

PREDICTING ACCURACY: A MODEL FOR ASSESSING CHILDREN'S TESTIMONIAL
COMPETENCE

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Jessica Zoe Klemfuss

August 2011

© 2011 Jessica Zoe Klemfuss

PREDICTING ACCURACY: A MODEL FOR ASSESSING CHILDREN'S TESTIMONIAL
COMPETENCE

Jessica Zoe Klemfuss, Ph. D.

Cornell University 2011

Three million children were subjects of at least one abuse or neglect report in the United States in 2009. When legal cases result from these reports, child testimony is usually the only source of prosecuting evidence. If there is a question about a child's ability to provide legal testimony, his or her testimonial competence may be assessed. When the presiding judge deems a child incompetent, the child is not allowed to testify, or if the child's status is unclear, the judge may provide a warning to the jury about giving less weight to the testimony. However, there are only skeletal legal guidelines in place to aid judges in these decisions and there is little empirical research in this area. The present study was designed to assess new techniques for determining children's testimonial competence. Sixty-four 3 to 5-year-old children completed sections of the Wechsler Preschool and Primary Scale of Intelligence- Third Edition (WPPSI-III), the Test of Language Development- Primary- Fourth Edition (TOLD-P-4), the Child Memory Scale (CMS), and the Video Suggestibility Scale for Children (VSSC) and answered questions designed to approximate the types of questions typically asked in competency hearings. Children also participated in a series of staged events with a confederate and were interviewed about the staged events immediately and after a delay of several days. Children's performance on the WPPSI-III, CMS, and VSSC predicted the ratio of correct to incorrect details children provided about the staged events at the delayed interview. Analyses comparing children who had accuracy ratios above 1 to those with accuracy ratios of 1 or below showed that children who gave more correct information than incorrect information scored higher on every language and memory variable and were less likely to yield to suggested items on the VSSC. Implications for the legal system are discussed.

BIOGRAPHICAL SKETCH

Jessica Zoe Klemfuss earned her bachelor's degree in Psychology in 2004 from the University of California, Berkeley. After completing her undergraduate education she worked as a research assistant for Dr. Liat Ayalon, studying the cognitive effects of untreated sleep apnea at the University of California, San Diego and the Veterans Administration Hospital in San Diego, California. She entered the graduate program in the Department of Human Development at Cornell University a year later and earned her master's degree in Developmental Psychology in January of 2009. In her time at Cornell Zoe worked primarily in the labs of Drs. Stephen Ceci and Qi Wang. Affiliation with these labs contributed to both the content of her research and her methodological approaches.

Zoe's primary area of study involves factors influencing children's testimonial accuracy. The factors she has examined thus far include the format and content of interviewer questions, children's narrative skill, mother-child conversation style, and components of testimonial competency. She is also currently involved in projects examining adult perceptions of child accuracy, the influence of emotion and self-relevance on adult memory accuracy, and the development of crypto-plagiarism in young children. After completion of her doctorate, Zoe will start a position as a post-doctoral fellow at the University of California, Irvine and the University of Southern California with Drs. Jodi Quas and Thomas Lyon.

To my family,
and in memory of Sarah Kulkofsky- my informal mentor and close friend.

ACKNOWLEDGMENTS

This work was funded by a dissertation grant from the Department of Human Development at Cornell University and by a contribution from Dr. Stephen Ceci.

I would like to thank my committee members Drs. Steve Ceci, Qi Wang, and Charles Brainerd. Their guidance and encouragement through the years have shaped me as an academic and made me feel like a rising colleague. And though she was not formally a member of my committee, I would be remiss if I did not also thank Dr. Chrissie Schelhas-Miller, who trained me to be a teacher and supported me as a friend through the ups and downs of graduate school.

I consider myself fortunate to have had the exceptional advising and resources provided by my committee and the department, but I would not have gotten to this point without the mentoring and friendship of Sarah Kukofsky. Sarah took me under her wing even before I decided where I would go for graduate school, and she made my decision to come to Cornell an easy one. Sarah became my housemate during her final two years as a graduate student and remained my informal mentor, collaborator, and close friend through graduate school and after she took a faculty position at Texas Tech University. She helped me get through tough times, helped me celebrate good times, and taught me everything from minute functions of Excel to complex theory and research methods. I am lucky to have known her and I continue to model myself after her.

I am also grateful to the undergraduate research assistants who helped with my doctoral research: Alexandra Beauchamp, Hilary Oran, Kristen Spitaletta, Maclen Stanley, Michelle Elsner, Nathalie Dupart, Patricio Martinez, and Zoe Luscher. I was consistently impressed with their abilities and their dedication to the project. I could not have done it without them. I am also appreciative of Eva Dixon and Maanami Ransom, who assisted with initial piloting of the study, and of course, to the children, families, and school staff who volunteered time and resources to make this research possible.

I would like to thank my friends who kept me happy and motivated throughout my graduate career. I was lucky to find wonderful people in my graduate cohort, most notably, Sarah

Hertzog and Andy Reed, who listened to my research ideas, screened my talks and presentations, edited my written work, and were there to lean on when life got tough. I am also grateful for the love and support of the many other exceptional friends I have gotten to know during my time in Ithaca. Thank you to Cagla Aydin for being the best lab mate anyone could ask for, to the “Archit crowd” for being my Friday night outlet, to Ian Schachner, Tim Brown, and Kyle Emich for sharing my love of Ithaca, good food, and fun events, and to so many others who have made the past six years some of the best of my life.

Finally, thank you to my family for listening to me, loving me, and believing in me. One of the perks of being at Cornell was getting to be near my extended family and getting to feel the warmth and support of so many family members at once. One of the most exciting parts of moving back to the west coast is that I will be near my nuclear family again. Mom, Dad, and Nola, thank you for making me feel like you are always there for me no matter how far away I am.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgments	v
List of Figures	viii
List of Tables	ix
Chapter 1 – Introduction	
<i>History of Children’s Legal Testimony</i>	1
<i>Previous Research on Testimonial Competency</i>	4
<i>Understanding and Answering Simple Interview Questions</i>	5
<i>Observing and Recalling Important Events and Information</i>	7
<i>Knowledge about Truth and Lies</i>	15
<i>Summary</i>	19
Chapter 2 – Methods	
<i>Participants</i>	20
<i>Materials</i>	20
<i>Procedure</i>	22
<i>Coding</i>	24
Chapter 3 – Results	
<i>Preliminary Analyses</i>	27
<i>Language and Accuracy Ratio</i>	27
<i>Memory and Accuracy Ratio</i>	28
<i>Full Model</i>	29
<i>Performance Consistency</i>	30
<i>Language and Consistency</i>	32
<i>Memory and Consistency</i>	33
<i>Competency Cutoffs</i>	34
Chapter 4 – Discussion	39
Appendix A – Forensic Memory Interview	48
Appendix B – Staged Event	49
Appendix C – Staged Event Interview	50
References	51

LIST OF FIGURES

Number of Children by Language Percentile Group and Report Accuracy Group38

LIST OF TABLES

1. Visit Descriptions.....	24
2. Full Model Predicting Open-Ended Accuracy Ratio	31
3. Full Model Predicting Closed-Ended Accuracy Ratio.....	31
4. Age and Language and Memory Variables by Open-Ended Low- vs. High Accuracy	35
5. Age and Language and Memory Variables by Closed-Ended Low- vs. High Accuracy	35

CHAPTER 1

INTRODUCTION

History of Children's Legal Testimony

In the 1980's and 1990's there was a dramatic increase in the prosecution of child abuse cases in the United States. This was partially due to an increased awareness of the seriousness and prevalence of crimes against children (Hall & Sales, 2008), and partially due to increasing confidence in young children's ability to provide testimony (Child Victims and Child Witnesses Rights Act, 1990). While there is some evidence that abuse reports are declining (Finkelhor & Jones, 2006), 3 million children, or 40.3 out of every 1,000 children, were subjects of abuse or neglect reports in 2009 and many of these children were subjects of multiple reports. In one fifth of these reports there was evidence of maltreatment (U.S. Department of Health and Human Services, 2009). Because maltreatment cases often rely on child testimony as the primary source of evidence, the legal system should have a vested interest in determining which children's testimony should be admitted and believed (Ceci & Bruck, 1995).

In the literature on child development, children's believability is usually decided by measuring the objective accuracy of a child's report. In the legal context, there is often no empirical evidence with which to directly assess accuracy. Therefore, inferences about a child's accuracy are made at two levels. At the first level the trial judge must decide whether a child's testimony should be admissible. To do this, he or she determines whether the testimony is *reliable* and, in some cases, whether the witness is *competent*. Legally, reliability is defined broadly and refers to all potential evidence, including testimony (Rosenthal, 2002). To consider evidence reliable, there must be reason to believe that it is what it purports to be. In the case of testimony the judge must decide that the witness had the opportunity to witness what he or she

purports to have witnessed and that the witness's testimony was not tampered with through coaching or suggestion. Testimonial competence refers to whether or not a witness has the cognitive ability to provide useful testimony. Testimonial competency guidelines are set in both federal and state law and competency decisions are made by the presiding judge (Myers, 1997).

