitness truthfully recall a low-stakes event (e.g., Gentle et al., 2013; Mattison, Dando, & Ormerod, 2014; McCauley & Fisher, 1995). However, actual child eyewitnesses are interviewed about the potentially traumatic crimes of others that they know, and thus, they often attempt to conceal the transgression (London, Bruck, Ceci, & Shuman, 2005; Smith et al., 2000). To adequately assess the efficacy of specific questioning strategies, it is therefore necessary for experimental forensic research to incorporate a generalizable target event whereby participants witness a serious transgression, and at the same time, feel internal and external pressure to conceal the crime. The aim of the current study was to evaluate the efficacy of the CI, in comparison to a Standard Interview (SI), in cases where children were asked to falsely deny the crime of an adult. The design of the study was to create, within ethical bounds, an analogue to situations where children may be coached to conceal the transgressive behaviour of another.

Cognitive Load Questioning with TDC

The revised version of the CI (Fisher & Geiselman, 1992) used in the current study consists of a chronological free-recall question and four cognitive load questions. A standard *Free Recall* question (n = 1) is asked at the beginning of the interview to encourage the witness to describe in as much detail as possible everything they remember, in chronological order, about the event in question. Substantial forensic research supports the use of free-recall questions, instead of closed-ended recognition questions, for encouraging eyewitnesses to provided detailed recollections of an event in their own words (e.g., Quas et al., 2000; Wyman et al., 2019). However, standard free-recall questions can be overly broad in nature, which can result in uninformative reports (Lyon, 2014). Moreover, deceptive witnesses are more likely to intentionally mislead interviewers on this question (Pipe & Wilson, 1994). Given these limitations with free-recall reports, four cognitive load questions were developed to further enhance eyewitness memory recall, while at the same time, discourage false reports.

First, a *Mental Context Reinstatement* question inquires about all the senses the witness experienced at the time of the event, including their sights, smells, auditory memories, tastes, thoughts and things they touched. Recalling multiple senses experienced at the time of an event results in increased activation of important brain functions, such as the amygdala and hippocampus, involved in memory processing and recall (Manning et al., 2011). Second, witnesses are asked to recall their experiences according to *Another Perspective*. Recalling experiences from multiple perspectives, including first-person and third-person memory recall, has shown to increase the vividness (Rice & Rubin, 2009) and emotional intensity (Bernsten & Rubin, 2006) of a memory. Third, a *Reverse-Order Recall* question involves freely recalling an event, but in reverse-order. In addition to enhancing memory recall of incidental information for

honest speakers (Geiselman & Callot, 1990), this question has been found to be especially effective at increasing the cognitive effort associated with telling and maintaining a false report (Vrij et al., 2008). Fourth, an additional chronological-order free-recall question (i.e., a *Re-tell* question) is asked to promote increased memory recall. Finally, open-ended *prompts* (e.g., "Can you tell me more?") are asked after each free-recall and cognitive load question to encourage the witness to further elaborate on their prior statements (Hershkowitz et al., 2012).

The CI has shown to be very effective with adult respondents for increasing the amount of new information disclosed (see Memon et al., 2010, for a metanalytic review), as well as for discouraging false reports due to the cognitively taxing nature of these questions (Vrij et al., 2011, 2012). However, these strategies have produced mixed results with children (see Holliday et al., 2009, for a review). When analyzing children's truthful reports, some studies found that the CI increased eyewitness recall performance (e.g., Gentle et al., 2013; McCauley & Fisher, 1995). Conversely, other studies found that the cognitive load questions did not significantly improve the quantity or quality of children's eyewitness reports (Saykaly et al., 2016), especially for those under the age of 8 (Memon et al., 2010; Wyman et al., 2019). To better understand the value of cognitive load questions with TDC, the current study will address three notable gaps in the forensic literature.

Developmental considerations. In adult eyewitness studies with typical populations, participants often have commensurate developmental characteristics. However, there is more variability in the developmental characteristics of children, especially when a broad age range is used. For example, the ability to produce detailed (e.g., Jack et al., 2014; Wyman et al., 2019), accurate (Eisen, Goodman, Qin, Davis, & Crayton, 2007) and persuasive (Talwar & Lee, 2008) recollections increases with child age and cognitive development. Some CI studies have

assessed for age and verbal test score differences on the CI (e.g., Wyman et al., 2019), but it is unclear whether other developmental factors should also be considered.

Prior deception research indicates that the ability to tell consistent and convincing lies across multiple questions is related to a child's cognitive development, namely their improving inhibitory control and theory-of-mind capabilities (see Talwar & Crossman, 2011, 2012, for reviews). To tell a persuasive story, whether true or false, a person must use inhibitory control to restrain from giving any information that contradicts their narrative (Vrij et al., 2005). At the same time, a speaker must consider the perspective of the interviewer when trying to convince them with their report (Talwar & Crossman, 2011, 2012). More specifically, the speaker must utilize their theory of mind and perspective taking skills to effectively monitor the verbal (e.g., tone of voice) and nonverbal behaviours (e.g., eye-contact and body language) of the interviewer to ensure that they are telling a clear and convincing narrative. The current study will assess whether inhibitory control and theory of mind functioning, in addition to age and verbal ability, are related to children's performance on the CI and SI.

Some adult CI research suggests that women produce longer and more coherent testimonies on the CI than men (Suckle-Nelson et al., 2010). Moreover, police officers and mock-jurors are more likely to believe the testimonies of female (adult and child) eyewitnesses (Mann et al., 2004; Talwar et al., 2006; Wyman, Foster et al., 2018). Developmental research also suggests that there are gender differences in communication. Namely, girls have been found to outperform boys on tests of verbal ability (Hyde, 2016; Petersen, 2018), scholastic language aptitude (Voyer & Voyer, 2014) and interpersonal communication (Hall & Matsumoto, 2005; Hampson et al., 2006). Given the potential for gender differences in testimony performance, the current study will assess whether girls and boys perform differently on the CI and SI.

Question-specific benefits. Few studies to date have compared the efficacy of the individual cognitive load questions. Most studies have evaluated the CI as a whole when compared to a SI (e.g., Akehurst, Milne, & Kohnken, 2003; Gentle et al., 2013) or individual cognitive load questions, such as only studying the efficacy of mental context reinstatement (Hershkowitz et al., 2001) or reverse-order recall questions (Sayakly et al., 2016; Vrij et al., 2012). But a recent study by Wyman and colleagues (2019) suggests that some cognitive load questions may be more effective than others. Child truth-tellers and lie-tellers (accusers and deniers) in their study produced the most detailed disclosures about a theft in response to a free-recall question. The reverse-order and re-tell questions encouraged children to recall new information about the theft that had not been previously discussed, especially for those who truthfully accused the researcher of a theft. In contrast, the mental context reinstatement and another perspective questions had minimal benefits as children typically recalled the least amount of information on these questions (Wyman et al., 2019). The current study will expand on Wyman and colleagues (2019) by comparing the CI questions to those in a SI, which does not include cognitive load questions. Comparing the CI to a SI will permit for a greater understanding of the benefits of the CI in its entirety, while also providing important information on the efficacy of the individual cognitive load questions.

Generalizable research paradigm. Forensic interviewers typically question children about novel and stressful events whereby the eyewitness is most often the observer and/or victim of a transgression. Additionally, child victims often fail to disclose another's transgression in cases of witness coaching, even when free-recall questioning is used (Pipe & Wilson, 1994; Wyman et al., 2019). Despite this important concern, there is relatively little deception research,

beyond promising to tell the truth (Lyon & Dorado, 2008), that specifically examines ways to improve the honesty and accuracy of children's reports.

Among the few deception eyewitness studies, the transgression of another is very brief and the child is normally just an observer. In Wyman and colleagues (2019), children simply observed a researcher unexpectedly steal money from a wallet; the length of the transgression (< 1-minute), and the child's lack of involvement correspond to other deception studies (e.g., Pipe & Wilson, 1994; Talwar, Lee, Bala, & Lindsay, 2004). In cases of crimes against children, however, the transgressions children report on (e.g., physical and sexual abuse) can take place for several minutes, or even hours. Rather than simply being bystanders or victims, children often feel complicit in cases of maltreatment (Romeo, Otgaar, & Landstrom, 2018). To improve the overall generalizability of the findings to actual investigations of child maltreatment, children in the current study will witness a prolonged transgression that lasts several minutes in which they and another bystander indirectly participates. Afterwards, children will be asked to conceal the crime by the transgressor prior to being interviewed by an unfamiliar adult using the CI or SI. This research paradigm can better generalize to real life situations in which a child feels internal and external pressure to deny the criminal behaviour of someone they know when being interviewed by an unfamiliar police officer or forensic interviewer.

Present Study

The current study was designed to provide a comprehensive evaluation of the benefits and limitations of different cognitive load questions when children are asked to provide detailed eyewitness reports about an adult's transgression. Children (N = 104) between the ages of 8 and 13 years, with typical cognitive functioning, witnessed (and indirectly participated in) a prolonged transgression involving the theft of money from a stranger's wallet by a researcher

(E1). Children were then asked to falsely deny the theft by E1. Afterwards, children were interviewed by another researcher (E2) using either the CI (n = 52) or SI (n = 52) protocols.

The first objective of this study was to assess for potential developmental factors, including age, gender and performance on cognitive measures of inhibitory control, verbal production and theory of mind, that may impact a child's performance on the following eight dependent measures of testimony performance: (1) *Report Honesty* (i.e., whether child told a truth or lie about the theft), (2) *Response Length* (i.e., number of words disclosed), (3) *number of Transgression Details* (i.e., unique, repeated and total theft-related details), (4) *Events Recalled* (i.e., number of unique and repeated experiences with E1), (5) *Temporal Order Accuracy* (i.e., accuracy of the order in which children recalled the events with E1), (6) *Disclosure Frequency* (i.e., willingness to report the theft to E2), (7) *Disclosure Forthcomingness* (i.e., how quickly children disclose the theft), and (8) *Testimony Consistency* (i.e., the consistency of children's closed-ended responses throughout the interview). Based on prior developmental research (Jack et al., 2014; Memon et al., 2010; Wyman et al., 2019), performance on dependent measures two through eight is expected to increase with child age and cognitive test performance. On the other hand, child age is not expected to be a significant predictor of overall testimony veracity. Finally, girls are expected to produce longer and more detailed testimonies than boys, irrespective of interview condition and testimony veracity, given that they typically perform better on verbal tasks (Hyde, 2016; Petersen, 2018; Voyer & Voyer, 2014).

To assess the benefits of the cognitive load questions with children, the second objective of this study was to examine whether the CI, when compared to the SI, increased child performance across the eight dependent measures of testimony performance. Prior research with adult respondents suggests that the cognitive load questions provide additional opportunities to

openly recall an event, while utilizing mnemonic strategies to improve recall and discourage dishonesty (see Memon et al., 2010, for a review). Thus, it is expected that those interviewed using the CI would provide more honest, informative, detailed and forthcoming theft reports with more events discussed. Children in the CI group are also expected to disclose the theft with greater frequency given that they will have more opportunities to openly recall the events from their time with E1. Due to the mentally taxing nature of the cognitive load questions (Vrij et al., 2011, 2012), children interviewed using the CI are nevertheless expected to have lower overall Temporal Order Accuracy and Testimony Consistency in their reports compared to those interviewed with the SI.

The third objective of this study was to compare the efficacy of the individual free-recall and cognitive load questions. The free-recall question is expected to elicit the most honest reports, with more information and unique Transgression Details and Events Recalled, since it is the first question asked in each interview. The reverse-order and re-tell questions are also expected to encourage children to recall more unique (i.e., new) and repeated (i.e., previously stated) Events Recalled with E1, as well as more Transgression Details (Wyman et al., 2019). However, in accordance with the concerns highlighted in prior research (Saywitz, Geiselman, & Bornstein, 1992; Wyman et al., 2019), the mental context reinstatement and another perspective questions are expected to be less effective due to the increased cognitive sophistication required to benefit from these questions. Finally, Temporal Order Accuracy is expected to be lower on the cognitive load questions due to the mentally taxing nature of these questions.

The fourth objective of this study was to examine differences in the testimony content of truth-tellers and lie-tellers across the dependent measures (2 to 5, 7 and 8). Based on research with adults (Colwell et al., 2007, 2013) and children (Wyman et al., 2019), truth-tellers are

expected to produce statements that are more forthcoming, longer and detailed (Transgression Details and Events Recalled) about the theft. In contrast, lie-tellers will omit the theft in their open-ended responses, and instead, rely on the closed-ended questions to falsely deny the theft (Wyman et al., 2019). Lie-tellers are also expected to have lower Temporal Order Accuracy and Testimony Consistency on the CI due to the increased cognitive effort required to tell and maintain a plausible lie (Christ, Van Essen, Watson, Brubaker, & McDermott, 2009).

Method

One-hundred and four children, between the ages of 8 and 13 years (*Mage* = 10.52, *SD* = 1.67), were recruited from a large metropolitan area (i.e., population > 3,000,000) to participate in the current study. This study received ethics approval from the residing university, and parents provided verbal and written consent to participate. While the parents were aware of the true nature of the study, they nevertheless were asked by the recruiter to not discuss the theft, request to lie, or follow-up interview with their children prior to the study. Instead, parents were simply asked to tell their children that they were going to be doing some school work with a university student. Children were tested individually, and no participant stated that they were aware of the experimental nature of this study during the testing. Parents remained in a private waiting room and did not interact with their child for the duration of the study.

Participants were randomly placed by E1 into either the Cognitive Interview (CI; *n* = 52, *Mage* = 10.26) or Standard Interview (SI; *n* = 52, *Mage* = 10.77) groups prior to the study. E1 and the interviewer (E2) did not change the selected interview group at any point during the study after meeting a child. Furthermore, child gender was coded dichotomously (boys versus girls) as no parent selected the “Other” gender category on the standard demographic form they completed.

Materials and Procedures

Rapport building and cognitive tasks. Two rapport building activities, including naming three things about each other and asking each other three questions, were completed for the purpose of increasing children's rapport with E1. Three subtests from the Developmental Neuropsychological Assessment- Second Edition (NEPSY-II), a widely used neurocognitive assessment instrument for children aged 3 to 16 years (Korkman, Kirk & Kemp, 2007), were then administered to evaluate children's inhibitory control, verbal productivity and theory of mind skills. The subtests in the NEPSY-II have appropriate reliability and validity to be used with children as a comprehensive measure of neuropsychological development (Davis & Matthews, 2010). Overall, the three cognitive tasks took between 25-and 35-minutes to complete. Higher scores represented better performance.

The *Inhibition* (IN) subtest consisted of three tasks that assessed children's ability to inhibit automatic responses in favor of novel responses, as well as their ability to switch between both response types. The *Naming* task required children to name a series of shapes (Item 1) or the direction of different arrows (Item 2) as quickly as possible. The *Inhibition* task asked children to say the "opposite" shape (Item 1) or arrow direction (Item 2) to what they saw. For the *Switching* task, children were asked to mentally switch between response types in accordance to whether the shape and arrow direction were white or black. For example, when children saw a black shape, they were asked to say the shape's name; however, if they saw a white shape, they said the name of the other shape. For each task, children were evaluated using four criteria. First, children received a standard score for each of the three Inhibition subtests, ranging from one to 20, which combined their completion time and response accuracy (i.e., self-corrected + uncorrected errors) performance when compared to a normed sample of similar aged peers.

Second, children's scores on the three tasks were added together to achieve a total IN score (out of 60). Third, children received a separate *Total Error* standard score, ranging from 1 to 20, which represented their total number of uncorrected and self-corrected errors across the three tasks when compared to a normed sample of similar-aged peers.

The *Word Generation* (WG) subtest consisted of two tasks that evaluated children's verbal production skills. The *Initial Letter Word Generation* task required children to name as many words that began with the Letter F as possible within a 60-second period, as well as words that began with the Letter S. The *Semantic Word Generation* task required children to name as many animals as possible within a 60-second time limit, and then, as many foods or drinks. Children received a standard score for each task, ranging from one-to 20, that was based on the total number of unique and correct words they provided when compared to a normed sample of similar aged peers. A total WG score out of 40 combined the children's scores on the two tasks.

The *Theory of Mind* (ToM) subtest consisted of two tasks that examined children's perspective taking skills for understanding the feelings, perceptions and intentions of others. For the *Verbal* task, children answered questions (n = 15) about a series of vignettes whereby they were asked to think from the perspective of another. The discontinue rule for this task was four consecutive scores of 0, and the total possible score on this task was 22. On the *Contextual* task, children were shown six pictures depicting various social situations in which the face of the target individual was not shown. Children were then asked to select a photo from four options that depicted the appropriate affect for the target individual. The total possible score for this task was 6. A total ToM raw score out of 28 combined the children's scores on these two tasks.

Initial wallet discovery. E1 and the children completed a filler activity (i.e., cup shuffle guessing game) for 2-minutes. During the game, a third researcher (E3) entered the room to find

his/her missing book. E3 greeted the children and introduced him/herself. S/he also asked if E1 had seen his/her book, to which E1 replied "no". E3 then found a wallet while looking for the book behind a desk. E3 did not open it, but s/he asked both E1 and the child separately if either of them knew to whom the wallet belonged. After E1 and the child said "no", E3 placed the wallet on the desk and left the room. The children and E1 continued the filler activity for an additional three-to five-minutes.

Theft. After the filler activity, E1 stated that s/he needed to go to another room to complete the next activities. Before leaving, E1 stated that s/he needed to clean up the room and grab his/her jacket that was located on the desk near the wallet. While grabbing the jacket, E1 picked up the wallet and showed it to the children. To include the children on the interaction with the wallet, E1 asked them to look through the wallet to find out who it belonged to while E1 cleaned up the testing materials. Inside the wallet, the children found a bus ID of someone named Sarah, along with the picture of this person. E1 stated that s/he did not know who this person was. E1 then proceeded to look through the wallet; s/he then took out a twenty-dollar bill and showed it to the children. E1 then said, "I'm going to take it", and s/he put the money in his/her back pocket. E1 ensured that the children were looking at them during this interaction to guarantee that they saw the money and the subsequent theft. E1 and the children then went to another room to complete the other activities.

Lie request. In next room, E1 and the children engaged in a cooperative filler activity (i.e., building a tower with blocks). The interviewer (E2) then entered the room, introduced him/herself to the children, and asked to speak with E1 privately. E1 then left the room with E2 for 2-minutes while the children completed the rest of the activity by themselves. When E1 returned, s/he notified them that E2 was aware of the missing money from the wallet. E1 then

asked the children if they remembered the theft; all the children recounted that E1 had taken the money from the wallet. E1 then stated that s/he took the money because s/he "really needed it", and s/he asked the children to say that E1 did not take it if E2 asked about it. Children did not have to agree to the request, but E1 asked the children to repeat the lie request to ensure that they understood what was asked of them. Overall, 79.8 % (n = 83) of children told E1 that they would deny that E1 took the money, with there being no significant differences in the pre-interview agreement rate between the CI (75%) and SI (84.6%) groups, $\chi^2(1) = .22$, $p = .329$. After, E2 entered the room and asked to speak to the children privately. E1 then left the room.

Interview. In line with best practice child interviewing protocols, such as the NICHD protocol (Orbach et al., 2000), E2 introduced the CI and SI interviews by stating that there were no right or wrong answers to their questions, and that children should answer each question honestly and to the best of their ability. Children were also told that they could stop the interview at any time. Every E2 was trained by the primary investigator to only ask the questions, as written, in each interview protocol. E2 was also trained to maintain a neutral disposition, and to refrain from providing any verbal or nonverbal (e.g., nodding or shaking head) behaviours that could reinforce or discourage particular response types.

Children were interviewed using either the CI (n = 52), which included developmentally-appropriate modifications to the wording of the cognitive load questions, or the SI (n = 52) that did not include the cognitive load questions. A *Free Recall* question (n = 1) was asked in both interviews to allow children to describe in as much detail as possible everything they remembered from their time spent with E1. In the CI group only, cognitive load questions (n = 4) were used to enhance the memory of the interviewees. First, children were asked to recall the senses they experienced (sights, smells, auditory memories, taste and things they touch) during

their time with E1 (*Mental Context Reinstatement* question). Second, children recalled their experiences with E1 from the perspective of their parent (*Another Perspective* question). Third, children described everything they remembered about their time with E1, but in reverse-order (*Reverse-Order Recall* question). Fourth, children were asked to provide a final free-recall recollection about their time with E1 (*Re-tell* question). One *prompt* ("Is there anything else?") was asked after the free-recall and cognitive load questions to encourage the children to elaborate on their prior statements.

In both interviews, children were asked *Non-direct closed-ended* questions (n = 18) whereby they were asked to provide short-answers regarding the wallet situation (e.g., "Did someone take money from the wallet?") and E1's (e.g., "Did E1 do something they were not supposed to do?") and E3's general behaviour; however, children were only questioned about the specific identity of the transgressor on the final *Direct Inquiry* question (i.e., "Did E1 take money from the wallet?"). Table 1 displays the questions in both interview protocols. As seen in Table 2, there were no significant developmental differences between the CI and SI groups, except for ToM Total [$F(1) = 5.67$, $p = .019$] and ToM Verbal Task [$F(1) = 8.95$, $p = .003$] scores.

Debrief. E2 thanked the children at the end of the interview, and they were debriefed about the true nature of the study by E1 and E2. During the debrief, children were told that E1 should not have taken the money or asked them to lie; but, by participating in this study, they helped kids in real life when they are questioned by adults about difficult topics. Finally, children were told that it is not okay if an adult asks them to lie, and that they should always tell the truth to an adult they trust if they witnessed a transgression, like their parent(s) or a teacher. All participants and their parents reported no adverse feelings following the conclusion of the study.

Coding Procedures

Interviews were video recorded and transcribed to assess for possible narrative differences between the CI and SI. The coding protocols for this study were based on some of the measures used by the Assessment Criteria Indicative of Deception (ACID), which is an SVA tool that is used for detecting systematic and measurable differences in the content of true and false eyewitness reports (see Colwell et al., 2013, for a review).

First, *Report Honesty* was assessed two ways. Children's *Attempts* to deceive the interviewer was determined by whether they omitted or directly denied the theft at any point during the interview (0 = Yes; 1 = No). Next, children's *Final Decision* to deceive the interviewer was based on their response to the Direct Inquiry question by asking if E1 took the money (0 = Lied; 1 = Told the truth). No child falsely accused E3 of taking the money when asked about that researcher's involvement with the theft.