If testimony is determined reliable and the child providing the testimony is either assumed to be competent or deemed to be competent by the judge, it is up to the court to determine the child's *credibility*. Credibility is the court's opinion about whether or not a child's testimony is believable. Therefore, reliability, competency, and credibility are all components of determining accuracy in a context in which there is question about how the original event unfolded. The present work focuses on the determination of a child's testimonial competency. This is a complex legal issue that has received little empirical attention.

In the landmark 1779 decision of *The King v. Brasier*, the defendant was accused of assault with intent to rape a young girl, Marry Harris, who was under seven years of age at the time of the alleged attack. This early ruling gained widespread acceptance for its determination that whether a child may testify at trial should not be based on a minimum age but rather on the child's ability to take the oath, which in turn depends on an assessment of the child's understanding of the importance of telling the truth.

The *King v. Brasier* dictum was later elaborated in the 1895 Supreme Court case of *Wheeler v. U.S.* In this case a 5 year old boy was the only witness to a murder, and the case could not be decided without his testimony. The judge ruled that the boy's testimony was admissible because he was both *sufficiently intelligent* to serve as a witness and because he demonstrated the *ability to distinguish between truths and lies*, and the *understanding that he was morally obligated to tell the truth*. These standards were further elaborated so that in many

states the competency standard now requires that a witness be able to 1) understand and answer simple questions, 2) observe and recall events pertinent to a case, 3) understand the difference between the truth and a lie, and 4) understand that they are morally obligated to tell the truth on the stand. Though the first of these standards, the ability to understand and answer simple interview questions, is the only explicit federal competency standard (Child Victims and Child Witnesses Rights Act, 1990).

Historically, courts have moved between extremes of allowing child testimony. Previously, children were not allowed to testify as witnesses because they were not believed competent as a group. However, today children as young as three years of age are presumed competent in most courts unless a particular child's competence is brought into question in a case (Child Victims and Child Witnesses Rights Act, 1990; Myers, 1992). According to Rule 601 of the Federal Rules of Evidence, age alone may not be considered a compelling reason to question a person's competency. If a child's competence is in question the trial judge can probe that child's competence in the presence of the jury. If a child seems particularly incompetent this could result in a competency hearing with only the judge, lawyers, and a supportive adult for the welfare of the child present. The judge can also request counsel from a psychologist or social worker, but the final decision is made by the judge (Gershman, 2001; Myers, 1997).

While competency hearings have become less common, trial judges make other legal decisions which previously would have fallen into the category of competency determination (Myers, 1986-1987). For example, a trial judge can exclude a child's testimony because it is not relevant or because the child lacks sufficient "personal knowledge" of the case. The judge can also decide to reserve a child's testimony and use it only if it is necessary to the case. These alternate decisions are legal even given Rule 601, but they can be used to reach the same end of

excluding a child's testimony because of perceived incompetence. In short, competency decisions are more ubiquitous than they seem at first glance because legal loopholes allow judges to make competency decisions without official competency hearings.

No matter which legal mechanism is used to exclude a child's testimony, the final decision is always made by a judge, who often has no special training involving children or developmental research (Cashmore & Bussey, 1996; Myers, 1992). To compound this issue, there are no empirically validated tests of testimonial competence currently used in courts (Cashmore & Bussey, 1996; Seton Hall L. Review, 2010). In fact, as will be discussed in detail below, there are few standards for the administration of competency exams and consequently, there are few studies examining the predictive value of these exams. What little research there is suggests that current exams are not predictive of the quality of children's testimony.

Previous Research on Testimonial Competence

There have been some efforts to develop reliable tests of children's competence. In an unpublished pilot, Hansen (1990) designed a functional test of two components of competency 1) *understanding of the obligation of an oath* and 2) *sufficient intelligence and capacity*. The test was piloted on a non-representative sample of 17 children including 1st-4th graders, and 7th graders. The test had high inter-rater reliability and high internal consistency on all measures. However, attempts at external validity were not successful, and the pilot did not include measures of predictive validity.

Lyon and Saywitz (1999, 2000) have developed a theoretically reliable test of children's knowledge of truths and lies and understanding of the distinction between truths and lies. For each task children are presented with four images with accompanying stories. For example, to test children's knowledge of the distinction between the truth and a lie, children see an image of

two children who are looking at a truck. Speech bubbles above the children's heads indicate that one child says that the truck is a truck and the other child says it is an airplane. Participants are then asked which child is telling the truth. This is repeated four times, each with different scenarios.

The process for testing understanding of the consequences of lying is quite similar. Participants are shown images in which two children are standing in front of an adult authority figure. For example, in one scenario, they both stand in front of a judge. Children are told that one of the children told the truth to the judge and the other child told a lie to the judge and are then asked which child would likely get into trouble. Children's passing scores require that they answer questions about all four scenarios correctly. These tests are useful because they reduce children's reliance on language and passing scores on the tests mean that children scored significantly above chance because they respond to four different scenarios rather than a single yes-no question.

Understanding and Answering Simple Interview Questions

When a child's testimonial competence is called into question, their ability to understand and answer simple interview questions is generally assessed through a short interview with the presiding judge with or without council or jury present (Myers, 1997). The problem, however, is that as basic as this competency may seem, the ability to effectively communicate in a legal setting is extremely complex, and can be challenging even for competent children (Saywitz & Goodman, 1996; Saywitz, Snyder, & Nathanson, 1999; Walker, 1993).

Walker (1993) has analyzed court transcripts with an eye to the linguistic complexities involved in a standard case involving child witnesses. Her findings illustrate that the language used in court is "informal, illogical, ungrammatical... full of blunders and grievous errors and

mutations... and characterized by endless sentences, false starts... and other crudities” (Walker, 1985, p. 115). Walker provides a detailed case analysis in which she examines the full court transcripts from interviews with a 5-year-old girl. She gives concrete examples of instances of semantic, syntactic, and pragmatic complexity that are far beyond the developmental level of a preschool-aged child. In a particularly dramatic example of syntactic complexity, the girl was asked “Do you also recall driving in a car a day or two after Doug-you found out that Doug-that something had happened to him and telling and pointing out houses as being the place where the people or one of the people who hurt Doug lived” (p. 68). As phrased, many adults would likely have difficulty unpacking the intent and providing a meaningful response, and research has demonstrated that young children have more trouble understanding the language of the courtroom than do adults (Saywitz, Goodman, & Lyon, 2002).

When the language of the interview is developmentally inappropriate or too complex for a particular child, children are more likely to assent to an interviewer’s question. Fritzley and Lee (2003) asked preschool-aged children questions that were answerable, and questions that were nonsensical to see whether they would respond in predictable ways. Overall, children rarely answered “I don’t know” to the nonsensical questions, and the youngest children exhibited a tendency to acquiesce to the experimenter. This finding suggests that perhaps limitations in language skills and vocabulary constrain children’s abilities to answer questions that contain advanced language and may lead children to falsely assent to an interviewer. In keeping with this, children’s overall language abilities are often correlated with decreased suggestibility (see Bruck & Melnyk, 2004 for a review).

The developmentally inappropriate language used in court makes it difficult for any child to provide reliable testimony, no matter what their relative competence at understanding and

answering questions may be. However, competency examinations may still be helpful in determining which children can provide quality testimony, and to what degree, in instances where they are asked developmentally appropriate questions. In addition, competency determinations could potentially guide well-trained forensic interviewers to ask questions at appropriate levels of complexity in interviews that occur outside of court. Children with low language scores may require interviewers to be particularly mindful of the vocabulary and structure used in interview questions. These interviews can be critical pieces of evidence in cases involving child witnesses and in some cases out of court testimony may be the only evidence from a child witness (see Hall & Sales, 2008). A reliable measure of competency may be particularly important because as will be discussed below, often the children who seem competent and sound credible are not accurate.

Validly and reliably determining individual differences in children's abilities to understand and answer simple interview questions is a daunting task. Children are often interviewed by professionals with experience interviewing children at some point in the pre-trial process, and these professionals are advised in the literature to make note of children's communicative abilities (ex: Myers, 1992, p. 52). However, these evaluations are merely recommendations to the judge, and are not necessarily deciding factors in a child's competency determination. The guidelines for judicial competency examinations are also extremely open-ended. Although judges have the responsibility to assess a child's ability to understand and answer questions, they are not given specific questions or tests to administer. Cashmore and Bussey (1996) found that there was a wide range of techniques used by judges to determine competency. While this study was conducted with Australian judges and magistrates, the competency guidelines are quite similar between Australia and the United States. Because there

is no standard test of children's communicative competency to be used in legal settings, we are limited in determining the validity of these tests, and communicative competency determinations are not likely to be reliable. Courts are in need of a practical, valid, and reliable test of communicative competency.