Second, the total number of words (*Response Length*), as calculated by a standard word processing program, children disclosed on the free-recall and cognitive load questions was calculated. Third, the number of specific details about the wallet situation (*Transgression Details*) that were disclosed on the free-recall and cognitive load questions were calculated to determine children's willingness to discuss the alleged crime. Transgression Details included any piece of information that related to the situation with the wallet; thus, each unique (i.e., not previously stated) and repeated (i.e., previously stated) detail about the theft situation was scored. The Total Transgression Detail score for the free-recall and cognitive load questions was calculated by adding the unique and repeated transgression detail scores together. For example, the sentence, "We played these games with shapes, and E1 took the money. No wait, we also played story games, and then E1 took the money." contains 24-words (Response Length), three

unique Transgression Details and three repeated Transgression Details (6 total details). Two coders, who were blind to the experimental condition, child age and gender, coded the Transgression Detail scores with very high reliability (ICC = .90, $p < .001$).

Fourth, the total number of *Events Recalled* was calculated for the free-recall and cognitive load questions to assess children's abilities to recall all their experiences with E1, beyond just the wallet situation. As seen in Table 3, children participated in seven activities with E1 prior to their first interaction with the interviewer. For the free-recall and cognitive load questions, children received a point if they referred to a new activity (unique Events Recalled) or repeated a previously recalled activity (repeated Events Recalled), and a point if they recalled the event in the correct temporal order (Temporal Order Accuracy). For example, if a child recalled the Word Generation, Theory of Mind, Theft and then the Inhibition (incorrect order) tasks in this order, they would receive a score of 4 unique Events Recalled, and a Temporal Order Accuracy score of 3 out of 4 (75% accuracy) for this question. If the child recalled any of these activities on the following open-ended questions, then they would receive a repeated Event Recalled score each time they mentioned the activity. The correct temporal order was reversed for the reverse-order question (see Table 3). Lastly, children received Total Events Recalled scores, which was calculated by summing the unique Events Recalled and repeated Events Recalled scores. Total interview Temporal Order Accuracy scores was calculated by dividing the total temporal accuracy scores across the five questions by the total Events Recalled scores. For instance, if a child recalled 12 events (unique and repeated) across the five questions and 7 of them were in the correct order, then they would receive a total accuracy score of 7 out of 12 (58.3% accuracy). There was good reliability between coders on the Total Events Recalled (ICC = .90, $p < .001$) and Temporal Order Accuracy (ICC = .78, $p < .001$) variables.

Fifth, *Disclosure Frequency* was recorded for each of the free-recall and cognitive load questions. A disclosure entailed a child indicating that E1 took the money from the wallet. Instances whereby children mentioned the wallet and/or money, but did not indicate that E1 took the money, were not scored as a transgression disclosure. Each transgression disclosure was recorded, even if it was repeated; therefore, this variable was analyzed as a continuous measure. There was perfect reliability (ICC = 1.00, $p < .001$) between the two coders' Disclosure Frequency ratings.

Sixth, *Disclosure Forthcomingness*, as used in Malloy and Mugno (2016), referred to the categorical section of the interview wherein a child first made a voluntary disclosure or false denial about the theft (0 = did not make a disclosure; 1 = made a truthful or false disclosure). As seen in Table 1, children's forthcomingness to make a disclosure was scored across four levels in the CI condition, in accordance to whether they made their first disclosure on the free-recall (level 1), cognitive load (level 2), non-direct closed-ended (level 3) or direct inquiry closed-ended (level 4) questions. Participants in the SI group were scored on three levels (levels 1, 3 and 4) given that they were not asked the cognitive load questions (level 2). There was perfect inter-rater reliability among the two raters who scored this measure, Kappa = 1.00, $p < .001$.

Seventh, *Testimony Consistency*, as used in Saykaly and colleagues (2016), referred to the consistency of the participants' responses across eight closed-ended questions that inquired directly or non-directly about the theft. Participants received a score of 1 to 8 pertaining to how well their responses aligned with their final answer about the theft on the direct inquiry question; higher scores indicated a better level of response consistency (see Table 4).

Results

Statistical Procedures

The presentation order of the results for both studies are as follows. The first set of analyses assessed for potential developmental differences, namely child age (continuous), gender (categorical) and cognitive test performance, on the following eight dependent measures: (1) Report Honesty, (2) Response Length, (3) Transgression Details (unique, repeated and total), (4) Events Recalled (unique, repeated and total), (5) Temporal Order Accuracy, (6) Disclosure Forthcomingness, (7) Disclosure Frequency, and (8) Testimony Consistency. The second set of analyses examined interview protocol (CI versus SI) differences on these eight dependent measures. The third set of analyses evaluated testimony narrative differences between the free-recall and cognitive load question-types. The fourth set of analyses measured differences between veracity groups (truth versus lie-tellers) across the dependent measures.

Developmental Predictors of Testimony Performance

Preliminary analyses assessed whether child age and the cognitive measures (Inhibition, Word Generation and ToM) should be treated as important predictors when examining interview group and question-type differences on the dependent measures. Bivariate correlation analyses revealed a small-to-moderate correlation between child age and total interview (i.e., free-recall and cognitive load question scores combined) Temporal Order Accuracy (r = .30, p = .002); however, there were no significant correlations between age, Transgression Details and total Events Recalled. There were also no significant correlations between the Inhibition, ToM and the dependent measures; whereas, only small correlations were found between Word Generation scores and total Response Length (r = .27, p = .006), Transgression Details (r = .19, p = .048) and Events Recalled (r = .29, p = .003). Given these weak correlations, child age and the

cognitive measures were not used as covariates when examining interview and question-type differences on the dependent measures.

Preliminary MANOVAs revealed significant gender differences on the total interview Response Length [$F(1, 100) = 6.52$, $p = .012$], Transgression Details [$F(1, 100) = 7.92$, $p = .006$], Events Recalled [$F(1, 100) = 5.81$, $p = .018$], and Disclosure Frequency [$F(1, 100) = 7.32$, $p = .008$] measures. These gender differences were still significant for Response Length, Transgression Details and Disclosure Frequency when analyzing just the participants in the CI group ($n = 52$); marginally significant gender differences were found on the Events Recalled measure [$F(1, 50) = 3.473$, $p = .068$] for CI participants. Conversely, there were no significant gender differences when analyzing just the participants in the SI group ($n = 52$). There were also no significant gender differences in children's attempts and final decisions to deceive the interviewer, and with respect to Temporal Order Accuracy and Testimony Consistency. Based on these findings, gender was inputted as an additional predictor for the following analyses.

Cognitive Interview versus Standard Interview

Binary logistic regressions evaluated interview group and gender differences (predictor variable) in *Report Honesty* on children's *Attempts* to deceive the interviewer on the free-recall and/or cognitive load questions (dependent variable; analysis 1) and on their *Final Decision* to deceive in response to the direct inquiry question (i.e., "Did E1 take money from the wallet?") at the end of the interview (analysis 2). Overall, 72.1% ($n = 75$) of children attempted to deceive the interviewer on the free-recall or cognitive load questions by directly denying E1's involvement in the theft or by omitting the wallet situation altogether. In contrast, only 41.3% ($n = 43$) followed through with the false denial in response to the direct inquiry question. Neither binary logistic regression analysis revealed significant interview group differences in children's

attempts [$\chi^2(1, N = 104) = .43$, $p = .512$, Nagelkerke R Square = .006] or final decisions [$\chi^2(1, N = 104) = .36$, $p = .550$, Nagelkerke R Square = .005] to deceive the interviewer. Additionally, there were no gender differences on either Report Honesty measures.

An ordinal logistic regression assessed for interview group and gender differences in Disclosure Forthcomingness (categorical variable; 4-levels). For this analysis, the SI group and boys were the reference groups for the two predictors. The overall model was not significant, $\chi 2(2, N = 104) = 1.66$, $p = .436$, Nagelkerke $R^2 = .02$. The majority of first disclosures were made on the free-recall (30.8%) and direct inquiry (46.2%) questions, which were asked in both interviews. There were also no gender differences on this measure. Figure 1 shows the Disclosure Forthcomingness frequency on each question-type on the CI and SI.

Independent samples *t*-tests examined for interview and gender differences in Disclosure Frequency. Overall, children in the CI (*M disclosures* = 2.72) group disclosed the theft significantly more throughout the interview compared to the SI (*M disclosures* = 0.63), $t(59) = 4.62$, $p < .001$. However, since the cognitive load questions were not asked in the SI, interview group differences regarding Disclosure Frequency on just the free-recall question was examined separately. Results from this independent samples *t*-test revealed no significant differences between the CI and SI groups in Disclosure Frequency on the free-recall question, $t(59) = -.02$, $p = .982$. Lastly, girls disclosed the theft significantly more throughout the interview than the boys [*Mdifference* = +2.05 disclosures; $t(27) = -2.48$, $p = .020$].

A MANOVA assessed for interview and gender group (predictor variables) differences in the total interview Response Length (dependent measure 1), Transgression Details (dependent measure 2), Events Recalled (dependent measure 3) and Testimony Consistency (dependent measure 4). The CI group was the reference category for the interview group variable, and boys

were the reference category for the gender variable. The overall models for interview group and child gender were significant for this MANOVA. As seen in Table 5, the CI group produced significantly more words [Response Length; $F(1, 100) = 40.32$, $p < .001$], Transgression Details [$F(1, 100) = 9.18$, $p = .003$] and Events Recalled [$F(1, 100) = 118.93$, $p < .001$] throughout their interview compared to the SI. In contrast, there were no significant differences between the CI and SI groups on the Testimony Consistency measure. In accordance with the preliminary analyses, girls produced testimonies that included more words [Response Length; $F(1, 100) = 4.77$, $p = .031$], Transgression Details [$F(1, 100) = 6.54$, $p = .012$] and more Events Recalled [$F(1, 100) = 4.52$, $p = .036$] than boys, irrespective of interview group. Refer to Table 6 for the gender differences across these dependent measures.

Follow-up MANOVAs evaluated interview and gender group differences in unique and repeated Transgression Details throughout the interview. As expected, the CI group recalled significantly more unique [F (1, 100) = 5.72, $p = .019$] and repeated [F (1, 100) = 16.99, $p < .001$] Transgression Details than the SI group (see Table 5). As shown in Table 6, girls produced significantly more unique [F (1, 100) = 6.44, $p = .013$] and repeated [F (1, 100) = 4.48, $p = .037$] Transgression Details than the boys. A second follow-up MANOVA assessed for interview and gender group differences in unique and repeated Events Recalled, and differences in Temporal Order Accuracy. While the CI group produced significantly more unique [F (1, 100) = 26.50, $p < .001$] and repeated events than the SI [$F(1, 100) = 148.72$, $p < .001$], there were no interview differences in Temporal Order Accuracy (see Table 5). Girls recalled significantly more unique [$F(1, 100) = 5.39$, $p = .022$], but not repeated, events. However, there were no gender differences in Temporal Order Accuracy (see Table 6).

Given that both interviews include the free-recall question, a MANOVA examined dependent measure differences between interview and gender groups on this specific question. Overall, there were no significant differences between interview groups across the dependent measures when analyzing just the free-recall question responses. There were significant gender differences in children's total Transgression Details [$F(1, 100) = 5.91$, $p = .017$] and Disclosure Frequency [$F(1, 100) = 4.00$, $p = .049$] on this question. Notably, girls produced free-recall responses with higher Transgression Details (*M difference* = +6.21 details, $p = .017$) and Disclosure Frequency (*M difference* = +.28 disclosures, $p = .049$) scores than the boys.

Question-type Differences

Repeated measures MANOVAs with Greenhouse-Geisser corrections evaluated differences in Response Length (analysis 1), Transgression Details (analysis 2), Events Recalled (analysis 3) and Disclosure Frequency (analysis 4) between the free-recall question and each of the four cognitive load questions. Since only the CI included the cognitive load questions, children in the SI group ($n = 52$) were excluded from these analyses. Additionally, it was not possible to assess for question-type differences in Temporal Order Accuracy given that children did not report an event on every question. Child gender was again inputted as a second independent variable for the following analyses.

Response Length. The first repeated measures MANOVA revealed significant differences between the five questions in overall Response Length, $F(1.65, 82.25) = 13.79$, $p < .001$. According to post-hoc tests using the Bonferroni correction, children produced significantly more words on the free-recall question when compared to the mental context reinstatement (*M difference* = 70.47, $p < .001$), another perspective (*M difference* = 79.10, $p < .001$), reverse-order (*M difference* = 40.56, $p = .025$) and re-tell questions (*M difference* = 37.07,

$p = .001$). The reverse-order question also had longer responses than the mental context reinstatement (*M difference* = 29.92, $p < .001$) and another perspective (*M difference* = 38.55, $p < .001$) questions. Similarly, the re-tell question had significantly longer responses than the mental context reinstatement (*M difference* = 33.40, $p = .003$) and another perspective (*M difference* = 42.03, $p = .001$) questions. Finally, post-hoc tests using the Bonferroni correction revealed no significant gender differences in Response Length on the individual questions.

Transgression Details. Significant differences between questions in the total number of Transgression Details reported were found on the second MANOVA, $F(2.88, 142.95) = 7.03$, $p < .001$. As shown in Table 7, post-hoc tests using the Bonferroni correction indicated that children gave significantly more Transgression Details on the free-recall and re-tell questions when compared to the mental context reinstatement and another perspective questions. Children also gave significantly more Transgression Details on the re-tell question when compared to the reverse-order recall question; whereas, children produced significantly more Transgression Details on the reverse-order recall question than on the another perspective question (see Table 7). Significant gender effects were found, $F(1, 50) = 5.14$, $p = .028$. Girls gave significantly more Transgression Details than the boys on the free-recall question ($B = 8.76$, $p = .017$), and marginally more details on the mental context reinstatement question ($B = 4.08$, $p = .054$).

Given the significant differences between questions in the number of Transgression Details reported, follow-up repeated measures MANOVAs indicated significant differences between questions in unique [F (2.65, 132.41) = 6.32, $p = .001$] and repeated Transgression Details [F (1.62, 80.97) = 11.25, $p < .001$]. Post-hoc tests using the Bonferroni correction determined that children gave significantly more unique Transgression Details on the free-recall question when compared to the mental context reinstatement and another perspective questions

(see Table 7). Responses on the re-tell question also included significantly more unique details than the another perspective question. Children gave significantly more repeated details on the re-tell question compared to the other four questions; whereas, reverse-order responses included significantly more repeated Transgression Details than the free-recall responses (refer to Table 7). Lastly, there were significant gender effects for both unique [$F(1, 50) = 25.61$, $p = .033$] and repeated [$F(1, 50) = 4.46$, $p = .040$] Transgression Details. More specifically, girls produced significantly more unique details than the boys on the free-recall ($B = 8.66$, $p = .074$) and mental context reinstatement ($B = 3.18$, $p = .047$) questions. Moreover, girls gave significantly more repeated details than the boys on the reverse-order recall question ($B = 2.12$, $p = .039$).

Events Recalled. For the third MANOVA, there were significant differences between question-types in the total number of events (unique + repeated) recalled throughout the interview, $F(3.16, 216.72) = 63.00$, $p < .001$. Children recalled significantly more total (unique + repeated) events on the free-recall, reverse-order and re-tell questions when compared to the mental context reinstatement and another perspective questions; reverse-order responses also included more events than the free-recall and re-tell responses (see Table 8). While there were marginally significant gender differences in the total number of Events Recalled [$F(1, 50) = 3.47$, $p = .068$], no significant gender differences were found when examining the Events Recalled on the individual free-recall and cognitive load questions.

Follow-up repeated measures MANOVAs revealed significant differences between question-types in the total number of unique [$F(1.87, 131.23) = 78.37$, $p < .001$] and repeated [$F(2.05, 102.28) = 73.04$, $p < .001$] Events Recalled throughout the interview. As seen in Table 8, children recalled significantly more unique events on the free-recall question when compared to each of the four cognitive load questions. Moreover, children recalled significantly more

unique events on the reverse-order question when compared to the other three cognitive load questions. With regards to repeated Events Recalled, children repeated more events on the reverse-order and re-tell questions when compared to the free-recall, mental context reinstatement and another perspective questions (see Table 8). Similarly, significantly more repeated events were recalled on the mental context reinstatement and another perspective questions when compared to free-recall responses. There were no significant gender differences in unique and repeated Events Recalled on any of the questions.

Disclosure Frequency. For the final MANOVA, there were significant differences between question-types in Disclosure Frequency, $F(2.82, 27.10) = 3.45$, $p = .023$. Children disclosed the theft significantly more on the free-recall question when compared to the mental context reinstatement question (*M difference* = .27, $p = .041$). Significant gender differences were also found, $F(1, 27) = 5.54$, $p = .026$. Girls disclosed the theft significantly more than the boys on the free-recall ($b = .71$, $p = .014$) and reverse-order ($b = .53$, $p = .052$) questions.

True versus False Narrative Characteristics

Truth-tellers were compared with the lie-tellers on the Response Length, Transgression Details, Events Recalled, Temporal Order Accuracy, Disclosure Forthcomingness, Disclosure Frequency and Testimony Consistency dependent measures. Truth-tellers included children who stated that E1 took the money on the final direct inquiry question; whereas, lie-tellers said that E1 did not take the money on this question. There were no significant differences between the truth-tellers and lie-tellers with respect to gender distribution and cognitive test scores; however, there was a marginally significant differences between truth-tellers (*M age* = 10.79 years) and lie-tellers (*M age* = 10.13 years) in age, $F(1, 103) = 4.00$, $p = .048$.

A MANOVA examined differences between truth-tellers and lie-tellers on the Response Length, Transgression Details, Events Recalled, Temporal Order Accuracy and the Testimony Consistency dependent measures for the entire sample (i.e., both the CI and SI groups). Overall, there were significant differences between veracity groups on the Transgression Details [$F(1, 98) = 8.64$, $p = .004$] and Testimony Consistency [$F(1, 98) = 18.49$, $p < .001$] dependent measures. Truth-tellers produced significantly more details about the theft (*M difference* = 14.95, $p = .004$) than the liars. Given these results, a follow-up MANOVA indicated significant differences between veracity groups in unique [$F(1, 98) = 8.35$, $p = .005$] and repeated [$F(1, 98) = 7.10$, $p = .009$] Transgression Details. Truth-tellers gave significantly more unique (*M difference* = 10.91, $p = .005$) and repeated (*M difference* = 4.05, $p = .009$) information about the theft than the lie-tellers. Refer to Figure 2 for the mean number of unique, repeated and total details for the truth-tellers and lie-tellers. Unexpectedly, however, lie-tellers ($M = 7.81$) were significantly more consistent than the truth-tellers ($M = 6.91$) in their responses on the eight closed-ended questions (*M difference* = .90, $p < .001$). There were no significant differences between veracity groups on the Response Length, Events Recalled and Temporal Order Accuracy dependent measures.

Next, an ordinal regression examined for differences between veracity groups in Disclosure Forthcomingness. The overall model was significant, $\chi2(2, N = 104) = 29.78$, $p = <$ .001, Nagelkerke $R^2 = .27$. Four follow-up logistic regression analyses were then conducted, whereby disclosures made on each question-type (free-recall, cognitive load, non-direct closed-ended and direct inquiry) were the dependent measures; the lie-tellers were the reference category for each analysis. The binary logistic regression model was significant for the free-recall question [$\chi2(1, N = 104) = 22.21$, $p = <$ .001, Nagelkerke $R^2 = .27$], with 69.2% of the cases being correctly classified. Overall, the lie-tellers were significantly less likely ($B = -2.49$, p

< .001) than the truth-tellers to make a disclosure (i.e., falsely deny the theft) on the free-recall question. The model was also significant for the direct inquiry question [$\chi 2(1, N = 104) = 24.45$, $p = < .001$, Nagelkerke $R^2 = .28$], with 74.0% cases being correctly classified. In contrast to the free-recall question, lie-tellers were significantly more likely to make their first disclosure on the direct inquiry question ($B = 2.10$, $p < .001$) when compared to the honest reporters. There were no significant veracity group differences in Disclosure Forthcomingness on the cognitive load and indirect closed-ended questions. Figure 3 shows the Disclosure Forthcomingness frequency (%) on each question type for the truth-tellers and lie-tellers.

Follow-up MANOVA analyses were conducted within the CI (analysis 1) and SI (analysis 2) groups. Because few children first disclosed the theft on the cognitive load questions (see Figure 3), veracity group differences in Disclosure Forthcomingness was not examined within the CI and SI groups. For these analyses, veracity group was the predictor and Response Length, Transgression Details, Events Recalled, Temporal Order Accuracy and Testimony Consistency scores were the dependent variables. On the first analysis (CI group only), significant veracity group differences were found on the Transgression Details [$F(1, 48) = 6.01$, $p = .018$], Events Recalled [$F(1, 48) = 5.55$, $p = .023$] and Testimony Consistency [$F(1, 48) = 16.67$, $p < .001$] measures. However, these differences were not present in the SI group (analysis 2). In the CI group, truth-tellers gave significantly more Transgression Details (*M difference* = + 21.73, $p = .018$) and Events Recalled (*M difference* = + 2.62, $p = .023$) than the lie-tellers. Follow-up MANOVAs indicated that truth-tellers interviewed using the CI produced significantly more unique [$F(1, 50) = 6.48$, $p = .014$; *M difference* = + 17.73 details] and repeated details [$F(1, 50) = 8.71$, $p = .005$; *M difference* = + 8.27 details] than the liars, as well as more unique [$F(1, 50) = 7.55$, $p = .008$; *M difference* = + 1.12 events] and repeated Events

Recalled [$F(1, 50) = 3.98$, $p = .052$; *M difference* = + 1.61 events]. Finally, lie-tellers in both the CI [*M difference* = + 0.98, $p = < .001$] and SI [*M difference* = + 0.79, $p = .035$] groups responded more consistently than the truth-tellers in response to the eight indirect closed-ended questions.

Two follow-up MANOVAs assessed which questions were most effective for distinguishing between the truthful and false responses. For the first analysis, Transgression Details and Events Recalled on the free-recall question only was the dependent measures, and veracity group was the predictor. The second analysis measured veracity group differences in Transgression Details reported on each of the cognitive load questions. Analysis three examined veracity group differences in Events Recalled on each of the cognitive load questions. Analysis 1 was separate from analyses 2 and 3 because it included the entire sample, since the free-recall question was asked in both the CI and SI groups. Analyses 2 and 3 only included the CI participants since they were the only ones administered the cognitive load questions.