Observing and Recalling Important Events and Information

Research on children's general memory skills has demonstrated that in some ways, young children's memory is quite impressive (e.g. Fivush, Hudson, & Nelson, 1984; Saywitz, Goodman, Nicholas, & Moan, 1991). The memory system is online from even before birth (DeCasper & Spence, 1986), and infants can retain information even over long-periods (Shields & Rovee-Collier, 1992). By the early preschool years children are even able to organize their memories of personally-experienced events into cohesive narrative reports (Fivush, Haden, & Adam, 1995; Newcombe & Reese, 2004). However, buried within these generalities are pronounced individual differences; the memory process develops slowly and does not reach adult levels until adolescence for most children, though earlier for some and later for others (ex: Greenhoot, Ornstein, Gordon, & Baker-Ward, 1999).

One limitation early in memory development is that young children's memory reports contain fewer details on average than those of adults (e.g., Leippe, Romanczyk, Manion, 1991). These differences come about because of early limitations in processing speed as well as limitations in children's knowledge. Limitations in processing speed can hamper memory at both encoding and retrieval. At encoding, processing speed affects how much information gets into the system because the longer it takes to encode, the less likely it is that the information will make it into storage (Hale, 1990; Kail, 2007; Kail & Ferrer, 2007). Limitations in processing speed can also result in limitations in the volume of information retrieved. Because children

spend more cognitive resources trying to retrieve information, they tend to retrieve less than adults (Kail, 1986).

Limitations in children's knowledge base have also been demonstrated to hamper the amount of *relevant* information they are able to successfully encode, store and retrieve. Children have a different understanding than adults of which information is important and relevant. For example, while an adult is likely somewhat familiar with which pieces of information would be critical to provide when they are a witness in a criminal case, children are less likely to have legal knowledge (Saywitz & Goodman, 1996; Saywitz & Snyder, 1993). These differences in perception shape what children attend to when they witness events, which consequently shapes what they are able to retrieve. Perception of which information is important later shapes which information children decide to retrieve.

In a similar vein, children's knowledge of scripts--general knowledge of reoccurring events--plays a role in how they perceive and remember events. Familiarity with the elements of an event, or having a sense of the typical sequence of a witnessed event, contributes significantly to the interpretation and consequent memory for an event. Pillemer, Picariello, and Pruett (1994) interviewed children who had just experienced a fire drill at their school. Younger children, who were less familiar with the sequence of events common to fire drills, were more likely to make mistakes about the sequencing of events. For example, they were more likely to claim that they evacuated their classrooms before they heard the alarm. Older children did not make these types of mistakes. Thus, with experience, the number of spontaneous inaccurate statements decreased and the number of correct statements increased.

Another type of knowledge that can influence memory is metamemory knowledge. As children develop knowledge about how memory works and which strategies they can use to help

them remember, they remember more. The first strategy to develop is rehearsal, which involves mentally or verbally repeating to-be-remembered information to retain it in memory. Next is organization which entails rearranging information in a meaningful way to aid in retention. The final strategy to develop is elaboration in which information and structure is added to make a cohesive story out of the to-be-remembered information. These strategies do not develop one after another in a consistent trajectory, but overlap while children learn the memory benefits of each strategy and learn to use them efficiently (Siegler, 1996). As children begin using increasingly effective strategies there are resulting increases in memory abilities (Beuhring & Kee, 1987; Shlagmuller & Schneider, 2002). Complicating matters still further, research shows that the components of metamemory are affected by the child's representational knowledge: earlier competence is seen in domains in which the child has elaborate knowledge (Ceci, Fitneva, & Williams, 2010).

Children not only become better at remembering more accurate information over time, but they also remember a higher ratio of accurate to inaccurate information over time. Beuscher and Roebbers (2005) tested 6, 8, and 10 year old children's memories about a brief video one week after its presentation. They found that the number of correct responses about the video increased with age and that the 6 year olds included a higher proportion of false details into their reports than either of the older age groups. Further, older children were more likely to respond "I don't know" when they were asked leading questions for which "I don't know" was the most appropriate answer. The authors also tested whether children could effectively utilize social information to monitor their reports for accurate information. Half the children were told that the experimenter was not knowledgeable about the video prior to the interview. If children understood that the experimenter had no special knowledge, and could successfully use this

information to monitor their memories, then they should have been less influenced by any unanswerable questions the experimenter asked. However, the information made no difference in the accuracy of children's reports. The authors suggest that this null finding implies that young children have difficulties in monitoring their memories based on a social "warning".

Finally, susceptibility to suggestion can play a major part in children's memory accuracy. Suggestibility refers to a wide range of factors that may influence a person's memory report. Ceci and Bruck (1995) provide a comprehensive definition of suggestibility that will be the basis of the following discussion, "...suggestibility refers to the degree to which the encoding, storage, retrieval, and reporting of events can be influenced by internal and external factors" (p. 44). This brief definition encompasses critical distinctions that are missing in classic definitions of suggestibility. First, suggestibility may occur at any stage of the memory process, including at the point of reporting the event. Second, suggestibility may occur due to either internal or external factors. Implicit within this definition is the understanding that unlike previous conceptualizations of suggestibility, the current understanding is that a child's memory report may be inaccurate not only because the memory has been permanently altered (the cognitive or internal explanation), but it could also be because the child is making a superficial attempt to please the interviewer or some other interested adult (the social or external explanation).

The cognitive and social forms of suggestion can also work in concert so that while a child may initially provide a false report for social reasons, the act of rehearsing the false report could damage the memory trace, resulting in permanent alterations of the internal memory. For example, a suggestive interviewing technique that is frequently used in empirical work involves asking children to "pretend" that an event occurred and to describe what it might be like. Often children begin to incorporate these initial intentionally imagined details into later memory

reports (e.g. Ceci, Crotteau, Smith, & Loftus, 1994; Kulkofsky & Klemfuss, 2008). A similar mechanism may occur for children who provide false information for personal motivations. They may “turn a lie into memory’s truth” (Loftus, 1992).

Inaccurate reports can be particularly damaging in a legal setting. They may result in false allegations, these allegations can result in an unnecessary trial, and sometimes trials based on false claims can result in false convictions. Conversely, false reports may preclude trials when they should, in fact, be held. A consistent finding in the literature on children’s testimony is that in most cases children are more vulnerable to interviewer biases than are adults (see Ceci & Bruck, 1995 for a review and Brainerd, Reyna, & Ceci, 2008 for important exceptions).

Age is the most robust individual difference in the suggestibility literature. This has been a consistent finding since early in the 20th century. William Stern (1910, cited in Ceci and Bruck, 1993) showed images to children between the ages of 7 and 18 and asked them to memorize them. He then interviewed the participants with both open-ended questions and closed-ended, suggestive questions. He found that every age group provided more accurate responses to the open-ended, non-suggestive questions, but further, that there was a developmental trend so that the youngest children he interviewed were the most susceptible to suggestive questions about the pictures. This finding has been replicated more recently in varied contexts, including forensically relevant ones, and using varied types of suggestion. For example, Eisen, Qin, Goodman, and Davis (2002) found that preschool-aged children who were interviewed about an anogenital examination that they received as part of an ongoing abuse investigation were more susceptible to misleading questions than were their older counterparts (6-10yr olds and 11-14yr olds). This study is forensically relevant because the participants were children who were suspected of

having been abused and because they were interviewed about an event that involved intimate physical contact and stress.

As discussed above, much of the early work on suggestibility, and many of the early definitions of suggestibility focused on contexts in which an adult manipulated a child's memory report through biased techniques. The manipulation of the child's report in these situations is largely subconscious for the child and this process is generally the one associated with cognitive alteration of the child's memory. A cognitive change can permanently alter the child's memory trace so that she no longer has access to the untainted memory of the original event. However, children's reports can also be altered due to social motivations that result in an initial conscious decision to manipulate the memory report.

In summary, the accuracy of a child's report depends on several external and internal factors. This review has focused on developmental differences in children's memory reports based on other cognitive developments such as processing speed and knowledge and contextual factors such as social influence and motivation to lie. Developmental differences in each of these factors often makes young children more vulnerable to forgetting and distortion, but identifying the causes of developmental differences can also inform us about potential individual differences in remembering and forgetting beyond age.

As discussed above, the most consistent individual difference in memory ability and suggestibility is age. On average, older children and adults remember more accurate information and are less suggestible than are young children. However, competency decisions are not, and should not, be made solely on the basis of age because there are wide individual differences in accuracy within age groups and because in some contexts, young children are more accurate than young children and adults.