For analysis 1, there were significant differences between veracity groups in Transgression Details disclosed only [$F(1, 102) = 8.03$, $p = .006$]. Overall, truth-tellers disclosed significantly more details (*M difference* = + 6.57, $p = .006$) about the theft in their free-recall responses than the liars. On analyses 2 and 3, there were significant veracity group differences in Transgression Details disclosed [$F(1, 50) = 5.14$, $p = .028$] and Events Recalled [$F(1, 50) = 7.09$, $p = .010$] on the reverse-order question. Truth-tellers discussed the theft more than the liars (*M difference* = + 4.24, $p = .028$), and stated more events (*M difference* = + 1.02, $p = .010$), when recalling their experiences with E1 in reverse-order. Lastly, truth-tellers provided significantly more Transgression Details than the liars (*M difference* = + 3.34 details) on the mental context reinstatement question, $F(1, 50) = 11.57$, $p = .001$. No significant veracity group differences were found on the another perspective and re-tell questions.

Discussion

The present study was designed to provide researchers and forensic professionals with new information on the efficacy of the four cognitive load questions for encouraging credible eyewitness reports from children about an adult's crime. The first objective of this study was to evaluate whether developmental factors, namely child age, gender and cognitive test scores, predicted eyewitness performance across the eight dependent measures. In contrast to our hypotheses, child age and cognitive test performance were not meaningfully correlated with Response Length, Transgression Details, Events Recalled or Disclosure Frequency. Prior eyewitness research indicated that children's abilities to produce detail eyewitness reports increased with age and verbal test scores (e.g., Jack et al., 2014; Wyman et al., 2019). However, this research typically included younger children than those who participated in this study. In Wyman and colleagues (2019) for example, the age group differences in testimony content were most prevalent when comparing the 6-7 years olds to those above the age of 8.

Despite the fact that there were no gender differences in child age, verbal test scores or in report honesty, girls in the CI group had significantly higher Response Length, Transgression Details (unique and repeated), Events Recalled (unique only) and Disclosure Frequency scores than the boys (see Table 6). On the free-recall question, girls had significantly higher Transgression Details (unique only) and Disclosure Frequency scores than the boys. Further, girls gave significantly more Transgression Details on the mental context reinstatement (unique) and reverse-order questions (repeated); girls also reported the theft significantly more on the reverse-order question. These gender differences in eyewitness reporting were expected given that girls outperform boys on verbal assessments to a small degree (Hyde, 2016; Petersen, 2018), and in language scholastic courses (Voyer & Voyer, 2014). Girls also demonstrate more acute

interpersonal and social awareness than boys (Hall & Matsumoto, 2005; Hampson et al., 2006), which is important when attempting to communicate clearly with others. Finally, these findings reinforce some prior literature that suggests that girls benefit more from the CI. In a deception study with adults, Suckle-Nelson and colleagues (2010) found that women produced more coherent eyewitness statements with more words and details on the CI than men. In Wyman and colleagues (2018), mock-jurors perceive child female eyewitness responses on a modified version of the CI to be more convincing than male responses. Taken together, when given additional opportunities to provide open-ended recollections on the CI, school-age girls produce responses with more words, details, events and repeated disclosures than boys.

To better understand the efficacy of cognitive load questioning with children, the second objective of this study was to examine whether the CI increased eyewitness performance on the eight dependent measures of testimony performance. The current findings provide support for using some cognitive load questions with children. As seen in Table 5, children questioned using the CI had significantly higher scores on the Response Length, Transgression Details (unique and repeated), Disclosure Frequency and Events Recalled (unique and repeated) measures when compared to the SI. However, there were no interview group differences in Disclosure Forthcomingness. While children disclosed the theft more often on the CI, this protocol did not encourage children to first disclose the theft earlier in the interview.

When analyzing just the free-recall responses, there were no significant differences between interview groups on any of the dependent measures. However, the addition of the cognitive load questions resulted in children giving more relevant (higher Transgression Details and Disclosure Frequency scores) and irrelevant (higher Response Length and Events Recalled scores) information pertaining to the theft. This was expected given that CI group was given

four additional opportunities to openly recall the events with E1. While the CI enhanced children's memory recall of the events with E1, the increased cognitive effort associated with these questions surprisingly did not result in lower Temporal Order Accuracy or Testimony Consistency. Thus, the CI increased eyewitness performance, without compromising accuracy.

To better understand the benefits and limitations of cognitive load questioning with children, the third objective of this study was to compare the efficacy of the individual free-recall and cognitive load questions. The current findings reinforce prior literature which suggests that chronological free-recall questioning, such as the free-recall and re-tell questions used in this study, are very effective for encouraging children to provide detailed eyewitness reports (e.g., Guadagno et al., 2006; Wyman et al., 2019). Children in the CI group produced the most words, unique Transgression Details, unique Events Recalled and more frequent disclosures on the initial free-recall question (see Tables 6 and 7). Moreover, the re-tell question at the end of the interview encouraged children to provide significantly more words, total Transgression Details and repeated events. Consistent with Wyman and colleagues (2019), the reverse-order recall question was the most effective cognitive load question for increasing the number of words, Transgression Details (repeated) and Events Recalled (repeated).

In both veracity groups, children provided very little new information about the theft (Transgression Details and Disclosure Frequency) in response to the mental context reinstatement and another perspective questions. Thus, both the lie-tellers and truth-tellers simply omitted information about the theft on these questions likely because of the challenges associated with understanding the complex requirements of these two questions. Saywitz and colleagues (1992) argued that the mental context reinstatement and another perspective questions may be too mentally taxing for children, as these questions did not result in much new

information recalled. Future research should examine the efficacy of different modification strategies for increasing children's abilities to respond to these questions. For example, interviewers can break-up the mental context reinstatement question into several smaller parts, such as by asking children to recall each sense separately (e.g., "What did you smell? What did you see?"). Further, a study by Mattison and colleagues (2014) provided some support for having children, with and without an autism diagnosis, produce sketches in response to this question, instead of only verbal reports.

The current study was one of the first to assess whether these commonly used interviewing strategies can discourage false reporting by children. In contrast to the adult literature (Vrij et al., 2011, 2012), there were no significant differences between interview groups in children's attempts (CI = 75%; SI = 69%) and final decisions (CI = 44%; SI = 39%) to falsely deny the theft. Wyman and colleagues (2019) found a similar prevalence of attempts to falsely deny a theft (78%) or falsely accuse a researcher of a theft (72%) when children (ages 6 to 11) were interviewed using the CI. Moreover, their study found a higher rate of final decisions to falsely deny (63%) or falsely report (72%) the theft; however, there was no control (SI) group in this study (Wyman et al., 2019). These two studies indicate that the cognitive load questions did not increase the expected difficulty associated with telling a false story, likely because the children had difficultly responding to some of the cognitive load questions.

To accomplish the fourth objective, the testimony content of the truth-tellers and lie-tellers were compared. Consistent with the adult SVA literature (Colwell et al., 2007, 2013), as well as the relatively fewer child studies (Akehurst et al., 2001; Wyman et al., 2019), truth-tellers gave more forthcoming and detailed disclosures about the theft. The Disclosure Forthcomingness findings indicated that honest speakers were more willing to make a clear

disclosure about the theft earlier in the interview; whereas, the lie-tellers mostly waited until the final direct inquiry question at the end of the interview to make a clear false denial. Significant veracity group differences were also found on the unique, repeated and total Transgression Detail and Events Recalled dependent measures. Truth-tellers were more willing to discuss the wallet situation on the free-recall question, and they recalled new and repeated information about the theft on the reverse-order and mental context reinstatement questions. The deceivers instead lied by omission, and rarely discussed the wallet situation at any point during the interview. Furthermore, in contrast to the prior child SVA literature (e.g., Akehurst et al., 2001; Wyman et al., 2019), the current study measured veracity group differences in more than one interviewing protocol. Overall, the significant veracity group differences in the Disclosure Forthcomingness, Transgression Detail and Events Recalled were only found within the CI.

Surprisingly, the lie-tellers gave more consistent statements across the eight closed-ended questions. It was initially hypothesized that the truth-tellers would produce more consistent statements given that maintaining a false script is more cognitively taxing. This hypothesis was based on adult research that examined the coherence of their open-ended responses (e.g., Colwell, Hiscock, & Memon, 2002; Suckle-Nelson et al., 2010), rather than the consistency of their responses across a series of closed-ended questions. The Testimony Consistency coding used in this study was based on prior lie-maintenance research with children; albeit, this study required children to maintain their responses across more questions (i.e., 8-questions) than other studies, which typically had respondents answer 2-4 questions (e.g., Talwar & Lee, 2002ab; Talwar & Lee, 2008). In the current study, the lie-tellers simply had to answer “no” in response to the eight closed-ended questions that inquired about the theft. By answering “yes” to these

questions, however, the truth-tellers may have experienced more stress and guilt given that their responses implicated E1 in being responsible for a crime.

Directions for Future Research

The current study was designed to provide much needed information on the efficacy of the commonly used free-recall and cognitive load questions with children. This study expands on the prior interviewing literature (e.g., Gentle et al., 2013; Pipe & Wilson, 1994; Talwar et al., 2004; Wyman et al., 2019) by utilizing a research paradigm whereby children witnessed and participated in a prolonged theft. This target event generalizes to real-life settings whereby children are questioned about a familiar adult's crime, and to situations wherein they may feel complicit in the illegal activity (Romeo et al., 2018). In correspondence with real-life situations in which children feel internal and external pressure to falsely deny or omit the crimes of others (Black, Schweitzer, & Varghese, 2012; Schreiber et al., 2006), children in the present study were also coached by the transgressor to falsely deny the theft. Furthermore, the current study provided one of the most comprehensive evaluations of the efficacy of the CI versus the SI to date, along with the individual cognitive load and free-recall questions, using eight evidence-based indicators of testimony performance. Therefore, the current findings provide legal professionals and researchers with a much more complete picture of the strengths and weaknesses of the commonly used investigative interviewing strategies.

The current research methodology can be expanded upon in several ways. In the current study, typically-developing children reported on a theft that was carried out by a researcher they had gotten to know for 45-minutes. While this target event has greater real-life relevance compared to other forensic studies, there are several ways to further increase the generalizability of this paradigm. For example, Talwar, Lee, Bala and Lindsay (2004) had parents, the most

common perpetrators of crimes against children, coach their children to fabricate a false report. They also included a court-simulated procedure that utilized mock lawyers (e.g., prosecutors and defense) and judges to resemble real cases wherein children are interviewed after witnessing or experiencing a crime. Moreover, having children report on the transgression following a considerable time delay, such as several days or a week after the crime, is more consistent with actual forensic cases. This is particularly important because a time delay between witnessing an event and testifying can reduce the accuracy (Quas et al., 1999), consistency (Poole & White, 1993) and verbosity (Peterson & Whalen, 2001) of a disclosure.

Conclusions

The present study comprehensively evaluated the efficacy of the CI with children for producing longer, more detailed, accurate and honest eyewitness reports about an adult's crime. The current findings provide evidence for using free-recall and reverse-order questions for eliciting detailed disclosures from children about a crime they witnessed. Children gave the longest (higher Response Length), most detailed (higher number of Transgression Details and Events Recalled) and most frequent (higher Disclosure Frequency) disclosures about a theft in response to the free-recall question. At the same time, children recalled more repeated information about the theft and events with E1 in response to the reverse-order and re-tell questions. However, the mental context reinstatement and another perspective questions had minimal effects for encouraging children to provide new information about the theft. While the total CI interview responses were longer and more detailed when compared to the SI, this interview did not result in more honest or forthcoming reports about the theft. Surprisingly, the expected difficulty associated with the CI did not reduce children's overall Temporal Order Accuracy or Testimony Consistency.

The current study also provides some evidence of gender differences in children's abilities to provide longer and detailed disclosures about another's transgression. Girls provided significantly more words, transgression details, transgression disclosures and events recalled throughout the total sample. Within the CI group, girls had higher scores on the Response Length, Transgression Details, Events Recalled and Disclosure Frequency measures; these gender differences were particularly evident on the free-recall (unique details and Transgression Disclosures), mental context reinstatement (unique details) and reverse-order (repeated details and Transgression Disclosures) questions.

Furthermore, this study provides evidence that using the CI can lead to narrative differences among children's truthful and false reports. Notably, the truthful reporters produced more forthcoming disclosures on the CI, with a higher number of transgression details (unique, repeated and total) and events recalled (unique, repeated and total). Conversely, liars mainly lied by omission and rarely gave any new information about the wallet situation. These results suggest that the Transgression Details SVA coding measure, in collaboration with the new Disclosure Forthcomingness measure (Malloy & Mugno, 2016), can possibly be used by forensic professionals as methods for detecting truthful and false information in children's testimonies.

Taken together, professionals involved in the investigations of crimes against children should utilize repeated free-recall questioning for eliciting detailed disclosures from eyewitness about difficult topics. The current study also provides support for using the reverse-order cognitive load question for encouraging children to recall new and repeated information. Without adaptations, such as breaking-up the complex questions into smaller parts, the mental context reinstatement and another perspective cognitive load questions may be too cognitively taxing for children under the age of 13.

References

Akehurst, L., Köhnken, G., & Höfer, E. (2001). Content credibility of accounts derived from live and video presentations. *Legal and Criminological Psychology, 6*, 65-83. doi:10.1348/135532501168208

Akehurst, L., Milne, R., & Kohnken, G. (2003). The effects of children's age and delay on recall in a cognitive or structured interview. *Psychology, Crime & Law*, *9*, 97–107. doi:10.1080/10683160308140

Berntsen, D., & Rubin, D. C. (2006). Emotion and vantage point in autobiographical. *Cognition and Emotion, 20*, 1193-1215. doi.org/10.1080/02699930500371190

Black, F., Schweitzer, R., & Varghese, F. (2012). Allegations of child sexual abuse in family court cases: A qualitative analysis of psychiatric evidence. *Psychiatry, Psychology and Law, 19*, 482-496. doi:10.1080/13218719.2011.613905

Bull, R., Paulo, R.M., & Albuquerque, P.B. (2019). The impact of the cognitive interview in the UK and recent research in Portugal. In J.J. Dickinson, N.S. Compo, R. Carol, B.L. Schwartz, & M.R. McCauley (Eds). *Evidence-based investigative interviewing: Applying cognitive principles* (pp. 29-41). New York, NY: Routledge.

Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., & McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: Evidence from activation likelihood estimate meta-analyses. *Cerebral Cortex*, *19*, 1557-1566. doi:10.1093/cercor/bhn189

Colwell, K., Hiscock, C. K, & Memon, A. (2002). Interviewing techniques and the assessment of statement credibility**.** *Applied Cognitive Psychology, 16,* 287-300. doi:10.1002/acp.788Colwell, K., Hiscock-Anisman, C. K., & Fede, J. (2013).

Assessment Criteria Indicative of Deception (ACID): An example of the new paradigm of Differential Recall Enhancement. In B. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in eyewitness memory, credibility assessment, and investigative interviewing* (pp. 259–291). New York, NY: Springer.

Colwell, K., Hiscock-Anisman, C., Memon, A., Rachel, A., & Colwell, L. (2007). Vividness and spontaneity of statement detail characteristics as predictors of witness credibility. *American Journal of Forensic Psychology, 25*, 5-30. doi:10.07331290

Davis, J.L., & Matthews, R. (2010). NEPSY-II review. *Journal of Psychoeducational Assessment*, *28*(2), 175-182. doi:10.117/0734282909346716

Eisen, M. L., Goodman, G. S., Qin, J., Davis, S., & Crayton, J. (2007). Maltreated children's memory: Accuracy, suggestibility, and psychopathology. *Developmental Psychology*, *43*, 1275-1294. doi:10.1037/0012-1649.43.6.1275

Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C. Thomas Publisher Limited.

Geiselman, R. E., & Callot, R. (1990). Reverse versus forward recall of script-based texts. *Applied Cognitive Psychology*, *4*, 141-144. doi:10.1002/acp.2350040206

Gentle, M., Milne, R., Powell, M. B., & Sharman, S. J. (2013). Does the Cognitive Interview promote the coherence of narrative accounts in children with and without an intellectual disability? *International Journal of Disability, Development, and Education, 60*, 30-43. doi:10.1080/1034912X.2013.757138

Guadagno, B. L., Powell, M. B., & Wright, R. (2006). Police officers' and legal professionals' perceptions regarding how children are, and should be, questioned about repeated abuse. *Psychiatry, Psychology and Law, 13*, 251-260. doi:10.1375/pplt.13.2.251

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion, 4*, 201-206. doi:10.1037/1528-3542.4.2.201

Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A girl advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behaviour, 27,* 401-416. doi:10.1016/j.evolhumbehav.2006.05.002

Hershkowitz, I., Lamb, M. E., Orbach, Y., Katz, C., & Horowitz, D. (2012). The development of communicative and narrative skills among preschoolers: Lessons from forensic interviews about child abuse. *Child Development, 83,* 611-622. doi:10.1111/j.1467-8624.2011.01704.x

Hershkowitz, I., Orbach, Y., Lamb, M. E., Sternberg, K. J., & Horowitz, D. (2001). The effects of mental context reinstatement on children's accounts of sexual abuse. *Applied Cognitive Psychology, 15*, 235-248. doi:10.1002/acp.699

Hershkowitz, I., & Terner, A. (2007). The effects of repeated interviewing on children's forensic statements of sexual abuse. *Applied Cognitive Psychology, 21*, 1131-1143. doi:10.1002/acp.1319

Holliday, R. E., Brainerd, C. J., Reyna, V. F., & Humphries, J. E. (2009). The cognitive interview: Research and practice across the lifespan. In R. Bull, T. Valentine, & T. Williamson (Eds.), *Handbook of the psychology of investigative interviewing: Current developments and future directions* (pp. 137–160). Oxford: Wiley.

Hyde, J. S. (2016). Sex and cognition: Gender and cognitive functions. *Current Opinion in Neurobiology, 38,* 53-56. doi:10.1016/j.conb.2016.02.007

Jack, F., Leov, J., & Zajac, R. (2014). Age-related differences in the free-recall accounts of child, adolescent, and adult witnesses. *Applied Cognitive Psychology*, *28,* 30-38. doi:10.1002/acp.2951

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: Administration manual*. San Antonio, TX: The Psychological Corporation.

London, K., Bruck, M., Ceci, S. J., & Shuman, D. W. (2005). Disclosure of child sexual abuse: What does the research tell us about the ways that children tell? *Psychology, Public Policy, and Law*, *11,* 194-226. doi:10.1037/1076-8971.11.1.194

Lyon, T. D. (2014). Interviewing children. *Annual Review of Law and Social Science*, *10,* 73-89. doi:10.1146/annurev-lawsocsci-110413-030913

Lyon, T. D., & Dorado, J. S. (2008). Truth induction in young maltreated children: The effects of oath-taking and reassurance on true and false disclosures. *Child Abuse & Neglect*, *32,* 738-748. doi:10.1016/j.chiabu.2007.08.008

Malloy, L. C., & Mugno, A. P. (2016). Children's recantation of adult wrongdoing: An experimental investigation. *Journal of Experimental Child Psychology*, *145,* 11–21. doi:10.1016/j.jecp.2015.12.003

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, *89*, 137–149. doi:10.1037/0021-9010.89.1.137

Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B., & Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search.

Proceedings of the National Academy of Sciences, 108, 12893-12897. doi:10.1073/pnas.1015174108

Mattison, M. L., Dando, C. J., & Ormerod, T. C. (2014). Sketching to remember: Episodic free recall task support for child witnesses and victims with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 45,* 1751-1765. doi:10.1007/s10803-0142335-z

McCauley, M. R., & Fisher, R. P. (1995). Facilitating children's eyewitness recall with the revised cognitive interview. *Journal of Applied Psychology, 80*, 510-516. doi:10.1037/0021-9010.80.4.510

Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law, 16*, 340-372. doi:10.1037/a0020518

Milne, R., & Bull, R. (2003). Does the cognitive interview help children to resist the effects of suggestive questioning? *Legal and Criminological Psychology, 8*, 21-38. doi:10.1348/135532503762871219

Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse & Neglect, 24,* 733-752. doi:10.1016/S0145-2134 (00)00137-X

Petersen, J. (2018). Gender difference in verbal performance: A meta-analysis of United States state performance assessments. *Educational Psychology Review, 30,* 1269-1281. doi:10.1007/s10648-018-9450-x

Peterson, C., & Whalen, N. (2001). Five years later: Children's memory for medical emergencies. *Applied Cognitive Psychology, 15*, S7-S24. doi:10.1002/acp.832

Pipe, M.-E., & Wilson, J. C. (1994). Cues and secrets: Influences on children's event reports. *Developmental Psychology, 30*, 515-525. doi:10.1037/0012-1649.30.4.515

Poole, D. A., & White, L. T. (1993). Two years later: Effect of question repetition and retention interval on the eyewitness testimony of children and adults. *Developmental Psychology, 29*, 844-853. doi:10.1037/0012-1649.29.5.844

Quas, J. A., Goodman, G. S., Bidrose, S., Pipe, M. E., Craw, S., & Ablin, D. S. (1999). Emotion and memory: Children's long-term remembering, forgetting, and suggestibility. *Journal of Experimental Child Psychology, 72*, 235-270. doi:10.1006/jecp.1999.2491

Quas, J. A., Goodman, G. S., Ghetti, S., & Redlich, A. D. (2000). Questioning the child witness: What can we conclude from the research thus far? *Trauma, Violence, & Abuse, 1*, 223-249. doi:10.1177/1524838000001003002

Rice, H. J., & Rubin, D. C. (2009). I can see it both ways: First-and third-person visual perspectives at retrieval. *Consciousness and cognition, 18*, 877-890. doi:10.1016/j.concog.2009.07.004

Romeo, T., Otgaar, H., & Landstrom, S. (2018). Coping with Sexual Abuse. *Psychological Research on Urban Society, 1*, 46-54. doi:10.7454/proust.v1i1.11

Saykaly, C., Crossman, A., Morris, M., & Talwar, V. (2016). Question type and its effect on children's maintenance and accuracy during courtroom testimony. *Journal of Forensic Practice, 18*, 104-117. doi:10.1108/JFP-01-2015-0010

Saywitz, K. J., Geiselman, R. E., & Bornstein, G. K. (1992). Effects of cognitive interviewing and practice on children's recall performance. *Journal of Applied Psychology, 77,* 744-756. doi:10.1037/0021-9010.77.5.744

Schreiber, N., Bellah, L. D., Martinez, Y., McLaurin, K. A., Strok, R., Garven, S., & Wood, J. M. (2006). Suggestive interviewing in the McMartin Preschool and Kelly Michaels daycare abuse cases: A case study. *Social Influence, 1,* 16-47. doi:10.1080/15534510500361739

Smith, D. W., Letourneau, E. J., Saunders, B. E., Kilpatrick, D. G., Resnick, H. S., & Best, C. L. (2000). Delay in disclosure of childhood rape: Results from a national survey. *Child Abuse & Neglect, 24,* 273-287. doi:10.1016/S0145-2134(99)00130-1

Suckle-Nelson, J. A., Colwell, K., Hiscock-Anisman, C., Florence, S., Youschak, K. E., & Duarte, A. (2010). Assessment criteria indicative of deception (ACID): Replication and gender differences. *The Open Criminology Journal, 3,* 23-30. doi:10.2174/1874917801003010023

Talwar, V., & Crossman, A. (2011). From little white lies to filthy liars: The evolution of honesty and deception in young children. *Advances in Child Development and Behaviour, 40,* 139-179. doi:10.1016/B978-0-12-386491-8.00004-9

Talwar, V., & Crossman, A. M. (2012). Children's lies and their detection: Implications for child witness testimony. *Developmental Review, 32,* 337-359. doi:10.1016/j.dr.2012.06.004

Talwar, V., & Lee, K. (2002a). Development of lying to conceal a transgression: Children's control of expressive behaviour during verbal deception. *International Journal of Behavioural Development, 26,* 436–444. doi:10.1080/01650250143000373

Talwar, V., & Lee, K. (2002b). Emergence of white lie-telling in children between 3 and 7 years of age. *Merrill-Palmer Quarterly, 48,* 160–181. doi:10.1353/mpq.2002.0009

Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behaviour. *Child Development, 79,* 866-881. doi:10.1111/j.1467-8624.2008.01164.x

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). Children's lie-telling to conceal a parent's transgression: Legal implications. *Law & Human Behaviour, 21,* 405-426. doi:10.1023/B:LAHU.0000039333.51399.f6

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

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140,* 1174-1204. doi:10.1037/a0036620

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

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science, 20,* 28-32. doi:10.1177/0963721410391245

Vrij, A., Leal, S., Mann, S., & Fisher, R. (2012). Imposing cognitive load to elicit cues to deceit: Inducing the reverse-order technique naturally. *Psychology, Crime & Law, 18* 579-594. doi:10.1080/1068316X.2010.515987

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse-order. *Law and Human Behaviour, 32,* 253-265. doi:10.1007/s10979-007-9103-y

Wyman, J., Foster, I., Crossman, A., Colwell, K., & Talwar V. (2018). The efficacy of free-recall, cognitive load, and closed-ended questions when children are asked to falsely testify about a crime. *Applied Cognitive Psychology*, Advance online publication,. doi:10.1002/acp.3494

Wyman, J., Foster, I., Lavoie, J., Tong, D., & Talwar, V. (2019). Detecting children's false allegations and recantations of a crime. *Psychology, Crime & Law*, *24*, 652-671. doi:10.1080/1068316X.2017.1402018

Table 1

Organization and Structure of the Cognitive Interview and Standard Interview.

Recall task	Phrase from recall task	Forthcomingness Category
1. Free recall	Please describe, in as much detail as possible, everything that happened upstairs with E1. - Did anything else happen?	1
2. CL #1: Mental context reinstatement (CI only)	Now, think about and tell me all the sights, sounds, smells, emotions, thoughts, or anything else you remember from your time with E1. - Is there anything else?	2
3. Non-direct closed-ended block 1	a. Did anything happen that you did not think would happen? b. If a police officer had been upstairs with you and E1, would they have noticed something wrong? c. Were there winter boots in the room upstairs? d. Was a crime committed?	3
4. CL #2: Another's perspective (CI only)	I want you to imagine that your mom (or dad) was in the room with you and E1. Can you describe in as much detail as possible everything your mom (or dad) would remember regarding the events with E1? - Would s/he remember anything else?	2
5. Non-direct closed-ended block 2	a. I saw that there was a wallet in the room. Was it closer to you or E1? b. Did you see any money in the wallet? c. Did E1 do something they were not supposed to do while in the room upstairs?	3

		d. Were you or E1 sitting closer to the door of the room? e. Did E3, the other researcher you met upstairs, do something they were not supposed to do while in the room upstairs?	
6.	CL #3: Reverse-order (CI only)	Please describe everything you remember from upstairs, but backwards. Beginning with the last event upstairs and ending with the first event upstairs. So, please describe everything you remember from upstairs with E1 but backwards. - Did anything else happen?	2
7.	Non-direct closed-ended block 3	a. Was there money in the wallet? b. Was the money in the wallet green or blue? c. If a police office had been upstairs with you and E1, would they have arrested someone? d. Do you think you could have been mistaken about anything you said so far?	3
8.	CL #4: Re-tell (CI only)	One last time, please describe in as much detail as possible, everything that happened with E1. - Did anything else happen?	2
9.	Non-direct closed-ended block 4	a. So just one more time, did you see a wallet? b. Did you see money in the wallet? c. Did someone take the money from the wallet? d. Did you take the money from the wallet?	3
10.	Direct inquiry	a. Did E1 take the money from the wallet? b. Did E3 take the money from the wallet?	4

Table 2

Preliminary MANOVA Results Regarding Interview Group Differences (Mean, SD) in Age and Cognitive Test Scores.

	Cognitive Interview (*n* = 52)	Standard Interview (*n* = 52)	Total (*n* = 104)
Age	10.26	10.77	10.52
Gender			
Boys (%)	44.2% (*n* = 23)	59.6% (*n* = 31)	51.9% (*n* = 54)
Girls (%)	55.8% (*n* = 29)	40.4% (*n* = 21)	48.1% (*n* = 50)
Inhibition Total Score	27.88	28.21	28.04
Naming Task	8.73	8.75	8.74
Inhibition Task	9.98	9.31	9.64
Switching Task	9.17	10.15	9.66
Total Errors	13.71	12.65	13.18
Word Generation (WG) Total	21.59	20.36	20.98
Initial Letter WG	9.17	8.61	8.89
Semantic WG	12.42	11.75	12.09
Theory of Mind (ToM) Total[1]	21.33	23.24	22.28
ToM Verbal Task[2]	16.50	18.31	17.40
ToM Contextual Task	4.83	4.93	4.88

Note. Overall differences between the CI and SI groups were not significant at $p < .05$ on any of the measures, except for ToM Total[1] [$F(1) = 5.67$, $p = .019$] and ToM Verbal Task[2] [$F(1) = 8.95$, $p = .003$].

Table 3

Events Recalled Scoring Procedures

Free-recall, mental context reinstatement, another perspective and re-tell questions scoring	Unique Event (0 or 1)	Repeated Event (0 or 1)	Order Accuracy (0 or 1)
1. Three Questions Game			
2. Inhibition			
3. Word Generation			
4. Theory of Mind			
5. E2 Interaction			
6. Guessing Game			
7. Theft Situation			
TOTAL			

Reverse-order event recall scoring procedures	Unique Event (0 or 1)	Repeated Event (0 or 1)	Order Accuracy (0 or 1)
1. Theft Situation			
2. Guessing Game			
3. E2 Interaction			
4. Theory of Mind			
5. Word Generation			
6. Inhibition			
7. Three Questions Game			
TOTAL			

TOTALS (combined scores across 5-questions)

Unique Event Score: _____

Event Order Score: _____

Repeated Event Score: _____

Table 4

Study 1 Testimony Consistency Scoring Procedures.

Questions	**False Denial** (Maintains false story that theft did not happen)	**True Accusation** (Maintains true story that E1 did take the money)
1. If a police officer had been upstairs, would they have noticed something wrong?	NO	YES
2. Was a crime committed?	NO	YES
3. Did E1 do something they were not supposed to do while in the room?	NO	YES
4. Did E3, the other researcher you met upstairs, do something they were not supposed to do while in the room?	NO	NO
5. If a police officer had been upstairs with you and E1, would they have arrested someone?	NO	YES
6. Did someone take the money from the wallet?	NO	YES
7. Did E1 take the money from the wallet?	NO	YES
8. Did E3, the other researcher you met upstairs, take the money from the wallet?	NO	NO

* Scored out of 8

Table 5

Mean (SD) Differences between the CI and SI in Total Interview Response Length, Transgression Details, Events Recalled, Temporal Order Accuracy, Disclosure Frequency and Testimony Consistency Dependent Measures.

	Cognitive Interview (n = 52)	Standard Interview (n = 52)	Total Sample (N = 104)
Response Length**	346.92 (264.53)	96.56 (78.28)	221.74 (231.31)
Transgression Details**	25.35 (35.56)	8.44 (10.83)	16.89 (27.50)
Unique Details*	18.78 (26.24)	8.27 (10.67)	13.53 (20.62)
Repeated Details**	6.57 (10.77)	.16 (.57)	3.37 (8.24)
Total Events Recalled**	9.15 (4.23)	2.33 (1.38)	5.74 (4.64)
Unique Events**	3.95 (1.55)	2.33 (1.38)	3.14 (1.67)
Repeated Events**	5.20 (2.97)	0	2.60 (3.34)
Temporal Order Accuracy	78.86%	73.93 %	76.44 %
Disclosure Frequency*	1.51 (2.27)	0.39 (0.66)	0.95 (1.76)
Testimony Consistency	90.13%	92.38%	91.25%

* Difference between interview groups significant at $p < .05$.
** Difference between interview groups significant at $p < .01$.

Table 6

Mean Gender Differences (SD) in Total Interview Response Length, Transgression Details, Events Recalled, Temporal Order Accuracy, Disclosure Frequency and Testimony Consistency.

	Girls (n = 50)	Boys (n = 54)
Response Length*	283.96 (290.27)	164.13 (138.01)
Transgression Details*	24.83 (35.12)	9.55 (14.67)
Unique Details*	19.36 (25.74)	8.13 (12.34)
Repeated Details*	5.47 (10.84)	1.42 (3.91)
Events Recalled*	6.95 (4.81)	4.62 (4.23)
Unique Events*	3.64 (1.68)	2.68 (1.54)
Repeated Events	3.31 (3.55)	1.95(3.04)
Temporal Order Accuracy (%)	76.93%	75.98%
Disclosure Frequency**	1.44 (2.18)	0.50 (1.10)
Testimony Consistency	89.75%	92.63%

* Difference between genders significant at $p < .05$.
** Difference between genders significant at $p < .01$.
Note[1]. A preliminary MANOVA indicated that boys and girls did not differ significantly in age or cognitive test performance on Word Generation subtest, nor in their attempts and final decisions to deceive the interviewer.
Note[2]. Gender differences in Total Transgression Details were still significant when analyzing just the participants in the CI group (n = 54); differences in Response Length (p = .055) and Events Recalled (p = .068) were marginally significant in this group. No gender differences were found within the SI group on any of the dependent measures.

Table 7

Pairwise Comparisons between Questions in the Mean Number of Unique, Repeated and Total Transgression Details.

Unique Details				
	1	2	3	4
1. Free recall	--	--	--	--
2. Mental context reinstatement	-4.76*	--	--	--
3. Another perspective	- 5.62*	-.86	--	--
4. Reverse-order	- 3.89	.87	1.73	--
5. Re-tell	- 1.99	2.78	3.64*	1.90
Repeated Details				
	1	2	3	4
1. Free recall	--	--	--	--
2. Mental context reinstatement	.57	--	--	--
3. Another perspective	.64	.07	--	--
4. Reverse-order	1.51**	.94	.87	--
5. Re-tell	2.91**	2.34*	2.27**	1.40*
Total Details				
	1	2	3	4
1. Free recall	--	--	--	--
2. Mental context reinstatement	-4.19**	--	--	--
3. Another perspective	-4.98**	-.79	--	--
4. Reverse-order	-2.38	1.81	2.60**	--
5. Re-tell	.925	5.12**	5.90**	3.31**

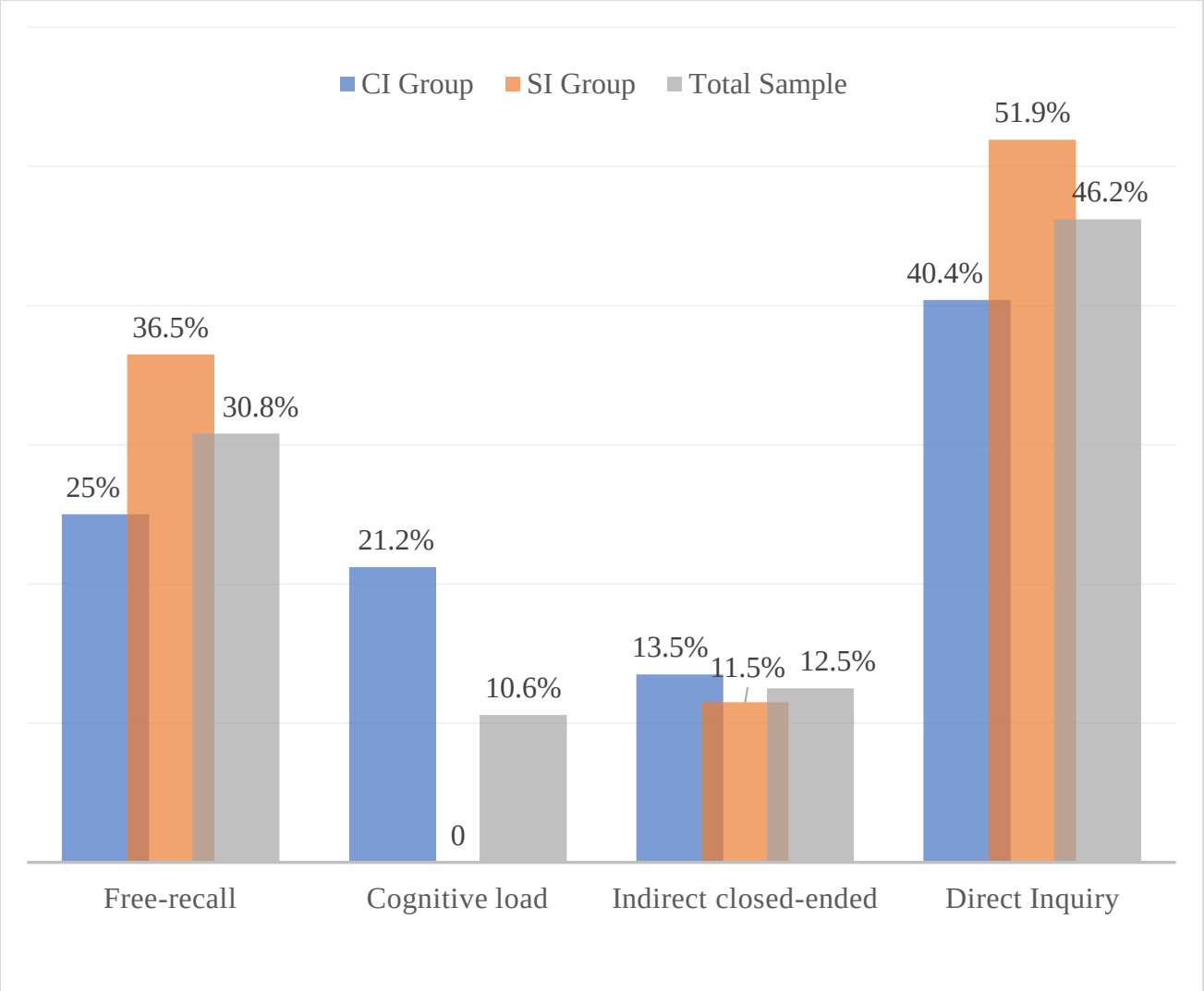
* Mean differences significant at $p < .05$.
** Mean differences significant at $p < .01$.

Table 8

Pairwise Comparisons between Questions in the Mean Number of Unique, Repeated and Total Events Recalled.

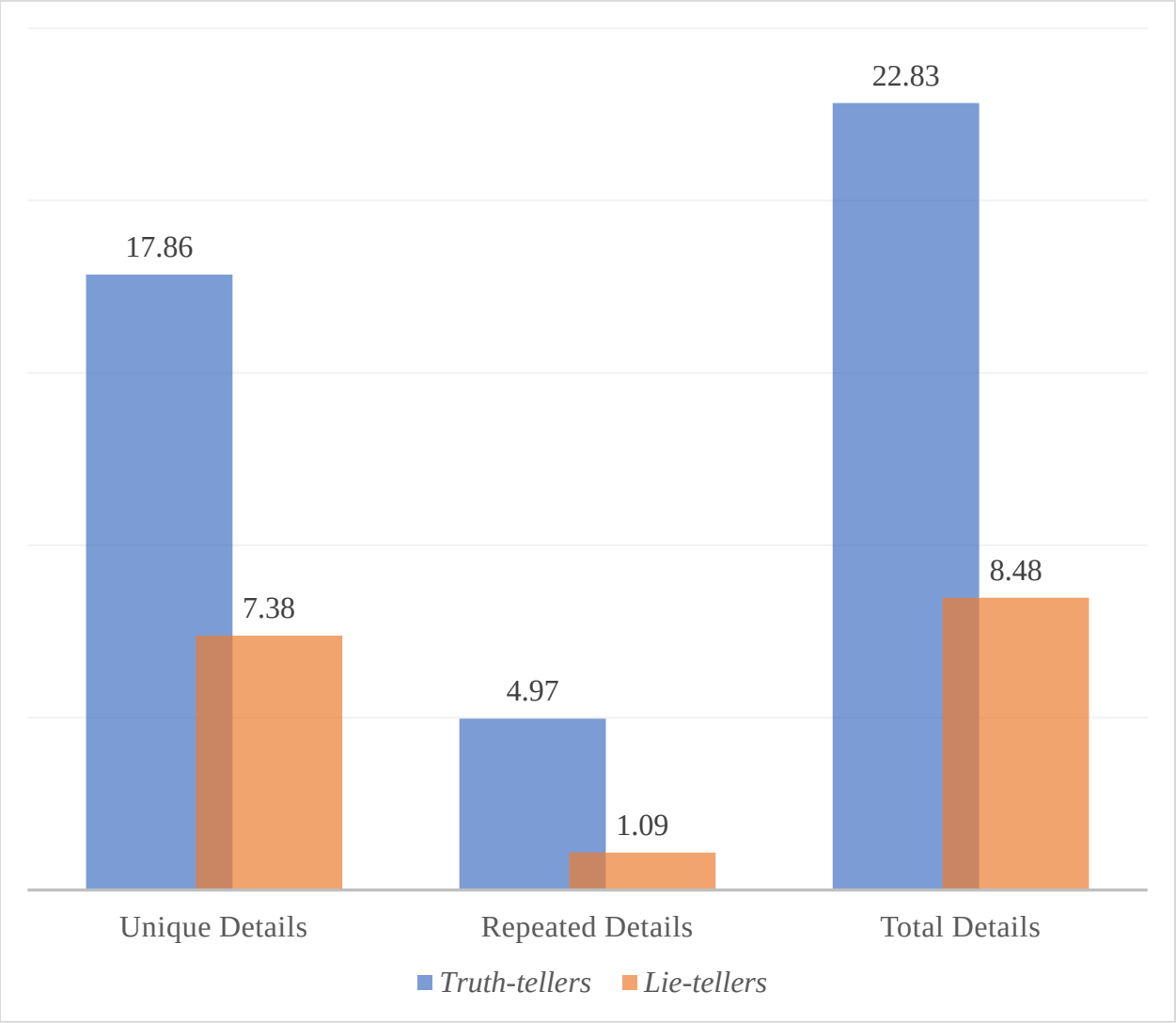
Unique Events Recalled				
	1	2	3	4
1. Free recall	--	--	--	--
2. Mental context reinstatement	-2.31**	--	--	--
3. Another perspective	-2.37**	-.06	--	--
4. Reverse-order	-1.64**	.66**	.73**	--
5. Re-tell	-2.25**	.05	.12	-.61**
Repeated Events Recalled				
	1	2	3	4
1. Free recall	--	--	--	--
2. Mental context reinstatement	.26**	--	--	--
3. Another perspective	.54**	.28	--	--
4. Reverse-order	2.08**	1.82**	1.54**	--
5. Re-tell	2.22**	1.97**	1.69**	.15
Total Events Recalled				
	1	2	3	4
1. Free recall	--	--	--	--
2. Mental context reinstatement	-2.05**	--	--	--
3. Another perspective	-1.83**	.22	--	--
4. Reverse-order	.43*	2.48**	2.26**	--
5. Re-tell	-.03	2.02**	1.80**	-.46*

* Mean differences significant at $p < .05$.
** Mean differences significant at $p < .01$.



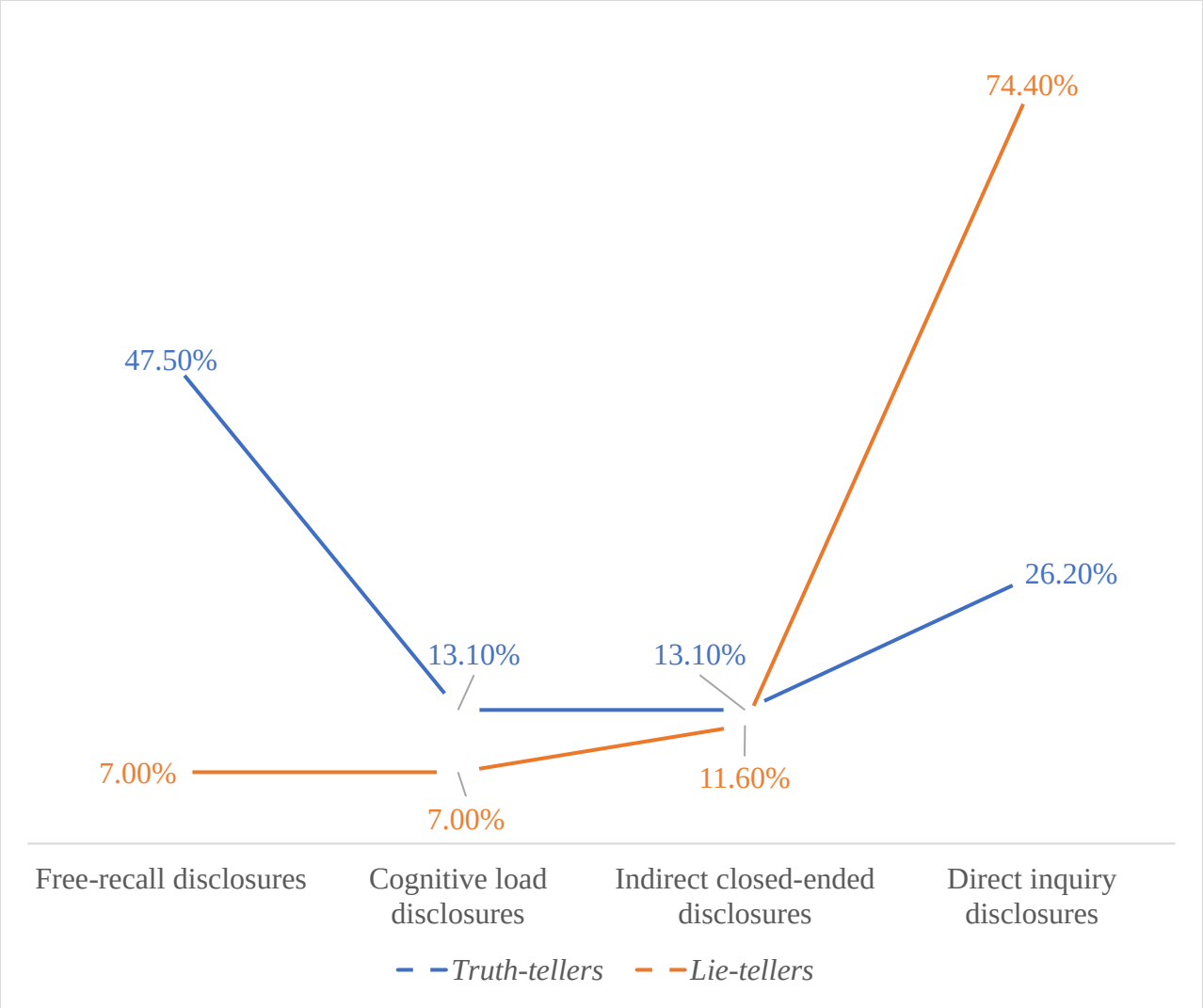
Note. Children in the SI group were not asked the cognitive load questions.