There is limited research which explores individual differences, other than age, in children's memory skills (Bauer, 2004). While it is clear that children vary considerably in their abilities to remember and to accurately report those memories, our ability to predict those differences has been limited. Research on individual differences in event memory in the preschool years has explored differences in gender (ex: Reese & Fivush, 1993), information processing (Schneider & Bjorklund, 1998), background knowledge (Goodman, Quas, Batterman-Faunce, Riddlesberger, & Kuhn, 1994; Ornstein, Merritt, Baker-ward, Furtado, Gordon, & Principe, 1998), socialization (Nelson & Fivush, 2004), narrative skill (Kulkofsky, Wang, & Ceci, 2007, Kulkofsky & Klemfuss, 2008, Nelson & Fivush, 2004), and language skills (Greenhoot, Ornstein, Gordon, & Baker-Ward, 1999; Gordon, Ornstein, Nida, Follmer, Crenshaw, & Albert, 1993; Reese & Brown, 2000; Walkenfeld, 2000). However, the results have been mixed and the measures used to assess individual differences vary widely. Further, in a majority of studies the measures are associated with the number of details children report or recognize rather than the accuracy of the report.

A majority of the limited research on individual differences and memory has focused on forensic issues like suggestibility (Bauer, 1996). For example, Leichtman and Ceci (1995) biased a group of children with stereotypes and interviewed the children repeatedly and suggestively to examine the effects on the children's memory reports. While some children eventually incorporated all four suggested, stereotype-consistent items into their memory reports, some children resisted all attempts at suggestion. Therefore, while some children are prone to incorporating suggestive details, not all children are suggestible. There are also reliable tests of individual differences in suggestibility which indicates that susceptibility to suggestion is a quantifiable individual difference (Gudjonsson, 1984; 1987; Scullin & Ceci, 2001). Bruck and

Melynck (2004) published a thorough review of the literature on individual differences and suggestibility and found that the cognitive factors that were most consistently related to suggestibility were language ability and creativity.

Secondly, there is a body of recent evidence demonstrating that in some contexts young children may be *more* accurate than their older counterparts. For example, children's limited knowledge of semantic associations has been shown to limit their suggestibility when the suggestions concern items that are semantically related to true items. Older children and adults are more susceptible to these types of suggestions than are younger children who lack this knowledge (Brainerd & Reyna, 2007; Brainerd et al., 2008; Ceci, Papierno, & Kulkofsky, 2007). Therefore, younger children are not always more suggestible than older children and adults.

In summary, while the memory system continues to develop well past the preschool years, young children can still provide accurate and detailed testimony. However, there is significant variability in preschool-aged children's abilities to provide accurate testimony and to resist suggestion. Because of the breadth of individual differences in children's abilities to recall events, predicting children's competency in this domain could prove critical in order to ensure a fair trial. However, the determination of a child's ability to observe and recall the events relevant to a case currently suffers from the same limitations in reliability and validity as the assessment of a child's ability to understand and answer interview questions. Additionally, the empirical literature on the topic is especially limited. Judges have the freedom to determine competence using their own discretion and are provided with only skeletal guidelines with which to make those assessments. This process necessarily leads to inconsistencies in assessment (Cashmore & Bussey, 1996).

Knowledge About Truth and Lies

In order to provide accurate testimony children must not only meet a threshold of language and memory skills, they must also be motivated to provide accurate testimony- they must be honest. As noted by the Brasier Court in 1779, "... [children's] admissibility depends upon the sense and reason they entertain of the danger and impiety of falsehood, which is to be collected from their answers to questions propounded to them by the Court" (202-203). Research has amply demonstrated that young children can, and will lie to adults for multiple reasons. Children will sometimes lie to avoid punishment, to sustain a game, to keep a promise/protect a loved one, to achieve personal gain, and to avoid embarrassment (see Ceci & Bruck, 1995 for a review). In fact, some research paradigms rely on lying behaviors in preschool-aged children (London & Nunez, 2002; Talwar, Lee, Bala, and Lindsay, 2002). A paradigm used by London and Nunez (2002) and Talwar et al (2002) to examine contextual effects on lying behavior involves instructing children not to peek at a concealed object or toy when the interviewer leaves the room. The room is surveyed by a hidden video camera to observe children's behaviors. In both studies the lying rates for the children who peeked were as high as 90%.

The most recent area of study involving children's competency to give testimony involves children's abilities to distinguish truths from lies and children's understanding of their moral obligation to tell the truth. While the debate about how early children are capable of lying is long-standing (Piaget, 1932), the relationship of their understanding of truth-telling and their truth-telling behavior has only recently been empirically tested. The research on children's truth and lie competency has taken two approaches. The first is designed to assess the predictive power of current competency exams for children's behavior, and the second strives to enhance the validity of the current exams.

It appears that from an early age children have a basic grasp of what it means to tell the truth and an understanding that it is good to tell the truth (Pipe & Wilson, 1994). In fact, young children often have more rigid definitions of morality than do adults (Kohlberg, 1963; Piaget, 1932). Haugaard, Repucci, Laird, and Naful (1991) showed preschool-aged children a video in which a young girl is coached by her mother to lie to the police or a video in which she is not coached. The children were then interviewed to see whether they thought the girl lied when she told the police incorrect information. The authors hypothesized that because of immature knowledge of what it means to lie, and because of children's susceptibility to social pressure by authority figures, children would not consider the false information provided by the mother as a lie. However, they found that even the preschoolers said that the girl was lying if she told the police incorrect information. It did not matter whether the girl in the video was instructed to lie by her mother. The authors also read children a vignette in which one child lied to protect a friend from punishment and they found that nearly every child (91%) agreed that saying something that was untrue was a lie even if it was to protect a friend.

It has been demonstrated that if children are interviewed in an age-appropriate manner they can be competent at answering questions testing their understanding of the truth/lie distinction and their understanding of the importance of telling the truth (Pipe & Wilson, 1994). Interviews that are appropriate for young children reduce the need for conceptual description of terms and rely on concrete demonstrations of truths and lies. For example, preschool-aged children are generally accurate at answering questions such as "If I told you I had purple hair, would that be the truth or a lie". On the other hand, young children have difficulty answering questions that call for abstract conceptualizations, such as "What is the difference between the truth and a lie". The downside to these types of competency questions is that they may underestimate

young children's appreciation of truth/lie distinctions. Children can also be accurate when answering simple, concrete questions about the morality of lying. For example, they are often accurate at answering a question such as "Is it good/bad to tell the truth/a lie" (Myers, 1997). The downside to these latter types of competency exams, however, is that they involve exclusively closed-ended and forced-choice questions, which are often more leading than the open-ended variety (ex: Fritzley & Lee, 2003).

Empirical work has failed to find a relationship with these competency tests and children's lie-telling behavior. While most children are able to pass the truth and lie sections of legal competency exams, their performance on these exams does not predict whether they will lie or tell the truth. It is unclear whether this is an artifact of the methods employed to test children's knowledge, or whether children's behavior simply does not reflect their knowledge in this context.

Talwar et al (2002) interviewed 3-7 year old children using hypothetical scenarios where either the child or an imaginary character had the opportunity to lie or tell the truth about a transgression. Nearly every child demonstrated knowledge of truths and lies, indicated that the imaginary character should not lie to conceal a transgression, and that they would not lie to protect themselves. However, performance on these measures did not predict truth-telling performance. Many of the children lied to conceal their own transgression, despite successfully passing the competency exam.

For half the children Talwar et al (2002) had a conceptual discussion about lies and truths before children were given the opportunity to lie, and for the other half children demonstrated their truth- or lie-telling behavior before they had this conceptual talk with the experimenter. The authors found that the order of presentation did not matter. This implies that truth and lie

components of competency exams do not predict behavior, and further, that having these conversations does not make children more truthful. However, when children promised to tell the truth, their truth-telling behavior increased. Therefore, Talwar et al. (2002) suggest that taking the oath is more likely to increase children's truthfulness than is measuring their understanding of truth vs. lies, or their understanding of the morality of lying. It is also more likely to influence their truth-telling behavior than is having a discussion about truths and lies.

London and Nunez (2002) similarly tested the relationship between children's truth-telling knowledge and performance. They found that children's performance on competency exams did not predict their truth-telling behavior but that children who participated in these discussions were less likely to lie to experimenters about minor transgressions. Thus, neither Talwar et al (2002) nor London and Nunez (2002) found a relationship between truth/lie competency and truth-telling performance. However, there is a discrepancy in their respective findings concerning the benefits of truth/lie discussions. Talwar et al (2002) found that truth/lie discussion had no impact on children's truth-telling behavior and London and Nunez (2002) found that the discussion improved children's memory performance. This is likely because the truth-lie discussions in London and Nunez (2002) included requests for children to tell the truth, which is similar to the beneficial effect of taking the oath in the Talwar et al (2002) study. Therefore, it may be the case that it was the oath alone that led to increased accuracy in the London and Nunez (2002) results.

Summary

The limited data on children's testimonial competence suggests that competency standards are in need of revision and refinement. I know of no research that directly assesses the types of questions that are typically used to measure children's ability to understand and answer

interview questions or those used to estimate children's memory skills. However, above I have reviewed evidence that individual differences exist in these domains. I have also reviewed the recent work on the truth and lie components of the competency exam. This data suggests that establishing a child's understanding of truths and lies and morality about lying is at best measured inappropriately in the current system and at worst an unnecessary and misleading component to the exam. Instead, requiring children to take a child-friendly oath may accomplish more than including these components on a competency exam (McGough, 1994).