Figure 1. Frequency (%) of first disclosures on each question-type in the CI and SI groups, as well as according to the total sample.



Note. MANOVA indicated that the truth-tellers disclosed significantly ($p < .01$) more unique, repeated and total details than the lie-tellers.

Figure 2. Mean number of unique, repeated and total transgression details among typically-developing truth-tellers and lie-tellers.



Note. Binary logistic regression analyses indicated that the truth-tellers were significantly more likely to make their first disclosure (i.e., accuse E1 of the theft) on the free-recall question when compared to the lie-tellers ($B = 2.49$, $p < .001$). Whereas, the lie-tellers were significantly more likely to make their first disclosure (i.e., falsely deny theft) on the direct inquiry closed-ended question ($B = 2.10$, $p < .001$).

Figure 3. Frequency (%) of first truthful and false disclosures on each question-type among typically-developing truth-tellers and lie-tellers.

Chapter 4: Bridging Manuscripts 1 and 2

The following manuscript expands on Study 1 by assessing whether the CI can be effective with CWID. The results from Study 1 indicated that the CI was effective for increasing the amount of words, transgression details and events recalled among TDC, without decreasing the accuracy or consistency of their statements. In particular, the two free-recall questions (initial and re-tell) and the reverse-order recall questions increased children's memory recall about the events during the study. The CI was also useful for distinguishing between children's true and false reports. Notably, truth-tellers produced more forthcoming and detailed disclosures about the researcher's transgression, while the lie-tellers typically omitted any information about the transgression until they were directly asked about it at the end of the interview.

Although these results provide important information about whether the CI can be used with TDC, it is presently unclear whether these findings generalize to populations with cognitive deficits. It is important to study the value of different questioning strategies with CWID given their increased vulnerability to being maltreated (Jones et al., 2012). At the same time, child welfare and forensic professionals report having inadequate training and procedures for questioning CWID (Taylor et al., 2016); thus, they often have low confidence in eyewitness credibility of this vulnerable population (Henry et al., 2011). CWID also have significant cognitive deficits that can greatly impact their eyewitness testimony performance when compared to TDC, including lower verbal, memory and cognitive processing capabilities. While the research objectives of the two studies are the same, the second manuscript provides child welfare and forensic interviewers with important guidance on how to question CWID about an adult's transgression.

Chapter 5: Manuscript #2

Using the Cognitive Interview to question children with intellectual disabilities about a transgression

Joshua Wyman[1], Christine Saykaly[2], & Victoria Talwar[1]

[1]Department of Educational and Counselling Psychology, McGill University.

[2]Summit School, Montreal, QC.

Abstract

The current study evaluated the efficacy of the Cognitive Interview with children (*N* = 50; ages 6 to 18) with mild and moderate intellectual disabilities. After watching a magic show wherein the magician made a major error, the children were asked by the magician to keep the transgression a secret. Next, children were interviewed using the Cognitive Interview (CI: free-recall, cognitive load and closed-ended questions) or a Standard Interview (SI: free-recall and closed-ended questions) by an unfamiliar researcher, whereby their statement quality (statement consistency, veracity, temporal order accuracy, and disclosure forthcomingness and frequency) and quantity (number of words, transgression details and events recalled) were evaluated. The CI group disclosed more total words, events, and transgression details and disclosures than the SI group, without compromising temporal order accuracy or testimony consistency. Nevertheless, there were no significant interview group differences in disclosure honesty or forthcomingness. Truth-tellers provided the most forthcoming and detailed statements on the free-recall question, irrespective of interview type. Conversely, lie-tellers rarely discussed the alleged transgression, and primarily lied on the final direct closed-ended question. These results suggest that the CI can be effective with children with intellectual disabilities for increasing eyewitness memory recall, without compromising accuracy.

Keywords: Forensic interviewing; children with intellectual disabilities; deception; eyewitness credibility.

Using the Cognitive Interview to question children with intellectual disabilities about a transgression

Child welfare and legal professionals have expressed the need to improve forensic interviewing training with atypical child populations (Aarons et al., 2004; Taylor, et al., 2016). To date, many of the commonly used interviewing practices, such as the Cognitive Interview (CI; Fisher & Geiselman, 1992), have been tested primarily with typically-developing children (TDC; Memon et al., 2010). However, the cognitive and adaptive-skill profile of children with intellectual disabilities (CWID) differs considerably when compared to their typically-developing counterparts (American Psychiatry Association, 2013), which can impact their ability to provide detailed eyewitness reports (see Wyman, Lavoie et al., 2018 for a review). Moreover, forensic professionals often feel unprepared to interview CWID, as they report having inadequate training and procedures needed to effectively interview this population (Taylor et al., 2016). In addition to the increased prevalence of maltreatment (Jones et al., 2012), CWID are more likely to be perceived as having lower eyewitness credibility (Henry et al., 2011). To address these important concerns, the current study analyzed the efficacy of free recall, cognitive load and closed-ended questions for encouraging CWID to produce honest, forthcoming, consistent and detailed disclosures about an adult's transgression.

Cognitive Load Questioning with CWID

CWID have significant cognitive deficits compared to TDC (American Psychiatric Association, 2013). According to the *Diagnostic and Statistical Manual of Mental Disorders: Fifth Edition* (DSM-5; American Psychiatric Association, 2013), an "intellectual disability" (i.e., Intellectual Developmental Disorder) is a disorder that begins early in life that interferes with a child's cognitive development and adaptive functioning. In most cases, an intellectual disability

is diagnosed when a child's cognitive functioning, as determined by standardized intelligence measures, is at least two standard deviations below the population mean. Deficits in cognitive functioning can affect a child's learning potential and ability to use age-appropriate communication, reasoning and judgment. Severe delays in learned adaptive behaviours, such as functional communication, applied academic and self-care skills, must also be present. Delays in adaptive functioning can limit a child's ability to meet community standards of personal independence and social responsibility (American Psychiatric Association, 2013). A child's cognitive delay can impact how they respond to standard forensic interviewing measures, given that recalling detailed and accurate information increases with age and cognitive development (Jack et al., 2014; Wyman, Lavoie, et al., 2018; Wyman et al., 2019).

While there is considerable CI research with adult eyewitnesses (see Bull et al., 2019; Memon et al., 2010, for reviews) and some research with TDC (e.g., Saykaly et al., 2016; Wyman et al., 2019), only two CI studies to date (Gentle et al., 2013; Milne, Sharman, Powell, & Mead., 2013) included a sample of CWID; whereas, two other studies tested children with high functioning autism (Mattison et al., 2014) and learning disabilities (Robinson & McGuire, 2006). In Gentle and colleagues (2013), CWID who were administered the mental context reinstatement and reverse-order recall questions produced disclosures about a videotaped magic show that were longer, more consistent and more accurate when compared to those who were given a standard interview. Similarly, Milne and colleagues (2013) found that CWID and TDC interviewed using the CI recalled more correct details about a videotaped magic show.

Although these studies provide insight into the efficacy of the CI with CWID, there are also some gaps in this research that the current study will address. Firstly, the children in these studies were truthfully recalling the events from memory that they did not actually witness in

person. Yet, children often falsely deny or omit experiences of maltreatment for prolonged periods of time (Smith et al., 2000), which can make it difficult for authorities to substantiate reports of abuse made by people (e.g., neighbors and school personnel) other than the victim. Much of the forensic literature to date with CWID has discussed the fact that this population is more susceptible to disclosing false information when being interviewed using suggestive questioning strategies (e.g., Henry & Gudjonsson, 1999, 2003). Nevertheless, when telling a lie, the child is knowingly attempting to deceive the interviewer, rather than simply communicating false information that they believe to be true. While much is known regarding TDC's deception skills (see Talwar and Crossman, 2011, for a review), the current study will be the first to analyze CWID's willingness and abilities to tell true and false eyewitness disclosures when being interviewed with generalizable forensic interviewing protocols.

Secondly, by having children witness a live transgression, instead of viewing a video recording of a positive event (Gentle et al., 2013; Milne et al., 2013), the present study can better generalize to real-life situations wherein children most often have to disclose about a transgression they witnessed in person. Witnessing a live versus a videotaped event enhances the accuracy of a child's testimony (Thierry & Spence, 2004), possibly because a video cannot capture many sensory stimuli experienced during a live event, such smells, tastes and tactile information, that can assist with memory recollection (Hershkowitz et al., 2001).

Thirdly, the CI has shown to be useful for discouraging false reporting from adult eyewitnesses (Vrij et al., 2011, 2012) given that imposing increased cognitive load makes producing and maintaining a false narrative more challenging (Vrij, 2000). Adults' truthful statements in response to cognitive load questions are therefore more vivid (more words and details) compared to their false statements (Vrij, Fisher, & Blank, 2017). However, no study to

date has assessed whether these truthful versus false report differences also exist in populations with intellectual deficits. The current study will therefore be the first to assess if the cognitive load questions can be used to distinguish between the true and false narratives of CWID.

Present Study

In the current study, CWID witnessed a transgression during a live magic show whereby the magician accidentally spilled ink on a pair of expensive gloves. Children were then asked by the magician to conceal the error. Next, children were interviewed using the SI or CI to determine the effectiveness of different question-types for encouraging truthful, detailed, accurate, consistent and forthcoming disclosures about the magician's error.

The first objective was to assess whether child developmental factors, such as age, gender and intelligence test scores, were related to their performance on eight dependent measures of testimony performance: (1) Report Honesty, (2) Response Length, (3) Transgression Details (unique, repeated and total), (4) Events Recalled (unique, repeated and total), (5) Temporal Order Accuracy, (6) Disclosure Frequency, (7) Disclosure Forthcomingness, and (8) Testimony Consistency. Scores on dependent measures two through eight are expected to increase with child age and cognitive test performance; however, these developmental factors will not predict report veracity. Consistent with developmental research with TDC (Petersen, 2018; Voyer & Voyer, 2014), girls are expected to produce longer and more detailed testimonies than the boys.

The second objective assessed whether the CI, when compared to the SI, increased children's performance on the eight dependent measures. In line with forensic studies to date (e.g., Gentle et al., 2013; Milne et al., 2013), CWID administered the CI will provide the most honest, vivid (higher Response Length, Transgression Details and Events Recalled), frequent and forthcoming disclosures compared to the SI group. Conversely, the CI group is expected to have

lower Temporal Order Accuracy and Testimony Consistency due to the increased mental effort required to respond to the cognitive load questions. Given that only 24 children were in the CI group, it was not possible to reliably examine question-type differences in testimony content.

For the third objective, the statements of truth-tellers and lie-tellers were examined across the dependent measures of eyewitness performance (2 to 5, 7 and 8). Based on the adult literature (see Colwell et al., 2013, for a review), truth-tellers will produce statements that are more forthcoming, consistent and detailed about the magician's error compared to the lie-tellers. Whereas, the lie-tellers will give shorter and less detailed open-ended statements given that they will lie by omitting the transgression altogether (Wyman et al., 2019).

Method

Participants

Fifty-six participants between the ages of 6 and 18 were recruited from a private school in public interest in a large metropolitan area (population approximately 3,000,000). The private school serves 600 students with a range of developmental disabilities, including intellectual disabilities, Autism Spectrum Disorder (ASD), Downs Syndrome and language impairments. The study methodology received ethics approval from the residing university and the private school. A research coordinator, employed by the private school, determined participant eligibility according to the study's recruitment requirements. All participants had to have a prior diagnosis of an intellectual disability and be fluent in English. However, children diagnosed with ASD (n = 18, 36%) or Downs Syndrome (n = 13, 26%), in addition to their ID diagnosis, were included in the study. Parental consent was obtained for the eligible participants.

All eligible participants were administered four subtests from the Wechsler Abbreviated Scale of Intelligence: 2nd Edition (WASI-II; Wechsler & Zhou, 2011) by a doctoral school

psychology student to assess their current level of cognitive functioning. In addition to a Full Scale Intelligence Quotient (FSIQ), the WASI-II provided Verbal Comprehension Index (VCI) and Perceptual Reasoning Index (PRI) t-scores. Higher *t*-scores reflected better performance.

The majority of participants (*n* = 46, 82%) continued to meet the diagnostic criteria for a mild (*n* = 29) or moderate (*n* = 17) intellectual disability. However, five children obtained a FSIQ score above 80 on the WASI-II, and thus, were excluded from the final analyses; moreover, one student was excluded because they were age 20. Four children performed in the borderline range of intellectual functioning (*t*-scores between 70 and 79), and given their prior ID diagnosis via full psychological evaluation, were included in the final analyses. The final sample included fifty participants (*Mage* = 12.48 years; *nmales* = 31, 62%), which is similar to other CI studies with atypical child populations, including Milne and colleagues (2013; *N* = 20 CWID, 26 TDC) and Robinson and McGuire (2006; *N* = 38 children with learning disabilities). As seen in Table 1, there were no significant age, FSIQ score, VCI score, or gender differences between the CI (*n* = 24) and SI (*n* = 26) groups. But, the CI group had higher PRI scores.

Materials and Procedures

Magic show. Participant testing took place in the private school's research center, which consisted of a classroom and two private interview rooms. The magic show was presented to groups of 4 to 8 children (participants and non-participants); whereas, the participating children completed the lie request, interview, debriefing and cognitive testing components separately. A teacher was present for the magic show, except during the transgression, to assist with any potential behaviour management issues.

The magic show procedures correspond to those in Pipe and Wilson (1994). After introducing herself as a "student magician", the magician completed four magic tricks, such as

making an object defy gravity, and having objects randomly appear, disappear and multiply. Additionally, the magician provided participants with a card deck, whereby they were taught a card trick for 10-minutes. Following the card trick, the classroom teacher informed the group that she was leaving the room. For the final trick, the magician presented the audience with a pair of "special" and "expensive" gloves. The magician then opened a small bottle of black ink and placed it next to the gloves. However, when reaching across the table to grab a ball, she knocked the ink onto the gloves. The magician appeared very distressed following the ink spill, and unsuccessfully attempted to clean the gloves. The magician then proceeded to show the children the damaged gloves and insisted that she could get in trouble because the ruined gloves were expensive. After cleaning up the materials and leaving the center, the classroom teacher returned and brought the non-participating students back to their class. Study participants remained in the research center; nevertheless, they were brought to separate interview rooms to prevent any interactions between participants about the magic show and ink spill. Overall, the magic show lasted 20 to 25 minutes.

Lie request. After the magician left the classroom, the interviewer entered the research center and greeted the participants. The interviewer then brought the children to a private interview room, but s/he had them sit outside the room for 3-minutes while s/he cleaned it; the interview room door was closed while the participant was seated outside. Given that the two interview rooms were in separate areas of the research center, the participants who watched the same magic show were not able to see or interact with each other while they were waiting.

Next, the magician walked by the participants and greeted them. The magician asked them if they remembered what happened to the special gloves. If a participant could not recall the event on their own, the magician briefly described it to them. The magician then reiterated

their regret about the ink spill and stated that she did not want to get in trouble. Finally, the participants were asked to keep the ink spill and ruined gloves a secret in case the interviewer asks because the magician did not want to get in trouble. The participants were then asked to repeat the secret request to confirm that they understood it; nevertheless, they did not have to agree to keep the secret. Overall, the majority of children (n = 43, 86%) in the CI (84.6%) and SI (87.5%) groups told E1 that they would keep the secret for them. After the participants demonstrated that they understood the request, the magician left the area.

Interview. One-minute later, the interviewer asked the participants to enter the interview room. The interviewer explained that s/he was interested in his/her memory of the magic show and the magician. Participants were interviewed using the Standard Interview (SI) or the Cognitive Interview (CI). Participants were randomly assigned to their respective interview groups by the primary investigator prior to the commencement of the study. Table 2 shows the interview questions used in the SI and CI protocols.

In both groups, participants first answered one free-recall question that incorporated them recalling everything they remembered from the magic show. In the CI group, participants were also asked mental context reinstatement, another perspective, reverse-order and re-tell cognitive load questions. One prompt (e.g., "did anything else happen?") was asked after the free-recall and four cognitive load questions. Non-direct closed-ended questions (n = 9) asked for short or forced-choice information regarding the gloves and whether a transgression had occurred, without questioning whether the magician had damaged the gloves. At the end of the interview, a direct inquiry question inquired about the magician's involvement with the damaged gloves. After being debriefed about the true nature of study, participants completed the cognitive testing with a different researcher. The interviews were video-recorded and later transcribed.

Coding Procedures

The coding procedures correspond to those in Study 1. First, *Report Honesty* referred to whether a participant attempted to deceive the interviewer (0 = no; 1 = yes) at any point during the interview and/or on the final direct inquiry question (0 = no; 1 = yes). Second, *Response Length* included the total number of words disclosed on each open-ended question (free-recall and cognitive load). Third, *Transgression Details* included each unique (new detail), repeated (previously mentioned) and total details (unique + repeated) disclosed about the glove trick. For example, the statement "the magician dropped ink on the gloves. After she dropped the ink, she said goodbye", includes fifteen words, six unique Transgression Details, and two repeated details (8 total details). There was substantial agreement between the two raters who coded the Transgression Details (Kappa = 0.72, $p < .001$).

Fourth, the total number of *Events Recalled* was calculated for the free-recall and cognitive load questions. Overall, children participated in seven different events throughout the magic show; refer to Table 3 for the scoring procedures for this measure. As in Study 1, each unique and repeated event recalled by the children on the free-recall and cognitive load questions was scored. Children were also scored on whether they recalled each event in the correct temporal order (*Temporal Order Accuracy*). For instance, if a child recalled the ball disappearance (correct order), ink spill (incorrect order) and jail escape (incorrect order) magic trick events on the free-recall question in this order, then they would have recalled three unique events with a 1/3 (33%) Temporal Order Accuracy score. Each subsequent mention of these events on the cognitive load questions would be scored as a repeated event recalled.

Fifth, *Disclosure Frequency* was scored every time a child specifically stated that the magician was responsible for the ink spill. Each transgression disclosure was recorded, even if it

was repeated. The total interview Disclosure Frequency score was calculated by summing the number of disclosures made on the free-recall and cognitive load questions.

Sixth, *Disclosure Forthcomingness* referred to the categorical section of the interview wherein a child first made a voluntary disclosure or denial about the magician's ink spill (0 = did not make a disclosure; 1 = made a truthful or false disclosure). Children's forthcomingness to disclose the ink spill was scored across four levels in the CI condition, in accordance to whether they made their first disclosure on the free-recall (level 1), cognitive load (level 2), non-direct closed-ended (level 3) or direct inquiry closed-ended (level 4) questions; participants in the SI group were scored on three levels (levels 1, 3 and 4). The inter-rater reliability among the two raters who scored this measure was found to be Kappa = 0.73 (p <. .001), indicating substantial inter-rater agreement. See Table 2 for the Disclosure Forthcomingness scoring procedures.

Seventh, *Testimony Consistency* referred to the consistency of the participants' responses across the five closed-ended questions that inquired directly or indirectly about the ink spill. Participants received a score of 1 to 5 pertaining to how well their responses aligned with their response on the last interview question. Higher scores indicated a higher level of response consistency. Refer to Table 4 for the Testimony Consistency scoring procedures for this study.

Results

A binary logistic regression was conducted to examine for potential developmental predictors (child age, gender and cognitive test performance) on children's attempts (analysis 1) and final decisions (analysis 2) to deceive the interviewer about the ink spill. Overall, 78% (n = 39) of participants attempted to deceive the interviewer at some point during the interview. Conversely, only 36% (n = 18) of participants maintained the secret by the end of the interview. The overall models for attempts and final decisions to deceive were both not significant. In both

analyses, child age, cognitive test performance and gender were not significant predictors of attempts or final decisions to deceive.

Preliminary correlation analyses analyzed the relationships between age, cognitive test performance (VCI, PRI and FSIQ scores), Response Length, Transgression Details, Events Recalled, Temporal Order Accuracy, Disclosure Frequency and Testimony Consistency. Beyond small correlations between total Response Length, VCI (r = .30, p = .04) and FSIQ scores (r = .29, p = .04), there were no correlations between the other developmental predictors and the dependent measures. Next, a MANOVA examined for gender differences on the Response Length, Transgression Details, Events Recalled, Temporal Order Accuracy, Disclosure Frequency and Testimony Consistency measures. Since preliminary ANOVAs revealed significant gender differences in PRI [F(1, 48) = 7.96, p = .007] and FSIQ [F(1, 48) = 4.06, p = .049] scores, these two cognitive measures were inputted as covariates. Overall, there were no significant gender differences on any of the dependent measures. Given these findings, child age, cognitive test scores and gender were not used as covariates for the following analyses.

Cognitive Interview versus Standard Interview

A logistic regression tested for differences between the CI and SI on whether the participants attempted to deceive the interviewer (0 = no; 1 = yes). A separate logistic regression evaluated differences between the CI and SI on whether participants maintained the secret until the end of the interview (0 = no; 1 = yes). There were no significant differences in deception attempts or maintenance between the SI and CI groups. For the truth versus false statement comparison analyses, the veracity group consisted of truth-tellers (n = 32) versus lie-tellers (n = 18) who maintained their story by the end of the interview. Chi-Square and ANOVA analyses indicated no age or cognitive differences between the veracity groups (see Table 5).