The present study is a first step toward developing a reliable and valid exam for use in testimonial competency evaluations. The goals for the study were to assess whether the types of competency exams currently in use can predict children's testimonial accuracy and to test whether an empirically informed test of competence can more robustly predict accuracy. The study focuses on language and memory assessment because these are the competency components that have been previously neglected in the literature. Children were interviewed both with measures that approximate the types of questions commonly used by judges to estimate children's testimonial competency and with a set of empirically validated measures selected to measure the same basic competencies. Individual differences in children's competency scores were then compared with the accuracy of children's testimony about a staged event.

CHAPTER 2

METHODS

Participants

Sixty-four preschool-aged children between the ages of 32 and 68 months ($M = 52.30$, $SD = 7.17$, 48% female) were recruited from two schools in the central New York state area.

Principals were contacted directly and agreed to have consent forms distributed to parents of preschool-aged children within each school. Children were included in the study if their parents agreed that they could participate and that the interviews could be audio recorded.

Materials

Children completed test batteries to assess language ability, memory skills, and suggestibility. One set of tests involved questions designed to approximate the types of items currently asked by judges to assess child competency. The other set of items included a battery of primarily normed exams and tasks used in the empirical literature.

Forensic Battery. Interviewers engaged children in a 1-5 minute open-ended conversation in order to elicit as much narrative from the child as possible. Two raters who were blind to children's performance on other test measures rated each child's receptive and productive language abilities on a three-point Likert scale (*not proficient, proficient, highly proficient*) and specified the reasoning for their ratings. Memory was assessed through a structured interview in which children were asked to discuss four past events- breakfast that morning, all events from the previous day, the child's last birthday, and the previous summer. First children were asked to provide as much information as possible and then the interviewer followed up with closed-ended prompts as needed (see Appendix A).

Pilot Battery. Children's language was also assessed with the Verbal Intelligence Quotient (VIQ) from the Wechsler Preschool and Primary Scale of Intelligence- Third Edition

(WPPSI-III) and the Relational Vocabulary and Syntactic Understanding components of the Test of Language Development-Primary, Fourth Edition (TOLD-P:4). The VIQ subtests for children between the ages of 2 years, 6 months, and 3 years, 11 months, are Receptive Vocabulary and Information. For children 4 years old and above, the subtests are Information, Vocabulary, and Word Reasoning. The younger children also completed the Vocabulary and Word Reasoning subtests so that their raw scores could be compared with those of the older children.

The WPPSI-III was normed with 1700 children between the ages of 2 years, 6 months and 7 years, 3 months. The reliability coefficients for the subtests range from .83 to .95, and the test-retest reliability ranges from .84 to .92. The WPPSI-III is also considered a valid measure based on previous data on the WPPSI-R. Scores on subtest composites range from .74 to .90 when compared to the WPPSI, WISC-R, Stanford Binet (4th ed.), and McCarthy Scales. There is also evidence of discriminant validity with the WPPSI-R. For both age groups the subtests took approximately 20 minutes to administer.

All of the reliability coefficients for the subtests of the TOLD-P:4 exceed .80 as do test-retest scores. Inter-rater reliability coefficients exceed .90. Validity was established through comparison with the Pragmatic Language Observation Scale, TOLD-I:4, and WISC-IV Verbal Composite. The TOLD-P:4 has also been shown to identify people known to have poor language skills. The TOLD-P:4 subtests took approximately 10-15 minutes to administer.

Children's memory skills were assessed with the six core subtests of the Children's Memory Scale (CMS, Cohen, 1997) and an event memory task. The six CMS subtests include Dot Locations, Stories, Faces, Word Pairs, Number, and Sequences. The total administration time for the CMS is 20 minutes. The CMS has been normed with children between the ages of 5

and 16 years. The average reliability coefficient is .91, test-retest reliability is .89, and inter-rater reliability is .94. Validity was assessed by comparison to the Wechsler.

Children's general event memory skills and individual differences in suggestibility were assessed with the Video Suggestibility Scale for Children (VSSC, Scullin & Ceci, 2001; Scullin & Hembrooke, 1998; Scullin, Kanaya, & Ceci, 2002). The VSSC is composed of two time points. At the first interview children watch a brief video of a child's birthday party. Then, after a delay of at least one day children are interviewed about the content of the video. First, children are asked non-leading open-ended questions. Next, children are asked a set of nine leading yes/no questions followed by the following mild negative feedback: "You missed a few of the questions. Let's go through them again and see if you can do better this time". The interviewer then repeats the nine leading questions. Finally, this procedure is repeated with a second set of nine leading yes/no questions. The Cronbach's alpha coefficient for the VSSC is .85 for Yield items and .75 for Shift items. The predictive validity of the VSSC is .62 (Warren, Scullin, & Ceci, 2000). The open-ended section was used as an assessment of children's event memory and the closed-ended section was used as a measure of children's susceptibility to suggestion.

Procedure

All 64 children were interviewed four times. The first two interviews were spaced an average of 3.09 ($SD = 2.37$) days apart so that there was a delay between the target event and the memory test (described below). The third and fourth interviews were spaced an average of 4.97 ($SD = 4.43$) days apart to accommodate the VSSC procedures. For the vast majority of children the delay between the first and second, and third and fourth interviews, were each within 2-5 days. However, due to absences and school closings some children had longer delays between interviews. All interviews were conducted in quiet rooms at children's schools. The interviewer

and child sat across from each other at a child-sized table or on the floor for each interview. Children were offered breaks between tasks as needed.

Visit 1. The first visit consisted of a 1-5 minute warm-up to help the child feel comfortable and to add to the estimate of language skills. Next, children participated in a staged event. The primary experimenter excused herself, leaving the child alone in the interview room. At this point a confederate entered the room, introduced herself to the child and engaged the child in a series of brief tasks including reading a book, playing “Simon Says”, and interacting with toys (see Appendix B). Finally, the confederate picked up a battery-operated toy that was left in plain sight and pushed the button that should activate the toy. When the toy did not work, the confederate said “...hmm, I better take it. Don’t tell anybody, okay”. After the child consented, the confederate left the room and the experimenter returned. Next, the interviewer asked the child about what occurred in her absence by asking the child “So, what happened while I was out of the room just now?” and following up with open-ended prompts such as “what else” and “please tell me one more thing”. Finally, the interviewer asked children closed-ended questions which included some leading questions (see Appendix C). After the target event children were tested with the TOLD-P:4 subcomponents. The warm-up conversation and the interview about the staged event were audio recorded.

Visit 2. During the second visit children were interviewed about the staged event again. As after the original event, interviewers asked open-ended questions about the event and followed up with the same battery of closed-ended questions and the interview was audio-recorded. Next, children completed the subtests of the WPPSI-III.

Visit 3. The third visit was not temporally linked to the first or second visit except that it occurred within the same general testing period, within the 3 weeks surrounding the other

interviews. During this visit children completed the CMS core tests and watched the five-minute video from the VSSC.

Visit 4. During the final visit children completed the VSSC interview, answered the forensic memory questions, and were thoroughly debriefed. The Forensic Memory (FM) interview was audio-taped and the interviewer also made note of whether the child noted information falling into the categories of “who”, “what”, “when”, “where”, and “why” about the event. If the child did not address each of these categories in his or her open-ended response, the interviewer prompted the child with a question from the neglected category(ies). For example, if a child discussed his or her birthday, but failed to mention anyone who was present, the interviewer might ask “Who was with you on your birthday”.

In the debriefing phase the interviewer explained that she learned she made some mistakes when interviewing the child. She then listed all of the leading items from the staged event interviews and from the VSSC interview and provided the correct responses. Finally, she said “remember that toy that I couldn’t find before? It turns out that [confederate name] had it the whole time! She just took it so she could fix it for me. Wasn’t that a nice surprise” and presented the working toy to the child to interact with.

Table 1. Visit Descriptions

Visit 1	Description	Visit 2	Description	Visit 3	Description	Visit 4	Description
Warm up	Unstructured conversation	Target Event Interview	Identical to Visit 1 Target Event Interview	CMS	Dot Locations, Stories, Faces, Word Pairs, Number, and Sequences	VSSC	Open- and Closed- ended questions about the VSSC video
Target Event & Interview	Staged Event, open- and closed-ended questions	WPPSI-III	Verbal Intelligence Quotient	VSSC Video	Video about a boy’s birthday party	Forensic Memory	Structured interview about 4 past events
TOLD-P-4	Relational Vocabulary and Syntactic Understanding					Debriefing	

Coding

Target Event Recall. Children's open-ended responses were coded for the number of correct person, action, object, and descriptive details recalled as well as for commissions in those categories. Commissions were incorrect elements that children spontaneously added into their event narratives. Closed-ended responses were coded for the number of correct person, action, and object details and for both commissions and omissions of event components. For example, if a child was asked "What was the name of the little bear in the book?" they would receive credit for a correct person detail if they responded "Bartholomew", a commission of a person detail if they responded "Bob", and an omission of a person detail if they responded "I don't remember". An additional code was included when coding children's responses at the second time point. A "shift" score was applied in cases where a child's response in the second interview changed from his or her response during the first interview. Shifts were further categorized into shifts to correct responses (from incorrect responses) and shifts to incorrect responses (from correct responses).

VSSC. The VSSC was scored with a coding scheme developed by Memon and Holliday (2006), which is based on Scullin's original coding scheme (Scullin & Ceci, 2001; Scullin & Hembrooke, 1998; Scullin, Kanaya, & Ceci, 2002). Correct items were scored as relating to persons, objects, actions, or surroundings. Children received one point for each correctly recalled item.

Forensic Memory. Children's responses to the four forensic memory topics were coded for inclusion of "who", "what", "when", "where", and "why" details. If a response addressed a category at least once, the child received a score of one for that category. Therefore, children's scores could range from 0-5 for each memory topic. Because interviewers were instructed to give children specific prompts for any categories they failed to mention spontaneously, the number of

categories interviewers prompted before the child provided a spontaneous answer were also coded. The interviewer score could also range from 0-5 for each memory topic. Finally, a Forensic Memory Ratio was calculated by dividing the child's score (+1) by the interviewer's score (+1). Therefore, total child and interviewer prompt scores could each range from 0 to 20 and Forensic Memory ratio scores could range from 0.05 $((0+1)/(20+1))$ to 21 $((20+1)/(0+1))$.

Reliability. Two independent coders coded approximately 20% of the transcripts for each task. Cohen's Kappa for Target Event Recall was .98, $p < .001$ and for the closed-ended section it ranged from .85 to .96, all p-values $< .001$. There was perfect agreement on the VSSC Open-Ended section, and for Forensic Memory Kappas ranged from .77 to .95 for child codes and from .63 to .92 for interviewer codes, all p-values $< .001$. Disagreements were discussed until final codes were agreed upon. One of the coders assigned to each task then completed the remaining coding.

CHAPTER 3

RESULTS

Preliminary Analyses

Preliminary analyses revealed that there were no relationships between child gender, interviewer, delay between first and second interview, or delay between third and fourth interview, and any of the variables of interest. Gender, interviewer, and interview delay variables were excluded from the remaining analyses. Overall, children in the sample provided high ratios of correct to incorrect details. Sixty-seven percent of the 3-year-olds, 78% of the 4-year-olds, and 71% of the 5-year-olds in the sample provided more accurate than inaccurate details in their open-ended responses at the second time point. Fifty-six percent of the 3-year-olds, 79% of the 4-year-olds, and 75% of the 5-year-olds provided more accurate than inaccurate details in their closed-ended responses at the second time point.

For several of the following analyses backward stepwise regressions were conducted to reveal the most minimal models that would predict the accuracy of children's memory reports. Language and memory variables are generally analyzed separately both to increase statistical power in each model and because language and memory are considered separate competencies in the legal system (Myers, 1997).

Language and Accuracy Ratio

Backward stepwise regressions were conducted predicting the ratio of children's correct (+1) to incorrect (-1) at the second time point. Child age in months, raw scores for VIQ (VI), RV, and SU, and Forensic Language were included as predictors in the original model for each analysis. The two raters' estimates of children's receptive language, $r = .34, p < .01$, and

productive language, $r = .30, p < .05$, were correlated and all four scores (two raters, two measures each) were summed to form the Forensic Language variable.

At the final step of the regression predicting children's Open-Ended Accuracy Ratios only VI remained, and it explained 16% of the variance in children's scores ($R^2 \text{ adjusted} = 14\%$). The overall relationship was significant, $F(1, 62) = 11.41, p = .001$. With other variables removed, Open-Ended Accuracy Ratio scores were positively related to VI, increasing by .12 ($SE = .04$) for each unit increase in VI.

As with the open-ended model, in the model predicting children's performance on the closed-ended questions only VI remained at the final step with one high-influence outlier removed (Cook's distance > 1). The final model explained 30% of the variation in children's Closed-Ended Accuracy Ratios ($R^2 \text{ adjusted} = 28\%$). The overall relationship was significant, $F(1, 61) = 25.49, p < .001$. With other variables removed, Closed-Ended Accuracy scores were positively related to VI, increasing by .03 ($SE = .01$) for each unit increase in VI.

Memory and Accuracy Ratio

Next, regressions were conducted with age, CMS, the VSSC Open-Ended response score, VSSC Suggestibility score, and Forensic Memory ratio as predictors. The Forensic Memory ratio for the Breakfast event was correlated with Yesterday, $r = .59, p < .001$ and Summer, $r = .35, p < .05$, but not Birthday, $r = .193, p = \text{ns}$. The Forensic Memory ratio for Yesterday was correlated with Summer, $r = .33, p < .05$, but only marginally correlated with Birthday, $r = .26, p = .08$. The Forensic Memory ratio for Summer was the only variable that was significantly correlated with Birthday, $r = .28, p < .05$. It is likely that the variation in temporal proximity between children's birthdays and the interview explains the inconsistent relationship between the Birthday Forensic

Memory variable and the Forensic Memory ratio for other events. The four events were collapsed to create one Forensic Memory ratio that was used for all of the following analyses.

In the final step of the regression predicting children's open-ended responses only CMS remained as a predictor. The overall relationship was significant, $F(1, 56) = 11.04, p < .01$. With other variables removed, Open-Ended Accuracy scores were positively related to CMS scores, increasing by .09 ($SE = .03$) for every point increase in a child's CMS score. The final step of the model explained 17% of the variation in children's Open-Ended Accuracy Ratios ($R^2_{adjusted} = 15\%$).

In contrast, the Open-Ended and Suggestibility components of the VSSC were the sole predictors remaining at the final step of the regression model predicting Closed-Ended Accuracy Ratios, $F(2, 55) = 4.50, p < .05$. With other variables removed, Closed-Ended Accuracy scores were positively related to VSSC Open and negatively related to VSSC Suggestibility, increasing by .03, $SE = .01$ for every extra detail recalled in the VSSC Open, and decreasing by .03, $SE = .01$ for every extra unit of suggestibility. The effect of VSSC Open was significant, $t(55) = 2.55, p < .05$, and the effect of VSSC Suggestibility was marginal, $t(55) = 1.79, p = .08$. Children who remembered higher numbers of correct details about the birthday party video and children who were less suggestible, had higher ratios of correct to incorrect details in their closed-ended responses about the staged event. The final model explained 14% of the variability in Closed-Ended Accuracy Ratio scores ($R^2_{adjusted} = 11\%$).

Full Model

Finally, two regressions were run- one predicting children's Open-Ended Accuracy Ratios with the all variables from the final steps of the relevant language and memory regressions above, and one predicting children's Closed-Ended Accuracy Ratios with all the

variables from the final steps of the Closed-Ended language and memory models above. For each model, language variables were entered at the first step, and memory variables were entered at the second step.

Table 2 presents the findings for the Open-Ended model. The full model explains 17% of the variation in children’s Open-Ended Accuracy Ratios (R^2 adjusted = 14%), however, adding CMS at the second step did not add to the predictive power of the model, F change (1, 56) = .92, p = ns. Verbal Intelligence alone explained 16% of the variance (R^2 adjusted = 14%). The pattern of findings was similar for the Closed-Ended model (Table 3). Verbal Intelligence alone explained 28% of the variation in children’s Closed-Ended Accuracy Ratios (R^2 adjusted = 26%) and the addition of the two VSSC measures did not have a significant influence on the model, F change (2, 56) = .79, p = ns. Therefore, while VI and model-specific memory variables can each predict children’s accuracy ratios in separate models, the predictive power is not additive.

Table 2. Full Model Predicting Open-Ended Accuracy Ratio

	B	SE B	β
Step 1			
Constant	-.99	1.95	
VI	.12	.04	.39**
Step 2			
Constant	-1.67	2.08	
VI	.10	.05	.31*
CMS	.03	.03	.96

Note: R^2 = .16 for Step 1, ΔR^2 = .01 for Step 2, $^+p < .10$, $*p < .05$, $**p < .01$

Performance Consistency

Consistency in testimony is another factor that might speak to the accuracy of a child’s memory (Bruck, Ceci, & Hembrooke, 1997; Conte, Sorenson, Fogarty, & Rosa, 1991; Gordon & Follmer, 1994; but see Malloy & Quas, 2009; Quas, Davis, Goodman, & Myers, 2007) and has been demonstrated to contribute to estimations of children’s testimonial accuracy (Leippe,

Manion, & Romanczyk, 1992, Leippe & Romanczyk, 1989). Overall, the ratio of accurate to inaccurate information in children's open-ended reports shifted significantly from Time 1 ($M = 8.50, SD = 5.63$) to Time 2 ($M = 5.36, SD = 4.52$) so that on average, children were less accurate at Time 2, $t = 3.72 (63), p < .001, r = .42$. However, children's Time 1 and Time 2 responses were highly correlated when correct responses, $r = .58, p < .001$ and incorrect responses, $r = .35, p < .01$ were considered separately. Children's accuracy ratios for closed-ended questions were consistent between time points, $r = .87, p < .001$, as were their correct responses, $r = .78, p < .001$, and incorrect responses, $r = .75, p < .001$.