An ordinal regression examined interview group differences in children's Disclosure Forthcomingness across the four question-types levels (1. free-recall, 2. cognitive load, 3. non-direct closed-ended, and 4. direct inquiry closed-ended). The overall ordinal regression model was not significant. Most participants made their first disclosure about the ink spill (truthful accusation or false denial) on the free-recall question (n = 11, 22%) or on the final direct inquiry closed-ended question (n = 36, 72%); whereas, few children first disclosed the ink spill on the cognitive load (n = 1) and indirect closed-ended (n = 2) questions.

Next, a MANOVA evaluated differences between the SI and CI groups, irrespective of statement veracity, in total interview Response Length, Transgression Details, Events Recalled, Disclosure Frequency and Testimony Consistency. The overall model was significant, $F(5, 44) = 4.01$, $p = .004$. Significant interview group differences were found on the Response Length [$F(1, 48) = 12.00$, $p = .001$], Transgression Details [$F(1, 36) = 5.11$, $p = .028$], Events Recalled [$F(1, 48) = 15.70$, $p < .001$], and Disclosure Frequency [$F(1, 48) = 7.22$, $p = .010$] dependent measures. As expected, the CI group produced significantly more words (Response Length; *M difference* = +60.82 words, $p = .001$), Transgression Details (*M difference* = + 4.33 details, $p = .001$), events (*M difference* = + 2.57 events, $p < .001$) and disclosures about the ink spill (*M difference* = + .89 disclosures, $p = .010$) than the SI group. Despite the expected increased cognitive effort associated with the CI, there were surprisingly no interview group differences in Testimony Consistency. Since twelve children did not recall an event during the interview, analyses of these participants' Temporal Order Accuracy could not be completed. For this reason, a separate ANOVA was conducted for participants who did disclose at least one event (n = 38) to examine interview group differences in Temporal Order Accuracy. Overall, there were

no significant interview group differences on this measure. Table 6 shows the mean interview group differences across these dependent measures.

Two follow-up MANOVAs assessed for interview group differences in unique and repeated details (analysis 1), and unique and repeated Events Recalled (analysis 2). The overall MANOVA models for the Transgression Details [$F(2, 47) = 3.78$, $p = .030$] and Events Recalled [$F(2, 47) = 7.96$, $p = .001$] were significant. The CI group gave significantly more repeated Transgression Details (*M difference* = +2.01 details, $p = .009$) than the SI group, $F(1, 48) = 7.48$, $p = .009$; however, there were no differences in unique details recalled. Significant interview group differences were also found on the unique [$F(1, 48) = 15.89$, $p = .003$] and repeated [$F(1, 48) = 15.15$, $p < .001$] Events Recalled measures. As expected, the CI group gave significantly more unique (*M difference* = +1.13 events, $p = .003$) and repeated events (*M difference* = +1.54 events, $p < .001$) than the SI group (see Table 6).

Since both the SI and CI groups included the free-recall question, a separate MANOVA examined interview group differences in Response Length, Transgression Details, Events Recalled and Disclosure Frequency on this question exclusively. Overall, there were no significant interview group differences in their free-recall responses on any of the dependent measures. Lastly, since there were only 24 participants in the CI group, there was not enough statistical power to examine testimony content differences between the individual questions.

True versus False Narrative Characteristics

Three MANOVAs evaluated veracity group differences in Response Length, Transgression Details, Events Recalled and Testimony Consistency in the total sample (analysis 1), and within the SI (analysis 2) and CI (analysis 3) groups. Since not all truth-tellers and lie-tellers reported an event during the interview, analyses of the differences between veracity

groups in Temporal Order Accuracy could not be performed. The MANOVA model was significant for the first [$F(4, 45) = 5.44$, $p = .001$] and second [$F(4, 21) = 3.54$, $p = .023$] analyses. When analyzing the entire sample, there were significant differences between truth-tellers and lie-tellers on the Transgression Details measure only, $F(1, 48) = 6.40$, $p = .015$. As seen in Table 7, truth-tellers produced significantly more information pertaining to the magician's error (*M difference* = + 4.99 details, $p = .015$) than the lie-tellers. Despite the overall MANOVA model being significant, there were no significant differences between truth-tellers and lie-tellers on any of the dependent measures in the SI group. Within the CI group, however, significant differences in the number of Transgression Details disclosed were found between the truth-tellers and lie-tellers, $F(1, 22) = 6.99$, $p = .015$. As with analysis 1, truth-tellers (*M difference* = + 9.12 details, $p = .015$) disclosed more information about the ink spill throughout the CI than the lie-tellers. There were no significant differences between veracity groups across the three analyses on the other dependent measures.

Given these veracity group differences in Transgression Details disclosed, follow-up MANOVAs were performed to evaluate differences in unique and repeated Transgression Details in the total sample (analysis 1), and within the SI (analysis 2) and CI (analysis 3) groups. The overall MANOVA models were significant for the total sample [$F(2, 47) = 3.52$, $p = .037$] and the CI group [$F(2, 21) = 3.59$, $p = .045$]. For the total sample, truth-tellers gave significantly more unique [$F(1, 48) = 7.16$, $p = .010$; *M difference* = + 4.06 details] and repeated [$F(1, 48) = 3.65$, $p = .015$; *M difference* = + 1.63 details] Transgression Details. Furthermore, truth-tellers in the CI group gave significantly more unique ($F(1, 22) = 7.53$, $p = .012$; *M difference* = + 6.23 details) and repeated ($F(1, 22) = 4.90$, $p = .038$; *M difference* = + 3.33 details) Transgression

Details than the liars. Refer to Table 7 for veracity group scores across the unique, repeated and total Transgression Detail dependent measures.

An ordinal logistic regression examined for differences between veracity groups, irrespective of interview condition, in Disclosure Forthcomingness. The overall model was significant, $\chi 2$ (1, N = 50) = 8.10, p = .004, Nagelkerke R Square = .19. Timing of forthcomingness, with direct inquiry disclosures and the truth-tellers as the reference groups, was significantly different for the free-recall question, Wald (1) = 4.06, p = .044, 95% CIs = -1.50 to -.021. Follow-up binary logistic regression analyses indicated that lie-tellers were less likely, with marginal significance, to disclose the ink spill on the free-recall question (B = -2.25) compared to the truth-tellers, Wald (1) = 3.47, p = .060, 95% CIs = .01 to 1.11. The model was also significant when analyzing disclosures on the direct inquiry question, $\chi 2$ (1, N = 50) = 8.34, p < .004, Nagelkerke R Square = .22. For this analysis, lie-tellers were significantly more likely (B = 2.45) to make their false denial disclosure (i.e., falsely deny the ink spill) on the direct inquiry question compared to the truth-tellers, Wald (1) = 5.07, p = .024, 95% CIs = 1.37 to 98.52. Table 8 shows the rate of disclosures on each question-type according to veracity group.

Discussion

The current study was designed to inform interviewers on the efficacy of the commonly used forensic interviewing questions with CWID. While some studies have analyzed the efficacy of free-recall (e.g., Brown, et al., 2012), cognitive load (e.g., Gentle et al., 2013; Milne et al., 2013) and closed-ended (e.g., Brown, Lewis, Stephens, & Lamb, 2017) questions with this population, there is little research available that discusses CWID's eyewitness capacity when disclosing the transgression of another, especially when they are asked to lie. The current study

was therefore the first of its kind to directly examine the efficacy of multiple questioning styles in situations wherein CWID were asked to lie about another's transgression.

The first objective of this study was to examine for potential developmental predictors of children's testimony performance. As in study 1, preliminary analyses revealed no meaningful correlations between child age, WASI-II cognitive test performance and any of the testimony content dependent measures. Unlike study 1, however, there were no gender differences on any of the dependent measures. These unexpected findings may have resulted from utilizing a sample with extremely low VCI, PRI and FSIQ scores (see Table 1), which is consistent with their mild to moderate intellectual disability diagnoses. More specifically, the mean cognitive test scores in both interview groups were below more than 99% of their respective same-aged peers. This low level of cognitive functioning limited their abilities to respond to the open-ended questions, irrespective of their age or gender, especially when compared to TDC.

The second objective of this study was to examine whether the CI, when compared to the SI, improved children's eyewitness performance. As expected, the CI group produced significantly more words (Response Length), Transgression Details (repeated and total), Events Recalled (unique, repeated and total) and transgression disclosures (Disclosure Frequency) than the SI group. In the SI group, the CWID did not provide eyewitness statements with many words, details, events or disclosures. For example, the CWID's overall free-recall statements on the SI were only one-to-two sentences in length (*M Response Length* = 20 words) and contained fewer than two Transgression Details and Events Recalled. In contrast, the addition of the cognitive load questions in the CI group resulted in these children giving much more information about the ink spill and the magic show overall (see Table 6). This was expected because the cognitive load questions utilize mnemonic strategies that have consistently shown in the adult

literature (Memon et al., 2010) to increase eyewitness memory recall. However, because of the lack of CI literature with children, particularly those with cognitive deficits, it was unclear whether the cognitive load questions were too sophisticated for this vulnerable population. While it was not possible to determine which cognitive load questions were the most effective due to low statistical power, the current findings reinforce the limited research with CWID (e.g., Gentle et al., 2013) that suggests that the CI as a whole can potentially be used to encourage CWID to recall new information about a transgression they witnessed.

Surprisingly, there were no significant interview group differences in Report Honesty (conceal attempts and final decision to deceive) or Disclosure Forthcomingness. While nearly two-thirds of children eventually told the truth at the end of the interview (64%), the majority of them (78%) omitted the transgression altogether on the free-recall question. In contrast to the adult literature (Vrij et al., 2011, 2012), the cognitive load questions did not increase the cognitive effort associated with telling a lie; this was demonstrated by the fact that there were no significant differences between the CI (62.5% honest reports) and SI (65.4% honest reports) groups in children's final decisions to honestly report the ink spill. With respect to Disclosure Forthcomingness, children directly disclosed or falsely denied the ink spill on the free-recall (22%) or direct inquiry questions (72%); whereas, few children made their initial disclosure on the cognitive load (2%) or indirect closed- ended (4%) questions. Therefore, the cognitive load questions did not encourage the CWID to make new, or more truthful, disclosures about the magician's error when compared to standard interviewing procedures.

The increased cognitive effort on the CI, unexpectedly, did not reduce the accuracy of the children's reports. Children in both interview groups answered the five Testimony Consistency questions that inquired about the ink spill with good consistency (*M* score = 3.76 out of 5; 75.2%

consistency). Moreover, the CWID had 89.2% Temporal Order Accuracy of the events they recalled; however, this accuracy score is likely inflated given that most children did not recall many events to begin with (*M* = 2.32 Events Recalled). These results signify that CWID can provide consistent open and closed-ended eyewitness reports about another's transgressions on the CI, even when they had been coached to lie.

Given that no research to date has specifically analyzed the narrative characteristics of truthful and false statements made by CWID, the third objective of this study was to compare truth-tellers and lie-tellers across the dependent measures of eyewitness capacity. Consistent with the SVA literature (see Colwell et al., 2013, for a review), truth-tellers produced more forthcoming disclosures with a higher number of Transgression Details than the lie-tellers. Truth-tellers more readily discussed specific Transgression Details about the gloves and the magician's ink spill on the CI earlier in the interview when given the opportunity to freely recall their experiences. On the other hand, the lie-tellers typically omitted the final trick altogether during their responses to the free-recall and cognitive load questions, and instead, 94.4% of them made their eventual false disclosure on the direct inquiry closed-ended question. In the CI group only, truth-tellers also gave significantly more unique, repeated and total Transgression Details than the liars. For this reason, the addition of the cognitive load questions encouraged more memory recall among the truth-tellers, particularly with respect to the ink spill transgression. Conversely, the lie-tellers lied by omission and rarely added new information beyond their initial lie-scripts. These results provide new evidence that the CI can be used, in collaboration with SVA procedures, to distinguish between the truthful and false responses of CWID.

Directions for Future Research

There are several ways to improve on the current methodology. Firstly, given the clinical sample used for this study, it was only possible to obtain a final sample of 50 CWID. While, this sample was comparable to other experimental studies on the CI with atypical child populations (e.g., Milne et al., 2013; Robinson & McGuire, 2006), a larger sample would enable more thorough and generalizable data. Secondly, given that some CWID gave short responses on the cognitive load questions, adaptations are likely needed to these questions in order for them to be more effective with this vulnerable population. Some CI research has assessed the benefits of using sketches (Mattison et al., 2014) and reducing post-interview misinformation (Holliday, 2003); however, research is needed to evaluate whether these adaptations can benefit CWID.

Conclusions

The current study suggests that CWID can produce true and false eyewitness reports about a transgression they witnessed. When telling the truth, CWID discussed the ink spill earlier in the interview on the free-recall question; whereas, lie-tellers mostly falsely denied the transgression on the final direct inquiry closed-ended question. Truth-tellers interviewed using the CI also provided more total specific information about the gloves and the ink spill (Transgression Details) compared to the lie-tellers. When analyzing the value of the CI, the individual cognitive load questions encouraged honest speakers to provide new information about their experiences during the magic show and the transgression they witnessed, without compromising Temporal Order Accuracy or Testimony Consistency. However, these questions did not result in more honest or forthcoming disclosures when compared to the SI. Therefore, professionals can use free-recall and cognitive load questions with CWID to increase their eyewitness memory recall. At the same time, more research on potential developmentally

appropriate adaptations to these questions can further improve the credibility of the eyewitness reports given by CWID.

References

Aarons, N. M., Powell, M. B., & Browne, J. (2004). Police perceptions of interviews involving children with intellectual disabilities: A qualitative inquiry. *Policing and Society*, *14*, 269-278. doi:10.1080/10439460420000241848

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.

Brown, D. A., Lewis, C. N., Lamb, M. E., & Stephens, E. (2012). The influences of delay and severity of intellectual disability on event memory in children. *Journal of Consulting and Clinical Psychology*, *80*, 829-841. doi:10.1037/a0029388

Brown, D., Lewis, C., Stephens, E., & Lamb, M. (2017). Interviewers' approaches to questioning vulnerable child witnesses: The influences of developmental level versus intellectual disability status. *Legal and Criminological Psychology*, *22*, 332-349. doi:10.1111/lcrp.12104

Bull, R., Paulo, R.M., & Albuquerque, P.B. (2019). The impact of the cognitive interview in the UK and recent research in Portugal. In J.J. Dickinson, N.S. Compo, R. Carol, B.L. Schwartz, & M.R. McCauley (Eds). *Evidence-based investigative interviewing: Applying cognitive principles* (pp. 29-41). New York, NY: Routledge.

Colwell, K., Hiscock-Anisman, C. K., & Fede, J. (2013). Assessment Criteria Indicative of Deception (ACID): An example of the new paradigm of Differential Recall Enhancement. In B. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in eyewitness memory, credibility assessment, and investigative interviewing* (pp. 259–291). New York, NY: Springer.

Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C. Thomas Publisher Limited.

Gentle, M., Milne, R., Powell, M. B., & Sharman, S. J. (2013). Does the Cognitive Interview promote the coherence of narrative accounts in children with and without an intellectual disability? *International Journal of Disability, Development, and Education, 60,* 30-43. doi:10.1080/1034912X.2013.757138

Henry, L. A., & Gudjonsson, G. H. (1999). Eyewitness memory and suggestibility in children with mental retardation. *American Journal on Mental Retardation, 104,* 491-508. doi:10.1352/08958017(1999)104<0491:EMASIC>2.0.CO;2

Henry, L. A., & Gudjonsson, G. H. (2003). Eyewitness memory, suggestibility, and repeated recall sessions in children with mild and moderate intellectual disabilities. *Law and Human Behaviour, 27,* 481-505. doi:10.1023/A:1025434022699

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. doi:10.1111/j.1365-2788.2011.01383.x

Hershkowitz, I., Orbach, Y., Lamb, M. E., Sternberg, K. J., & Horowitz, D. (2001). The effects of mental context reinstatement on children's accounts of sexual abuse. *Applied Cognitive Psychology, 15,* 235-248. doi:10.1002/acp.699

Holliday, R. (2003). Reducing misinformation effects in children with Cognitive Interviews: Dissociating recollection and familiarity. *Child Development, 74,* 728-751. doi:10.1111/1467-8624.00565.

Jack, F., Leov, J., & Zajac, R. (2014). Age-related differences in the free-recall accounts of child, adolescent, and adult witnesses. *Applied Cognitive Psychology, 28,* 30-38. doi:10.1002/acp.2951

Jones, L., Bellis, M. A., Wood, S., Hughes, K., McCoy, E., Eckley, L., ... & Officer, A. (2012). Prevalence and risk of violence against children with disabilities: A systematic review and meta-analysis of observational studies. *The Lancet, 380,* 899-907. doi:10.1016/S0140-6736(12)60692-8

Mattison, M. L., Dando, C. J., & Ormerod, T. C. (2014). Sketching to remember: Episodic free recall task support for child witnesses and victims with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 45,* 1751-1765. doi:10.1007/s10803-0142335-z

Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law, 16,* 340-372. doi:10.1037/a0020518

Milne, R., Sharman, S. J., Powell, M. B., & Mead, S. (2013). Assessing the effectiveness of the Cognitive Interview for children with severe intellectual disabilities. *International Journal of Disability, Development and Education, 60,* 18-29. doi:10.1080/1034912X.2013.757137

Petersen, J. (2018). Gender difference in verbal performance: A meta-analysis of United States state performance assessments. *Educational Psychology Review, 30,* 1269-1281. doi:10.1007/s10648-018-9450-x

Pipe, M. E., & Wilson, J. C. (1994). Cues and secrets: Influences on children's event reports. *Developmental Psychology, 30,* 515-525. doi:10.1037/0012-1649.30.4.515

Robinson, J., & McGuire, J. (2006). Suggestibility and children with mild learning disabilities: The use of the Cognitive Interview. *Psychology, Crime & Law, 12,* 537-556. doi:10.1080/10683160500337550

Saykaly, C., Crossman, A., Morris, M., & Talwar, V. (2016). Question type and its effect on children's maintenance and accuracy during courtroom testimony. *Journal of Forensic Practice, 18,* 104-117. doi:10.1108/JFP-01-2015-0010

Smith, D. W., Letourneau, E. J., Saunders, B. E., Kilpatrick, D. G., Resnick, H. S., & Best, C. L. (2000). Delay in disclosure of childhood rape: Results from a national survey. *Child Abuse & Neglect, 24,* 273-287. doi:10.1016/S0145-2134(99)00130-1

Talwar, V., & Crossman, A. (2011). From little white lies to filthy liars: The evolution of honesty and deception in young children. *Advances in Child Development and Behaviour, 40,* 139-179. doi:10.1016/B978-0-12-386491-8.00004-9

Taylor, J., Stalker, K., & Stewart, A. (2016). Disabled children and the child protection system: A cause for concern. *Child Abuse Review, 25,* 60-73. doi:10.1002/car.2386

Thierry, K. L., & Spence, M. J. (2004). A real-life event enhances the accuracy of preschoolers' recall. *Applied Cognitive Psychology, 18,* 297-309. doi:10.1002/acp.965

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140,* 1174-1204. doi:10.1037/a0036620

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and its implications for professional practice.* Chichester, England: Wiley.

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology, 22,* 1-21. doi:10.1111/lcrp.12088

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science, 20,* 28-32. doi:10.1177/0963721410391245

Vrij, A., Leal, S., Mann, S., & Fisher, R. (2012). Imposing cognitive load to elicit cues to deceit: Inducing the reverse-order technique naturally. *Psychology, Crime & Law, 18* 579-594. doi:10.1080/1068316X.2010.515987

Wechsler, D., & Zhou, X. (2011). *Wechsler Abbreviated Scale of Intelligence, second edition* (WASI-II) *manual*. San Antonio, TX: Pearson Clinical Assessment.

Wyman, J., Foster, I., Crossman, A., Colwell, K., & Talwar V. (2019). The efficacy of free-recall, cognitive load, and closed-ended questions when children are asked to falsely testify about a crime. *Applied Cognitive Psychology,* Advance online publication, doi:10.1002/acp.3494

Wyman, J., Lavoie, J., & Talwar, V. (2018). Best practices for interviewing children with intellectual disabilities in maltreatment cases. *Exceptionality,* 1-18. doi:10.1080/09362835.2018.1425623

Table 1

Interview Group Sample Characteristics.

	Standard Interview (n = 26)	Cognitive Interview (n = 24)
Demographic Characteristics		
M_{age}	12.83	12.11
n_{males} (%)	17 (65.4%)	14 (58.3%)
Cognitive Characteristics		
FSIQ (*M*)	55.53	60.08
VCI (*M)*	59.34	57.38
PRI (*M*)*	56.62	67.20

* Difference significant at $p < .05$

Note. WASI cognitive scores below 70 are considered to be Extremely Low and indicate the presence of an intellectual disability.

Table 2

Organization and Scoring of the Cognitive Interview and Standard Interview.

Recall task	Phrase from recall task	Forthcomingness Category
1. Baseline questions	a. What is your favourite thing to do at school? b. What games do you like? c. Tell me everything you can remember about the breakfast you had today?	---
2. Free recall	Please describe, in as much detail as possible, everything that happened during the magic show? - Did anything else happen?	1
3. CL #1: Mental context reinstatement (CI only)	Now, tell me all the sights, sounds, smells, tastes and things you touched during the magic show. - Is there anything else?	2
4. Non-direct closed-ended block 1	a. Did anything happen that you did not think would happen? b. Were there winter boots in the classroom? c. Did the magician do something they were not supposed to while they were in the room?	3
5. CL #2: Another's perspective (CI only)	I want you to imagine that your parent was watching the magic show with you. Can you describe in as much detail everything s/he would remember from the show? - Would s/he remember anything else?	2
6. Non-direct closed-ended block 2	a. Did the magician use gloves during the show? b. Was the magician standing close to the door? c. Did the magician do something that might get them in trouble?	3
7. CL #3: Reverse-order (CI only)	Can you please describe everything you remember from the magic show, but backwards. Beginning with the last thing that happened at the magic show and ending with the first thing that happened. Please describe everything you remember from the magic show, but backwards. - Did anything else happen?	2
8. Non-direct closed-ended block 3	a. Did something happen to the gloves? b. Was the magician's gloves black or white? c. Do you think you could have made any mistakes about the things you said to me?	3
9. CL #4: Re-tell (CI only)	One last time, please describe everything you remember from the magic show. - Did anything else happen?	2
10. Non-direct closed-ended block 4	a. Did you see the magician's gloves? b. Did someone spill something on the magician's gloves?	3
11. Direct inquiry	Did the magician spill something on the gloves?	4

Table 3

Events Recalled Scoring Procedures.