Table 3. Full Model Predicting Closed-Ended Accuracy Ratio

	B	SE B	β
Step 1			
Constant	.30	.27	
VI	.02	.01	.52***
Step 2			
Constant	.43	.36	
VI	.02	.01	.48**
VSSC Open	.01	.01	.14
VSSC Suggestibility	-.01	.01	-.08

*Note: $R^2 = .28$ for Step 1, $\Delta R^2 = .02$ for Step 2, $^+ p < .10, *p < .05, **p < .01, ***p < .001$*

To detect individual differences in performance consistency, regressions were conducted with measures of accuracy consistency as the outcome variables. For the open-ended section these variables consisted of the differences in the number of correct responses and incorrect responses between Time 1 and Time 2. The models were run with the directionality of shift maintained (ie: positive numbers indicated that children were less accurate at Time 2 and negative numbers indicated that children were more accurate at Time 2) and with absolute distances between scores at Time 1 and Time 2 (indicating the magnitude of change between Time 1 and Time 2).

For the closed-ended questions there were two change variables- shifts from correct to incorrect responses and shifts from incorrect to correct responses. Shifts to incorrect responses and shifts to correct responses were not correlated, $r = .05$, $p = ns$, so they were considered separately.

Language and Consistency

All of the following consistency analyses were conducted using backward stepwise regressions. The first set of regressions predicted 1) the direction of change and 2) the magnitude of change in children's open-ended responses between Time 1 and Time 2. Increases in SU scores were associated with decreases in open-ended correct items recalled from Time 1 to Time 2 and with the absolute difference in scores between time points. SU explained 7% of the variance in children's answer shifts between time points, $B = .24$, $SE = .11$, $F(1, 62) = 4.49$, $p < .05$ and 10% of the absolute difference in scores between time points, $B = .24$, $SE = .08$, $F(1, 62) = 8.00$, $p < .01$. There were no relationships between children's language scores and the change in incorrect responses between Time 1 and Time 2.

The next set of regressions used age and language variables to predict shifts in children's closed-ended responses between Time 1 and Time 2. Only Forensic Language scores remained in the final step of the model predicting children's shifts to correct responses. The regression was a poor fit ($R^2 \text{ adjusted} = 8\%$), but the overall relationship was significant, $F(1, 62) = 6.12$, $p = .05$. With other variables removed, shifts to correct closed-ended responses between Time 1 and Time 2 decreased by $-.10$, ($SE = .04$) for each point increase in children's Forensic Language scores.

RV score was the only variable that remained at the final step predicting children's shifts to incorrect responses between time points. The regression was a poor fit ($R^2 \text{ adjusted} = 7\%$), but

the overall relationship was significant, $F(2, 58) = 5.37, p < .05$. With other variables removed, the number of closed-ended shifts to incorrect responses decreased by .10 ($SE = .04$) for each unit increase in children's RV scores.

Memory and Consistency

Next, backward stepwise regressions were conducted with age and memory variables predicting shifts in open-ended responses. None of the variables predicted decreases in correct details between time points, but children's CMS scores did appear to predict the absolute difference in correct responses between Time 1 and Time 2. However, the regression was a poor fit ($R^2 \text{ adjusted} = 5\%$), and the overall relationship was only marginally significant, $F(2, 57) = 3.86, p = .05$. With other variables removed, the absolute difference between open-ended correct responses increased by .06, $SE = .03$ for each increase in the total points children were awarded on the CMS.

Children's shifts in incorrect responses to the open-ended prompt were negatively related to CMS, $F(2, 56) = 5.66, p < .05$, but CMS scores explained little variance in shifts in incorrect responses ($R^2 \text{ adjusted} = 9\%$). With other variables removed, children's incorrect responses on the open-ended section decreased by .03, $SE = .01$ between Time 1 and Time 2 for each increase in the total points they were awarded on the CMS. None of the variables were associated with the absolute difference between incorrect responses between Time 1 and Time 2.

In the regression predicting shifts to correct responses in the closed-ended section between time points VSSC Suggestibility remained as a predictor at the final step. The regression explained little variation in children's shifts ($R^2 \text{ adjusted} = 6\%$), but the overall relationship was significant, $F(2, 56) = 4.76, p < .05$). With other variables removed, children had .07, $SE = .03$ more shifts to correct responses for each unit increase in VSSC Suggestibility.

There was no association with shifts to incorrect responses and any of the variables entered in the model.

Competency Cutoffs

The previous analyses have demonstrated the linear relationships between language and memory measures and children's later accuracy. However, competency determinations require a cutoff above which a child can be considered competent to testify. To get a sense of which language and memory scores are associated with a sufficient level of accuracy, independent t-tests were conducted on each of the language and memory variables based on an accuracy ratio cutoff of 1. In other words, children who provided more accurate information than inaccurate information were compared against children who provided an equal number of accurate and inaccurate details, or who provided more inaccurate than accurate details.

In this sample, 31% of children provided either an equal number of correct and incorrect details, or more incorrect than correct details in their open-ended reports and 27% of children provided equal numbers of correct and incorrect details, or more incorrect details in their closed-ended reports. In other words, if the acceptable competency cutoff is the ratio of correct to incorrect details, 31% of children in the sample were not competent to provide open-ended testimony and 27% of children were not competent to answer direct questions. Twenty-seven percent of the sample was competent to provide only one form of testimony (open- or closed-ended) and 15% of the sample was not competent to provide either form of testimony. Children who were most accurate in their open-ended responses were older, higher on both of the forensic variables, and higher on VIQ, SU, CMS, and VSSC Open. Relational Vocabulary scores were marginally higher in the more accurate group (Table 4).

Children who were most accurate in their closed-ended responses had higher scores on every language and memory variable except for VSSC Total Suggestibility and Forensic Memory Ratio. However, when VSSC Total Suggestibility was broken into its component parts, children's average Yield scores were higher in the lower accuracy group. There was no difference in VSSC Shift scores between groups (Table 5).

Table 4. Age and Language and Memory Variables by Open-Ended Low- vs. High Accuracy

	High Accuracy Ratio	Low Accuracy Ratio		
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t (df)</i>	<i>r</i>
Age	53.93 (6.03)	47.84 (7.37)	3.42 (60)**	0.40
VI	56.23 (12.88)	43.37 (14.67)	3.47 (60)**	0.41
RV	7.33 (5.53)	4.47 (5.10)	1.98 (60) ⁺	0.25
SU	15.86 (6.56)	11.11 (7.43)	2.53 (60)*	0.31
CMS	85.65 (23.17)	63.74 (20.15)	3.54 (60)**	0.42
VSSC Open	10.37 (6.79)	6.41 (5.37)	2.11 (58)*	0.27
VSSC Yield	5.73 (3.63)	6.63 (4.04)	.86 (58)	0.11
VSSC Shift	5.63 (3.71)	4.79 (3.28)	.851 (58)	0.11
For. Lang.	14.51 (3.73)	11.16 (4.18)	3.15 (60)**	0.38
FM Ratio	.91 (.31)	.65 (.20)	3.09 (55)**	0.38

Note: $p < .10^+$, $p < .05^*$, $p < .01^{**}$

Table 5. Age and Language and Memory Variables by Closed-Ended Low- vs. High Accuracy.

	High Accuracy Ratio	Low Accuracy Ratio		
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t (df)</i>	<i>r</i>
Age	53.72 (6.16)	48.35 (8.43)	2.78 (62)**	0.33
VI	57.30 (11.97)	37.35 (10.71)	6.05 (62)***	0.61
RV	7.38 (5.60)	3.47 (3.97)	2.65 (62)*	0.32
SU	16.60 (6.18)	9.18 (6.89)	4.12 (62)***	0.46
CMS	87.21 (21.63)	58.53 (19.37)	4.84 (62)***	0.52
VSSC Open	10.62 (6.74)	5.87 (4.91)	2.51 (62)*	0.30
VSSC Yield	5.48 (3.55)	7.75 (3.84)	2.16 (60)*	0.27
VSSC Shift	5.78 (3.89)	5.06 (3.55)	.65 (60)	0.08
For. Lang.	14.23 (3.60)	11.35 (4.78)	2.58 (62)*	0.31
FM Ratio	.86 (.27)	.76 (.41)	1.05 (57)	0.14

Note: $p < .05^*$, $p < .01^{**}$, $p < .001^{***}$

The VIQ, RV, and SU have established norms for the age group tested so next, Chi-Squares Tests were conducted which examined the influence of language proficiency (above or below the 50th percentile) on Open- and Closed-Ended Accuracy Ratios (Accuracy Ratio > 1 or ≤ 1). On average, children scored above the 50th percentile on VIQ $t(58) = 3.32, p < .01$, below the 50th percentile on RV, $t(62) = 5.35, p < .001$, and at the 50th percentile on SU, $t(62) = .05, p = ns$.