Free-recall, mental context reinstatement, another perspective and re-tell questions scoring.	Unique Event (0 or 1)	Repeated Event (0 or 1)	Order Accuracy (0 or 1)
1. Trick #1: Magic Quarter			
2. Trick #2: Ball Disappearance			
3. Trick #3: Magnetized Wand			
4. Trick #4: Jail Escape			
5. Card Trick Teaching/Practice			
6. Ink Spill			
7. Concluding Magic Show			
TOTAL			

Reverse-order question scoring procedures.	Unique Event (0 or 1)	Repeated Event (0 or 1)	Order Accuracy (0 or 1)
1. Concluding Magic Show			
2. Ink Spill			
3. Card Trick Teaching/Practice			
4. Trick #4: Jail Escape			
5. Trick #3: Magnetized Wand			
6. Trick #2: Ball Disappearance			
7. Trick #1: Magic Quarter			
TOTAL			

TOTALS (combined scores across 5-questions)

Unique Event Score: _____

Repeated Event Score: _____

Event Order Score: _____

Table 4

Testimony Consistency Scoring Procedures.

Question	Truthful Response Consistency (out of 5)	Lie-telling Response Consistency (out of 5)
Did the magician do something they were not supposed to do while they were in room?	YES	NO
Did the magician do something that might get them in trouble?	YES	NO
Did something good or bad happen to the gloves?	BAD	GOOD (or nothing)
Did someone spill something on the magician's gloves?	YES	NO
Did the magician spill something on the magician's gloves?	YES	NO

Table 5

Veracity Group Sample Differences.

	Truth-tellers (*n* = 32)	Lie-tellers (*n* = 18)
Mage	12.28	12.85
Cognitive Characteristics		
FSIQ (*M*)	57.16	58.72
VCI (*M)*	58.13	58.89
PRI (*M*)	61.25	62.50

* There were no significant differences between veracity groups on the age and cognitive measures.

Table 6

Mean (SD) Differences between the CI and SI in Total Interview Response Length, Transgression Details, Events Recalled, Temporal Order Accuracy and Testimony Consistency.

	Cognitive Interview (n = 24)	Standard Interview (n = 26)	Total Sample (N = 50)
Response Length**	81.17	20.25	49.54
Transgression Details*	6.15	1.81	3.89
Unique Details	4.02	1.69	2.81
Repeated Details*	2.13	0.12	1.08
Total Events Recalled**	3.71	1.04	2.32
Unique Events**	2.17	1.04	1.58
Repeated Events**	1.54	0	0.74
Temporal Order Accuracy	86.63%	92.13%	89.23%
Disclosure Frequency*	1.08	0.19	0.62
Testimony Consistency	72.60%	77.80%	75.20%

* Difference between interview groups significant at $p < .05$.
** Difference between interview groups significant at $p < .01$.

Table 7

Mean Unique, Repeated and Total Transgression Details According to Veracity Group.

Standard Interview Condition (*n* = 26)			
	Unique Details	Repeated Details	Total Details
Truth-tellers	2.15	0.11	2.27
Lie-tellers	0.83	0.12	0.94
Total	1.69	0.12	1.81
Cognitive Interview Condition (*n* = 24)			
	Unique Details*	Repeated Details*	Total Details*
Truth-tellers	6.23	3.33	9.57
Lie-tellers	0.33	0.11	0.44
Total	4.02	2.13	6.15
Total Sample (*n* = 50)			
	Unique Details*	Repeated Details	Total Details*
Truth-tellers	4.06	1.63	5.69
Lie-tellers	0.58	0.11	0.69
Total	2.81	1.08	3.89

* Veracity group differences significant at $p < .05$.

Table 8

Disclosure Forthcomingness (%) According to Testimony Veracity Group.

	Truth-tellers (*n* = 32)	Lie-tellers (*n* = 18)
Free recall Disclosures*	10 (31.3%)	1 (5.6%)
Cognitive load Disclosures	1(3.1%)	0
Non-direct Closed-ended Disclosures	2 (6.3%)	0
Direct Inquiry Disclosures*	19 (59.4%)	17 (94.4%)

* Significant at $p < .05$

Chapter 6: General Discussion

Despite a child's eyewitness testimony being integral to investigations of child maltreatment, researchers and forensic professionals continue to express concerns with current forensic interviewing procedures (see Lyon, 2014; Wyman, Lavoie, et al., 2018, for reviews). While free-recall questioning is widely recommended given that it encourages children to disclose information openly in their own words (Guadagno et al., 2006; Orbach et al., 2000), responses to these questions can nevertheless be overly broad and uninformative (Lyon, 2014). Alternatively, cognitive load questions in the CI have shown to be effective with adults for increasing the amount of information recalled and for discouraging false reports (see Memon et al., 2010; Vrij et al., 2011, 2012). However, the relatively little forensic research on the CI with children is inconclusive. Some studies suggest that these questions improve memory recall performance (e.g., Gentle et al., 2013; McCauley & Fisher, 1995); whereas other studies indicate that these questions may be too complex for younger children (Saykaly et al., 2016; Wyman et al., 2019). Given these gaps in the forensic literature, the two studies in this research program were designed to provide researchers and forensic professionals with new and comprehensive information about the efficacy of cognitive load questioning with TDC and CWID.

Developmental Findings

The current research program was one of the first to analyze ways of improving the eyewitness capacity of both TDC and CWID. The findings from both studies indicate that school-aged children with and without intellectual disabilities can give detailed and accurate eyewitness reports when interviewed using the CI about an adult's transgression. In both studies, children recalled more information on the CI when given more opportunities to openly recall their experiences, including more relevant (i.e., Transgression Details and Disclosure Frequency)

and irrelevant (i.e., higher Response Length and Events Recalled) information pertaining to the transgressions. In contrast, both the TDC and CWID gave relatively little important information about the transgressions they witnessed on the SI. This was expected given that the SI was designed to resemble bad, yet frequently used, interviewing strategies as this interview protocol primarily consisted of closed-ended recognition questions. Despite the CI being more cognitively taxing than the SI, there were no differences between interview groups in the accuracy or consistency of the reports given by the TDC and CWID. Interviewers should therefore utilize open-ended questions with children, including those with limited intellectual capacities, as they are able to produce more informative eyewitness reports on these questions.

Both studies found that child age and cognitive test scores were not positively related to performance on any of the dependent measures. Children's cognitive skills and verbal performance usually improve as they get older, and thus, the TDC and CWID's testimony performance was expected to improve with age. In the second study, the relatively smaller sample size likely made age-related comparisons difficult. For the first study, an older sample was used compared to some of the other CI child studies that found age-related differences (e.g., Wyman et al., 2019). As noted in Memon and colleagues' (2010) meta-analysis, cognitive load questions are least effective with children under the age of seven. Since the first study included children over the age of 8, the TDC in this study benefited from the CI.

Furthermore, gender differences in eyewitness performance were found in the first study only. Notably, TDC girls gave statements on the CI that contained more words, Transgression Details, disclosures and Events Recalled when compared to the boys. These gender differences reinforce past developmental and adult eyewitness research (Suckle-Nelson et al., 2010) that

indicates that girls perform better on verbal recollection tasks than boys (Hyde, 2016; Petersen, 2018).

Interviewing Contributions

The primary purpose of both studies was to examine the efficacy of the CI, when compared to a SI, with TDC and CWID. In both studies, children produced significantly more words, Events Recalled, Transgression Details and disclosures on the CI when compared to the SI. Even though more information was disclosed, the Temporal Order Accuracy and consistency of their reports were not reduced on the CI. These findings were expected given that prior research with adults (see Bull et al., 2019; Memon et al., 2010, for reviews) and children (e.g., Gentle et al., 2013; McCauley & Fisher, 1995) indicated that the use of cognitive load mnemonic strategies increases eyewitness memory among truthful child witnesses. Surprisingly, however, the CI did not discourage false reports (Vrij et al., 2011, 2012), nor did it increase Disclosure Forthcomingness when compared to the SI. Among the TDC, the first free-recall question was the most effective for increasing the amount of words, unique Events Recalled, and unique Transgression Details and disclosures. Consistent with prior research (Saywitz et al., 1992; Wyman et al., 2019), the reverse-order recall and re-tell (second free-recall question) questions were more effective than the mental context reinstatement and another perspective questions for increasing Response Length, repeated Transgression Details and repeated Events Recalled.

Implications for Testimony Credibility Evaluations

The findings from both studies provide important information regarding how TDC and CWID formulate their truthful and false statements. Consistent with the adult deception literature (see Colwell et al., 2013, for a review), the use of cognitive load questioning in both studies led to significant narrative differences in the statements given by the truth-tellers and lie-

tellers. Among the TDC and CWID witnesses, the honest speakers were more willing to make a disclosure about the transgression they witnessed earlier in the interview, and they also gave significantly more unique and repeated information about the transgression. In contrast, the lie-tellers likely developed a lie-script since they were trying to convince the interviewer that the transgression they witnessed did not actually take place. To ensure that they consistently maintained their fabricated report, they discussed irrelevant experiences (e.g., the cognitive tasks and other magic tricks) and relied on the direct inquiry forced-choice question to falsely deny the transgression. Moreover, the TDC lie-tellers actually produced more consistent statements across the closed-ended questions than the truth-tellers. Therefore, the truth-tellers likely focused on recalling more information from memory given that they believed their honesty was transparent, while the lie-tellers were more concerned with giving short responses that were consistent with their initial lie-script (see Vrij, Granhag, & Porter, 2011, for a review). In both studies, these narrative differences were not present when analyzing the statements within the SI group. Thus, the addition of the cognitive load questions led to discernable differences in the testimony characteristics of children's true and false reports.

Both studies utilized evidence-based SVA lie-detection coding procedures (e.g., Response Length and Transgression Details) to evaluate the eyewitness performance of the TDC and CWID. Given the truth versus lie narrative differences in both studies, these SVA coding tools have the potential to be useful with TDC and CWID. In addition to these tools, Disclosure Forthcomingness is another variable that should be considered when examining differences between truthful accusations and false denials. While this dependent measure is a relatively recent addition to the eyewitness literature (Malloy & Mugno, 2016), results from the two

current studies and others (e.g., Wyman et al., 2019) indicate that truth-tellers are more willing to make a clear disclosure earlier in the interview when compared to lie-tellers.

Research Contributions

The research paradigms used in both studies have positive implications for future eyewitness research with children. First, both studies expand on the prior interviewing literature (e.g., Gentle et al., 2013; Pipe & Wilson, 1994) by using a research paradigm whereby children witnessed and participated in a transgression. This target event generalizes to real-life settings whereby children are questioned about a familiar adult's transgression, and to situations wherein they may feel complicit in the illegal activity (Romeo et al., 2018). Consistent with real-life situations whereby children feel pressure to falsely deny or omit the crimes of others (Black, Schweitzer, & Varghese, 2012; Schreiber et al., 2006), the children in both studies were also coached by the transgressor to conceal their misdeed. By using this paradigm, it was possible to assess whether the CI can both discourage and detect children's false eyewitness reports.

Second, the current findings provide legal professionals and researchers with a much more complete picture of the strengths and weaknesses of the commonly used free-recall and cognitive load questions. Prior research typically used a smaller number of dependent measures to examine the efficacy of each interview, such as only evaluating the accuracy and number of details recalled (e.g., Gentle et al., 2013; Robinson & McGuire, 2006). In addition to these measures, the current studies also examined other important measures of statement quantity and quality. By analyzing the honesty of children's disclosures for example, it was possible to assess whether the CI discouraged intentional dishonest reports. Given that many children fail to disclose another's transgression when first being interviewed (Hershkowitz & Terner, 2007), the Disclosure Forthcomingness and Disclosure Frequency dependent measures inform forensic

professionals about whether the CI increases children's willingness to reveal the misdeeds of others. Whereas, the Transgression Details, Response Length and Events Recalled measures provided necessary information about whether the CI increased the amount of relevant and irrelevant information recalled by the children. Finally, the Temporal Order Accuracy and Testimony Consistency measures advise how each interviewing tool affected the accuracy and consistency of children's reports. Future forensic research should utilize these dependent measures when examining ways to improve children's eyewitness performance.

General Limitations and Directions for Future Research

There are several ways to improve on the research methodology in both studies. As discussed in the first manuscript, it is recommended that future research have a parent coach the child into producing a false report (e.g., Talwar et al., 2004), given that they are the most common perpetrators of crimes against children (Wissink et al., 2015). Moreover, incorporating a time delay between the witnessed target event and the actual forensic interview (e.g., Peterson & Whalen, 2011) will better resemble real-life forensic cases wherein children are interviewed. With respect to Study 2, a larger sample of children with a broader range of intellectual deficits, including those with severe or profound intellectual disabilities (e.g., Brown et al., 2012), would enable for more thorough and generalizable data on the efficacy of cognitive load questioning with CWID.

In addition to these recommendations, both studies suggest that there is an important need for adapting some of the cognitive load questioning strategies in order for them to be more effective with typical and atypical child populations. As highlighted in the first study, as well as in other forensic child studies (e.g., Saywitz et al., 1992; Wyman et al., 2019), the Mental Context Reinstatement and Another Perspective questions were largely ineffective with TDC.

Since these questions are very effective with adults (Memon et al., 2010), future research should analyze ways of modifying these questions so that they can be beneficial to child witnesses. For example, breaking up these questions into smaller parts (Wyman, Lavoie, et al., 2018), utilizing sketches instead of only verbal recollections (Mattison et al., 2014), reducing post-target event misinformation (Holliday, 2003), and narrative elaboration strategies that emphasize organizing elements of the event into psychologically salient categories (Saywitz & Snyder, 1996), are potential tools for making these questions more accessible to children. More research should assess whether these adaptations can improve typical and atypical child eyewitness responsiveness to these complex, yet potentially effective, cognitive load questions.

In both studies, children were asked to falsely deny the transgression of an adult. While false denial lies are a common type of false reporting in the eyewitness literature (Smith et al., 2000), there are other types of lies that take place in investigative interviewing contexts. For instance, eyewitnesses can be pressured to falsely accuse an innocent person of a crime they did not commit, such as in cases of custody battles (Black, Schweitzer, & Varghese 2012). At the same time, child eyewitnesses may falsely deny a crime because they feel complicit or responsible for the transgression (Hershkowitz et al., 2007); this is especially relevant to cases of youth with behaviour difficulties, who are more likely to lie to conceal their own misdeeds (Lavoie, Wyman, Crossman, & Talwar, 2018). Finally, some research suggests that there may be narrative differences in children's CI reports based on the type of lie they are telling. In Wyman and colleagues (2019), children's true and false accusation reports included more recalled information on the CI than denial reports. For these reasons, future research should assess whether the CI can be used to discourage and detect multiple types of false reports, including false denials, false accusations and lies told to protect oneself and others.

In both studies, the CI resulted in important narrative differences in the testimony content of truth-tellers and lie-tellers. Nevertheless, it is unclear whether these narrative differences can lead to improved truth and lie-detection among forensic professionals and laypersons. Among the adult studies (e.g., Colwell et al., 2007, 2013; Vrij et al., 2011, 2012, 2017), the use of the CI led to improved lie-detection among law enforcement and mock-jurors who were trained using SVA deception detection coding procedures. In a recent child study (Wyman, Foster et al., 2018), however, the use of the CI did not significantly improve laypersons' detection of children's initial and recanted false reports above chance level. It is worth noting, though, that this study did not have a SI control condition, nor were the adults trained on evidence-based SVA coding procedures. Future lie-detection research should assess whether training on the SVA procedures used in both studies can lead to better child eyewitness credibility evaluations.

Lastly, the SI in both studies did not include the same number of questions as the CI. Whereas the SI only had one free-recall question, the CI included five open-ended questions (free-recall + four cognitive load questions). Thus, it could be argued that the CI as whole was only more effective than the SI because the TDC and CWID were given more opportunities to openly recall their prior experiences during the study. To address this limitation, the first study thoroughly analyzed potential question-type differences in children's recall performance on the individual free-recall and cognitive load questions; this was not possible in the second study due to the smaller sample of CWID. To permit for better interview protocol comparisons, future studies should ensure that the control interview group answers the same number of questions as the experimental group. Interview group differences will therefore be fully based on differences in the attributes of their respective questions, rather than the number of questions asked.

Conclusions

The two studies in this research program comprehensively evaluated the effectiveness of commonly used forensic interviewing procedures with two populations who are most vulnerable to being maltreated (Jones et al., 2012). In both studies, the CI increased the Response Length, Transgression Details and Events Recalled, without compromising Temporal Order Accuracy or Testimony Consistency. In the first study, the free-recall question was more effective for encouraging detailed reports. Moreover, the reverse-order and re-tell questions were the most effective cognitive load questions as TDC recalled significantly more repeated transgression details and events when completing these tasks. Conversely, the mental context reinstatement and other perspective questions showed minimal benefits for increasing the quantity and quality of children's eyewitness reports. Contrary to our hypotheses (Vrij et al., 2011, 2012), the CI did not dissuade false reporting in either study, nor did it encourage more forthcoming disclosures. These results indicate that the CI can be a potentially useful interviewing tool for increasing memory recall in TDC and CWID, without compromising accuracy. Nevertheless, modifications are needed to the mental context reinstatement and another perspective questions in order for TDC and CWID to benefit from these mnemonic strategies.

From a developmental perspective, findings from the second study suggest that CWID can provide detailed and accurate eyewitness reports when being interviewed using the CI. Notably, this often underestimated population recalled detailed, consistent and accurate information on the CI when honestly reporting about the magic show and transgression they witnessed. Legal professionals often ask unreliable and suggestive closed-end questions with CWID due to their lower confidence in these children's abilities to provide credible open-ended reports (Andrews et al., 2015; Guadagno et al., 2006). The current findings, however, indicate

that interviewers should utilize more reliable free-recall and cognitive load questions (Memon et al., 2010; Quas et al., 2000) with CWID as these participants were able to provide detailed and accurate recollections in response to these questions. Surprisingly, there was no meaningful relationship between child age and cognitive test performance on the eight dependent measures in either study. Among the TDC, however, girls produced open-ended statements with higher Response Length, Transgression Details, Events Recalled and Disclosure Frequency scores on the CI when compared to the boys.

Finally, both studies have important implications for lie-detection research with children. In correspondence with Wyman and colleagues (2019), there were significant differences in the statements provided by the honest and dishonest speakers. Among the TDC and CWID, truth-tellers gave more forthcoming and detailed disclosures about the transgressions they witnessed when compared to the liars. Conversely, liars mostly omitted information altogether about their transgression and instead discussed unrelated topics, such as the cognitive tasks or other magic tricks. Whereas liars made their false denial disclosure on the final direct inquiry question, truth-tellers were more willing to make their true accusation reports on the free-recall question. These findings indicate that the use of cognitive load questions can lead to discernable differences in the statements of truth-tellers and liars. Further, these findings suggest that the SVA lie-detection coding procedures used in these two studies, which are frequently used with adults (Colwell et al., 2007, 2013; Vrij et al., 2011, 2012, 2017), can also be effective methods for detecting truth and lies in the statements of TDC and CWID.

Bibliography

Aarons, N. M., Powell, M. B., & Browne, J. (2004). Police perceptions of interviews involving children with intellectual disabilities: A qualitative inquiry. *Policing and Society, 14,* 269-278. doi:10.1080/10439460420000241848

Akehurst, L., Köhnken, G., & Höfer, E. (2001). Content credibility of accounts derived from live and video presentations. *Legal and Criminological Psychology, 6,* 65-83. doi:10.1348/135532501168208

Akehurst, L., Milne, R., & Kohnken, G. (2003). The effects of children's age and delay on recall in a cognitive or structured interview. *Psychology, Crime & Law, 9,* 97–107. doi:10.1080/10683160308140

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.

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

Berntsen, D., & Rubin, D. C. (2006). Emotion and vantage point in autobiographical. *Cognition and Emotion, 20,* 1193-1215. doi:10.1080/02699930500371190

Berryhill, M. E., & Olson, I. R. (2008). The right parietal lobe is critical for visual working memory. *Neuropsychologia, 46,* 1767-1774. doi:10.1016/ j.neuropsychologia.2008.01.009

Black, F., Schweitzer, R., & Varghese, F. (2012). Allegations of child sexual abuse in family court cases: A qualitative analysis of psychiatric evidence. *Psychiatry, Psychology and Law, 19,* 482-496. doi:10.1080/13218719.2011.613905

Bowles, P. V., & Sharman, S. J. (2014). A review of the impact of different types of leading interview questions on child and adult witnesses with intellectual disabilities. *Psychiatry, Psychology and Law, 21,* 205-217. doi:10.1080/13218719.2013.803276

Brown, D. A., Lewis, C. N., Lamb, M. E., & Stephens, E. (2012). The influences of delay and severity of intellectual disability on event memory in children. *Journal of Consulting and Clinical Psychology, 80,* 829-841. doi:10.1037/a0029388

Brown, D., Lewis, C., Stephens, E., & Lamb, M. (2017). Interviewers' approaches to questioning vulnerable child witnesses: The influences of developmental level versus intellectual disability status. *Legal and Criminological Psychology, 22,* 332-349. doi:10.1111/lcrp.12104

Brown, S. W., & Smith-Petersen, G. A. (2014). Time perception and temporal order memory. *Acta Psychologica, 148,* 173-180. doi:10.1016/j.actpsy.2014.02.003

Bull, R. (2018). Credible accounts: What they are and how to obtain them. *Linguistics and the Human Sciences, 12*(2-3), 166-176. doi:10.1558/lhs.36990

Bull, R., Paulo, R.M., & Albuquerque, P.B. (2019). The impact of the cognitive interview in the UK and recent research in Portugal. In J.J. Dickinson, N.S. Compo, R. Carol, B.L. Schwartz, & M.R. McCauley (Eds). *Evidence-based investigative interviewing: Applying cognitive principles* (pp. 29-41). New York, NY: Routledge.

Bull, R., Valentine, T., & Williamson, T. (2009). *Handbook of psychology of investigative interviewing: Current developments and future directions*. West Sussex, UK: John Wiley & Sons.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11,* 73-92. doi:10.1002/icd.298

Cederborg, A. C., Danielsson, H., La Rooy, D., & Lamb, M. E. (2009). Repetition of contaminating question types when children and youths with intellectual disabilities are interviewed. *Journal of Intellectual Disability Research, 53,* 440–449. doi:10.1111/j.1365-2788.2009.01160.x

Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., & McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: Evidence from activation likelihood estimate meta-analyses. *Cerebral Cortex, 19,* 1557-1566. doi:10.1093/cercor/bhn189

Colwell, K., Hiscock-Anisman, C. K., & Fede, J. (2013). Assessment Criteria Indicative of Deception (ACID): An example of the new paradigm of Differential Recall Enhancement. In B. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in eyewitness memory, credibility assessment, and investigative interviewing* (pp. 259–291). New York, NY: Springer.

Colwell, K., Hiscock, C. K, & Memon, A. (2002). Interviewing techniques and the assessment of statement credibility**.** *Applied Cognitive Psychology, 16,* 287-300. doi:10.1002/acp.788

Colwell, K., Hiscock-Anisman, C., Memon, A., Rachel, A., & Colwell, L. (2007). Vividness and spontaneity of statement detail characteristics as predictors of witness credibility. *American Journal of Forensic Psychology, 25,* 5-30. doi:10.07331290

Davis, J.L., & Matthews, R. (2010). NEPSY-II review. *Journal of Psychoeducational Assessment, 28*(2), 175-182. doi:10.117/0734282909346716

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74–118. doi:10.1037/0033-2909.129.1.74

Eisen, M. L., Goodman, G. S., Qin, J., Davis, S., & Crayton, J. (2007). Maltreated children's memory: Accuracy, suggestibility, and psychopathology. *Developmental Psychology, 43*, 1275-1294. doi:10.1037/0012-1649.43.6.1275

Evans, A. D., Xu, F., & Lee, K. (2011). When all signs point to you: Lies told in the face of evidence. *Developmental Psychology, 47*, 39-49. doi:10.1037/a0020787

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

Faller, K. C. (2016). Disclosure failures: Statistics, characteristics, and strategies to address them. In W. T. O'Donohue & M. Fanetti (Eds.), *Forensic interviews regarding child sexual abuse: A guide to evidence-based practice* (pp. 123–139). Switzerland: Springer International Publishing.

Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C. Thomas Publisher Limited.

Geiselman, R. E. (1984). Enhancement of eyewitness memory: An empirical evaluation of the cognitive interview. *Journal of Police Science & Administration, 12*, 74-80. Retrieved from http://www.ncjrs.gov/App/publications/abstract.aspx?ID=93116

Geiselman, R. E., & Callot, R. (1990). Reverse versus forward recall of script-based texts. *Applied Cognitive Psychology, 4,* 141-144. doi:10.1002/acp.2350040206

Geiselman, R. E., & Padilla, J. (1988). Cognitive interviewing with child witnesses. *Journal of Police Science & Administration, 16,* 236-242. Retrieved from https://psycnet.apa.org/record/1990-19437-001

Gentle, M., Milne, R., Powell, M. B., & Sharman, S. J. (2013). Does the Cognitive Interview promote the coherence of narrative accounts in children with and without an intellectual disability? *International Journal of Disability, Development, and Education, 60,* 30-43. doi:10.1080/1034912X.2013.757138

Gilbert, R., Widom, C. S., Browne, K., Fergusson, D., Webb, E., & Janson, S. (2009). Burden and consequences of child maltreatment in high-income countries. *The Lancet, 373,* 68-81. doi:10.1016/S0140-6736(08)61706-7

Guadagno, B. L., Powell, M. B., & Wright, R. (2006). Police officers' and legal professionals' perceptions regarding how children are, and should be, questioned about repeated abuse. *Psychiatry, Psychology and Law, 13,* 251-260. doi:10.1375/pplt.13.2.251

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion, 4,* 201-206. doi:10.1037/1528-3542.4.2.201

Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A girl advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behaviour, 27,* 401-416. doi:10.1016/j.evolhumbehav.2006.05.002

Heider, K. G. (1988). The Rashomon effect: When ethnographers disagree. *American Anthropologist, 90,* 73-81. doi:10.1525/aa.1988.90.1.02a00050

Henry, L. A., & Gudjonsson, G. H. (1999). Eyewitness memory and suggestibility in children with mental retardation. *American Journal on Mental Retardation, 104*, 491-508. doi:10.1352/08958017(1999)104<0491:EMASIC>2.0.CO;2

Henry, L. A., & Gudjonsson, G. H. (2003). Eyewitness memory, suggestibility, and repeated recall sessions in children with mild and moderate intellectual disabilities. *Law and Human Behaviour, 27*, 481-505. doi:10.1023/A:1025434022699

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. doi:10.1111/j.1365-2788.2011.01383.x

Hershkowitz, I., Lamb, M. E., Orbach, Y., Katz, C., & Horowitz, D. (2012). The development of communicative and narrative skills among preschoolers: Lessons from forensic interviews about child abuse. *Child Development, 83,* 611-622. doi:10.1111/j.1467-8624.2011.01704.x

Hershkowitz, I., Lanes, O., & Lamb, M. E. (2007). Exploring the disclosure of child sexual abuse with alleged victims and their parents. *Child Abuse & Neglect, 31*, 111-123. doi:10.1016/j.chiabu.2006.09.004

Hershkowitz, I., Orbach, Y., Lamb, M. E., Sternberg, K. J., & Horowitz, D. (2001). The effects of mental context reinstatement on children's accounts of sexual abuse. *Applied Cognitive Psychology, 15*, 235-248. doi:10.1002/acp.699

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

Hershkowitz, I., & Terner, A. (2007). The effects of repeated interviewing on children's forensic statements of sexual abuse. *Applied Cognitive Psychology, 21,* 1131-1143. doi:10.1002/acp.1319

Holliday, R. (2003). Reducing misinformation effects in children with Cognitive Interviews: Dissociating recollection and familiarity. *Child Development, 74,* 728-751. doi:10.1111/1467-8624.00565.

Holliday, R. E., Brainerd, C. J., Reyna, V. F., & Humphries, J. E. (2009). The cognitive interview: Research and practice across the lifespan. In R. Bull, T. Valentine, & T. Williamson (Eds.), *Handbook of the psychology of investigative interviewing: Current developments and future directions* (pp. 137–160). Oxford: Wiley.

Howe, M. L., Toth, S. L., & Cicchetti, D. (2011). Can maltreated children inhibit true and false memories for emotional information? *Child Development, 82,* 967-981. doi:10.1037/a0025160

Hyde, J. S. (2016). Sex and cognition: Gender and cognitive functions. *Current Opinion in Neurobiology, 38,* 53-56. doi:10.1016/j.conb.2016.02.007

Jack, F., Friedman, W., Reese, E., & Zajac, R. (2016). Age-related differences in memory for time, temporal reconstruction, and the availability and use of temporal landmarks. *Cognitive Development, 37,* 53-66. doi:10.1016/j.cogdev.2015.12.003

Jack, F., Leov, J., & Zajac, R. (2014). Age-related differences in the free-recall accounts of child, adolescent, and adult witnesses. *Applied Cognitive Psychology, 28,* 30-38. doi:10.1002/acp.2951

Jones, L., Bellis, M. A., Wood, S., Hughes, K., McCoy, E., Eckley, L., ... & Officer, A. (2012). Prevalence and risk of violence against children with disabilities: A systematic review

and meta-analysis of observational studies. *The Lancet, 380*, 899-907. doi:10.1016/S0140-6736(12)60692-8

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: Administration manual.* San Antonio, TX: The Psychological Corporation.

Lamb, M. E. (1998). Assessments of children's credibility in forensic contexts. *Current Directions in Psychological Science, 7*, 43-46. doi:10.1111/1467-8721.ep13175602

Lamb, M. E., Orbach, Y., Hershkowitz, I., Esplin, P. W., & 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 & Neglect, 31*, 1201-1231. doi:10.1016/j.chiabu.2007.03.021

Lavoie, J., Wyman, J., Crossman, A. M., & Talwar, V. (2018). Lie-telling as a mode of antisocial action: Children's lies and behaviour problems. *Journal of Moral Education, 47*, 432-450. doi:10.1080/03057240.2017.1405343

Lewy, J., Cyr, M., & Dion, J. (2015). Impact of interviewers' supportive comments and children's reluctance to cooperate during sexual abuse disclosure. *Child Abuse & Neglect, 43*, 112-122. doi:10.1016/j.chiabu.2015.03.002

Lightfoot, E. B., & LaLiberte, T. L. (2006). Approaches to child protection case management for cases involving people with disabilities. *Child Abuse & Neglect, 30*, 381-391. doi:10.1016/j.chiabu.2005.10.013

Liu, M., Granhag, P. A., Landström, S., Hjelmsaeter, E. R. A., Strömwall, L., & Vrij, A. (2010). "Can you remember what was in your pocket when you were stung by a bee?': Eliciting cues to deception by asking the unanticipated. *Open Criminology Journal, 3*, 31-36. doi:10.2174/1874917801003010031

London, K., Bruck, M., Ceci, S. J., & Shuman, D. W. (2005). Disclosure of child sexual abuse: What does the research tell us about the ways that children tell? *Psychology, Public Policy, and Law, 11*, 194-226. doi:10.1037/1076-8971.11.1.194

Lyon, T. D. (1999). The new wave in children's suggestibility research: A critique'. *Cornell Law Review, 84*, 1004-1087. Retrieved from https://works.bepress.com/thomaslyon/1/

Lyon, T. D. (2007). False denials: Overcoming methodological biases in abuse disclosure research. In M. E. Pipe, M. Lamb, Y. Orbach, & A. Cederborg (Eds.), *Disclosing abuse: Delays, denials, retractions, and incomplete accounts* (pp. 41–62). Mahwah, NJ: Laurence Earlbaum.

Lyon, T. D. (2014). Interviewing children. *Annual Review of Law and Social Science, 10,* 73-89. doi:10.1146/annurev-lawsocsci-110413-030913

Lyon, T. D., & Dorado, J. S. (2008). Truth induction in young maltreated children: The effects of oath-taking and reassurance on true and false disclosures. *Child Abuse & Neglect, 32,* 738-748. doi:10.1016/j.chiabu.2007.08.008

Malloy, L. C., & Mugno, A. P. (2016). Children's recantation of adult wrongdoing: An experimental investigation. *Journal of Experimental Child Psychology, 145*, 11–21. doi:10.1016/j.jecp.2015.12.003

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology, 89*, 137–149. doi:10.1037/0021-9010.89.1.137

Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B., & Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search.

Proceedings of the National Academy of Sciences, *108*, 12893-12897. doi:10.1073/pnas.1015174108

Mattison, M. L., Dando, C. J., & Ormerod, T. C. (2014). Sketching to remember: Episodic free recall task support for child witnesses and victims with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *45*, 1751-1765. doi:10.1007/s10803-0142335-z

McCauley, M. R., & Fisher, R. P. (1995). Facilitating children's eyewitness recall with the revised cognitive interview. *Journal of Applied Psychology*, *80*, 510-516. doi:10.1037/0021-9010.80.4.510

Memon, A., Cronin, O., Eaves, R., & Bull, R. (1993). The cognitive interview and child witnesses. *Issues in Criminological & Legal Psychology*, *20*, 3-9. Retrieved from https://shop.bps.org.uk/publications/publication-by-series/issues-in-criminological-and-legal-psychology.html

Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, *16*, 340-372. doi:10.1037/a0020518

Milne, R., & Bull, R. (1996). Interviewing children with mild learning disability with the cognitive interview. *Issues in Criminological & Legal Psychology*, *26*, 44-51. Retrieved from https://psycnet.apa.org/record/1998-10274-006

Milne, R., & Bull, R. (2003). Does the cognitive interview help children to resist the effects of suggestive questioning? *Legal and Criminological Psychology*, *8*, 21-38. doi:10.1348/135532503762871219

Milne, R., Sharman, S. J., Powell, M. B., & Mead, S. (2013). Assessing the effectiveness of the Cognitive Interview for children with severe intellectual disabilities. *International Journal of Disability, Development and Education, 60,* 18-29. doi:10.1080/1034912X.2013.757137

Mirandola, C., & Toffalini, E. (2018). What happened first? Working memory and negative emotion tell you better: evidence from a temporal binding task. *Cognition and Emotion, 32,* 666-673. doi:10.1080/02699931.2017.1319340

Murphy, G. H., & O'Callaghan, A. L. I. (2004). Capacity of adults with intellectual disabilities to consent to sexual relationships. *Psychological Medicine, 34,* 1347-1357. doi:10.1017/S0033291704001941

Okanda, M., & Itakura, S. (2010). When do children exhibit a "yes" bias? *Child Development, 81,* 568-580. doi:10.1111/j.1467-8624.2009.01416

Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse & Neglect, 24,* 733-752. doi:10.1016/S0145-2134(00)00137-X

Paulo, R. M., Albuquerque, P. B., Saraiva, M., & Bull, R. (2015). The enhanced cognitive interview: Testing appropriateness perception, memory capacity and error estimate relation with report quality. *Applied Cognitive Psychology, 29,* 536-543. doi:10.1002/acp.3132

Petersen, J. (2018). Gender difference in verbal performance: A meta-analysis of United States state performance assessments. *Educational Psychology Review, 30,* 1269-1281. doi:10.1007/s10648-018-9450-x

Peterson, C., & Biggs, M. (1997). Interviewing children about trauma: Problems with "specific" questions. *Journal of Traumatic Stress, 10,* 279-290. doi:10.1023/A:1024882213462

Peterson, C., & Grant, M. (2001). Forced-choice: Are forensic interviewers asking the right questions? *Canadian Journal of Behavioural Science*, *33*, 118-127. doi:10.1037/h0087134

Peterson, C., & Whalen, N. (2001). Five years later: Children's memory for medical emergencies. *Applied Cognitive Psychology, 15*, S7-S24. doi:10.1002/acp.832

Pipe, M.-E., & Wilson, J. C. (1994). Cues and secrets: Influences on children's event reports. *Developmental Psychology, 30*, 515-525. doi:10.1037/0012-1649.30.4.515

Poole, D. A., & White, L. T. (1993). Two years later: Effect of question repetition and retention interval on the eyewitness testimony of children and adults. *Developmental Psychology*, *29*, 844-853. doi:10.1037/0012-1649.29.5.844

Quas, J. A., Goodman, G. S., Bidrose, S., Pipe, M. E., Craw, S., & Ablin, D. S. (1999). Emotion and memory: Children's long-term remembering, forgetting, and suggestibility. *Journal of Experimental Child Psychology*, *72*, 235-270. doi:10.1006/jecp.1999.2491

Quas, J. A., Goodman, G. S., Ghetti, S., & Redlich, A. D. (2000). Questioning the child witness: What can we conclude from the research thus far? *Trauma, Violence, & Abuse*, *1,* 223-249. doi:10.1177/1524838000001003002

Quas, J. A., Malloy, L. C., Melinder, A., Goodman, G. S., D'Mello, M., & Schaaf, J. (2007). Developmental differences in the effects of repeated interviews and interviewer bias on young children's event memory and false reports. *Developmental Psychology*, *43*, 823-837. doi:10.1037/0012-1649.43.4.823

Rice, H. J., & Rubin, D. C. (2009). I can see it both ways: First-and third-person visual perspectives at retrieval. *Consciousness and Cognition, 18,* 877-890. doi:10.1016/j.concog.2009.07.004

Robinson, J., & McGuire, J. (2006). Suggestibility and children with mild learning disabilities: The use of the Cognitive Interview. *Psychology, Crime & Law, 12,* 537-556. doi:10.1080/10683160500337550

Romeo, T., Otgaar, H., & Landstrom, S. (2018). Coping with Sexual Abuse. *Psychological Research on Urban Society, 1,* 46-54. doi:10.7454/proust.v1i1.11

Saykaly, C., Crossman, A., Morris, M., & Talwar, V. (2016). Question type and its effect on children's maintenance and accuracy during courtroom testimony. *Journal of Forensic Practice, 18,* 104-117. doi:10.1108/JFP-01-2015-0010

Saywitz, K. J., Geiselman, R. E., & Bornstein, G. K. (1992). Effects of cognitive interviewing and practice on children's recall performance. *Journal of Applied Psychology, 77,* 744-756. doi:10.1037/0021-9010.77.5.744

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. doi:10.1037/0022-006X.64.6.1347

Schreiber, N., Bellah, L. D., Martinez, Y., McLaurin, K. A., Strok, R., Garven, S., & Wood, J. M. (2006). Suggestive interviewing in the McMartin Preschool and Kelly Michaels daycare abuse cases: A case study. *Social Influence, 1,* 16-47. doi:10.1080/15534510500361739

Shannon, P., & Agorastou, M. (2006). Identifying children with developmental disabilities receiving child protection services: A national survey of child welfare administrators.

Families in Society: The Journal of Contemporary Social Services, *87*, 351–357. doi:10.1606/1044-3894.3539

Smith, D. W., Letourneau, E. J., Saunders, B. E., Kilpatrick, D. G., Resnick, H. S., & Best, C. L. (2000). Delay in disclosure of childhood rape: Results from a national survey. *Child Abuse & Neglect*, *24*, 273-287. doi:10.1016/S0145-2134(99)00130-1

Snow, P., & Powell, M. (2004). Interviewing juvenile offenders: The importance of oral language competence. *Current Issues Criminal Justice*, *16*, 220–225. doi:10.1080/10345329.2004.12036317

Suckle-Nelson, J. A., Colwell, K., Hiscock-Anisman, C., Florence, S., Youschak, K. E., & Duarte, A. (2010). Assessment criteria indicative of deception (ACID): Replication and gender differences. *The Open Criminology Journal*, *3*, 23-30. doi:10.2174/1874917801003010023

Sullivan, R. M., Wilson, D. A., Ravel, N., & Mouly, A. M. (2015). Olfactory memory networks: From emotional learning to social behaviours. *Frontiers in Behavioural Neuroscience*, *9*, 1-4. doi:10.3389/fnbeh.2015.00036

Tager-Flusberg, H. (1991). Semantic processing in the free recall of autistic children: Further evidence for a cognitive deficit. *British Journal of Developmental Psychology*, *9*, 417-430. doi:10.1111/j.2044-835X.1991.tb00886.x

Talwar, V., & Crossman, A. (2011). From little white lies to filthy liars: The evolution of honesty and deception in young children. *Advances in Child Development and Behaviour*, *40*, 139-179. doi:10.1016/B978-0-12-386491-8.00004-9

Talwar, V., & Crossman, A. M. (2012). Children's lies and their detection: Implications for child witness testimony. *Developmental Review, 32,* 337-359. doi:10.1016/j.dr.2012.06.004

Talwar, V., Gordon, H. M., & Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology, 43,* 804-810. doi:10.1037/0012-1649.43.3.804

Talwar, V., & Lee, K. (2002a). Development of lying to conceal a transgression: Children's control of expressive behaviour during verbal deception. *International Journal of Behavioural Development, 26,* 436–444. doi:10.1080/01650250143000373

Talwar, V., & Lee, K. (2002b). Emergence of white lie-telling in children between 3 and 7 years of age. *Merrill-Palmer Quarterly, 48,* 160–181. doi:10.1353/mpq.2002.0009

Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behaviour. *Child Development, 79,* 866-881. doi:10.1111/j.1467-8624.2008.01164.x

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). Children's lie-telling to conceal a parent's transgression: Legal implications. *Law & Human Behaviour, 21,* 405-426. doi:10.1023/B:LAHU.0000039333.51399.f6

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

Taylor, J., Stalker, K., & Stewart, A. (2016). Disabled children and the child protection system: A cause for concern. *Child Abuse Review, 25,* 60-73. doi:10.1002/car.2386

Thierry, K. L., & Spence, M. J. (2004). A real-life event enhances the accuracy of preschoolers' recall. *Applied Cognitive Psychology, 18,* 297-309. doi:10.1002/acp.965

Tulving, E. (1974). Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; It may only be inaccessible. *American Scientist, 62*, 74-82. Retrieved from http://www.jstor.org/stable/27844717

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373. doi:10.1037/h0020071

U.S. Department of Health & Human Services. (2016). Child Maltreatment 2016. Retrieved from https://www.acf.hhs.gov/sites/default/files/cb/cm2016.pdf

Vicari, S. (2004). Memory development and intellectual disabilities. *Acta Paediatrica Supplement, 93*, 60-64. doi:10.1111/j.1651-2227.2004.tb03059.x

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*, 1174-1204. http://dx.doi.org/10.1037/a0036620

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and its implications for professional practice.* Chichester, England: Wiley.

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

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology, 22*, 1-21. doi:10.1111/lcrp.12088

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science, 20*, 28-32. doi:10.1177/0963721410391245

Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological science in the public interest, 11*(3), 89-121. doi: 10.1177/1529100610390861

Vrij, A., Leal, S., Mann, S., & Fisher, R. (2012). Imposing cognitive load to elicit cues to deceit: Inducing the reverse-order technique naturally. *Psychology, Crime & Law, 18* 579-594. doi:10.1080/1068316X.2010.515987

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse-order. *Law and Human Behaviour, 32*, 253-265. doi:10.1007/s10979-007-9103-y

Walsh, W. A., Jones, L. M., Cross, T. P., & Lippert, T. (2008). Prosecuting child sexual abuse: The importance of evidence type. *Crime & Delinquency, 56*, 436-454. doi:10.1177/ 0011128708320484

Wechsler, D., & Zhou, X. (2011). *Wechsler Abbreviated Scale of Intelligence, second edition* (WASI-II) *manual*. San Antonio, TX: Pearson Clinical Assessment.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs. Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128. doi:10.1080/016502598384441

Wissink, I. B., Van Vugt, E., Moonen, X., Stams, G. J. J., & Hendriks, J. (2015). Sexual abuse involving children with an intellectual disability (ID): A narrative review. *Research in Developmental Disabilities, 36*, 20-35. doi:10.1016/j.ridd.2014.09.007

Wyman, J., Foster, I., Crossman, A., Colwell, K., & Talwar V. (2018). The efficacy of free-recall, cognitive load, and closed-ended questions when children are asked to falsely

testify about a crime. *Applied Cognitive Psychology*, Advance online publication. doi:10.1002/acp.3494

Wyman, J., Foster, I., Lavoie, J., Tong, D., & Talwar, V. (2019). Detecting children's false allegations and recantations of a crime. *Psychology, Crime & Law*, *24*, 652-671. doi:10.1080/1068316X.2017.1402018

Wyman, J., Lavoie, J., & Talwar, V. (2018). Best practices for interviewing children with intellectual disabilities in maltreatment cases. *Exceptionality*, 1-18. doi:10.1080/09362835.2018.1425623