Children who performed at or above the 50th percentile on the VIQ or RV were no more likely than children with low VIQ or RV to provide a high ratio of accurate to inaccurate details in response to the open-ended prompt. However, VIQ was associated with Closed-Ended Accuracy Ratio, $\chi^2(1) = 9.20, p < .01$. The odds ratio indicated that children were 6.05 times more likely to have a high Closed-Ended Accuracy Ratio if they were at or above the 50th percentile on the VIQ. Relational Vocabulary was also associated with the Closed-Ended Accuracy Ratio, $\chi^2(1) = 5.25, p < .05$. Children with RV scores above the 50th percentile were 5.56 times more likely to also have high Closed-Ended Accuracy Ratios. SU was related to both Open-, $\chi^2(1) = 5.07, p < .05$, and Closed-Ended Accuracy Ratios, $\chi^2(1) = 12.33, p = .001$. Children who were above the 50th percentile on SU were 3.55 times more likely to have high Open-Ended, and 8.5 times more likely to have high Closed-Ended Accuracy Ratios, compared with children who had SU scores below the 50th percentile.

Using the cross tabulation of the verbal scores and accuracy ratios we can assess how many children would be incorrectly classified as competent or incompetent when using the above cutoffs for language proficiency and testimonial competency. Using children's VIQ percentiles, 16.7% of children in the sample would have been incorrectly excluded from providing closed-ended testimony and 10% would be incorrectly classified as competent to provide closed-ended testimony. Using the 50th percentile of RV as a cutoff, 42.2% of the

sample would be incorrectly excluded from providing closed-ended testimony and 3.1% would be incorrectly deemed competent. Finally, if the 50th percentile of SU scores was used as the competency standard, 22.6% of the sample would have been incorrectly excluded from providing open-ended testimony, 20.3% would have been incorrectly excluded from providing closed-ended testimony, 11.3% would be incorrectly found competent to provide open-ended testimony and 6.3% would be incorrectly classified as competent to provide closed-ended testimony.

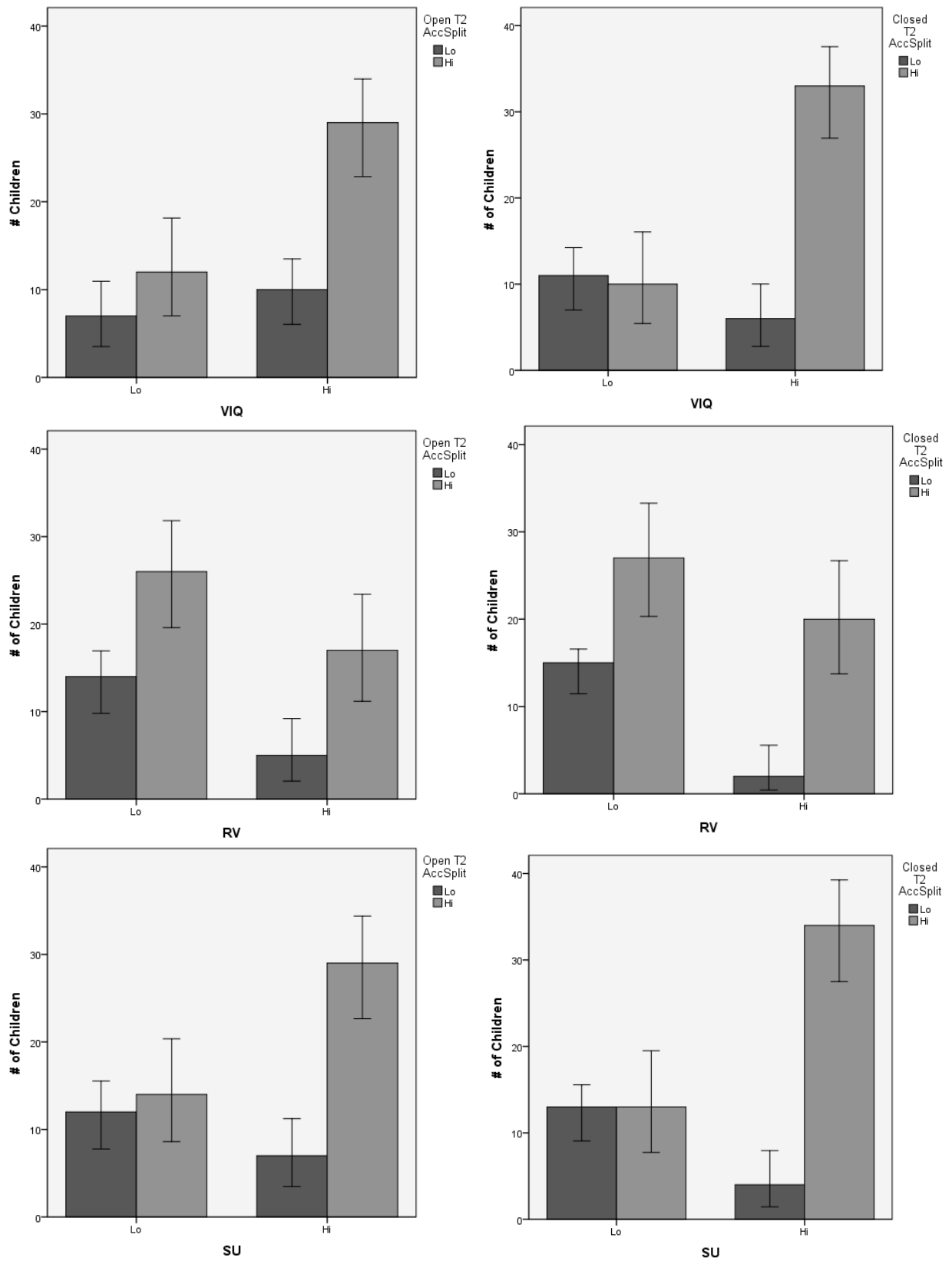


Figure 1. Number of Children by Language Percentile Group and Report Accuracy Group

CHAPTER 4

DISCUSSION

The first hypothesis of the present study was that language and memory variables that are used in developmental research could predict the accuracy of children's later statements about a staged event. The second was that these empirical measures of language and memory would predict children's performance better than the types of questions typically asked by courts to establish children's testimonial competency. The results of this study provided substantial support for the hypotheses. Several of the empirical measures predicted the ratio of accurate to inaccurate details in children's open- and closed-ended responses. There was not much evidence that the types of questions currently used in courts to determine linguistic and mnemonic competency could predict the accuracy of children's reports.

In the first set of analyses children's verbal intelligence scores predicted the accuracy of their open- and closed-ended reports. Children's Memory Scale scores predicted the accuracy of their open-ended responses and the open-ended section of the VSSC as well as the VSSC Suggestibility score predicted children's accuracy when answering mildly leading closed-ended questions. In line with the second experimental hypothesis, neither the forensic language questions nor the forensic memory questions predicted the accuracy of children's responses.

The findings indicating a link between children's language skills and memory accuracy is in line with previous work (Kulkofsky, 2010; Kulkofsky, Wang, & Ceci, 2008). However, overall there has been little empirical research examining individual differences which contribute to children's report accuracy, particularly pertaining to general memory abilities. There is a body of related literature which has focused on the individual differences that are associated with children's suggestibility, including the relationships between suggestibility and children's

performance on language and memory measures (see Bruck & Melnyk, 2004 for a review). The present results are consistent with this literature as well. Bruck and Melnyk (2004) report that studies which assess language with a comprehensive measure find a relationship between language and suggestibility such that children with higher language scores tend to be less suggestible (ie: more accurate). The WPPSI verbal intelligence score was the most comprehensive measure of child language used in the present study and it is therefore not surprising that it was the most robust language measure for predicting children's report accuracy.

Few studies that have examined the relationship between general memory scores and children's suggestibility found significant relationships between the two (Bruck & Melnyk, 2004). In line with the previously reported findings about individual differences in suggestibility, the comprehensive language variables in the present study explained more of the variation in children's report accuracy than did the memory variables. While the CMS, VSSC Open, and VSSC Suggestibility did independently predict children's accuracy, they explained no extra variance when VI was included in the model. The weak association between VSSC Suggestibility and children's report accuracy was inconsistent with the limited literature on the VSSC.

Previous work has demonstrated a negative relationship between children's VSSC Suggestibility scores and memory accuracy (McFarlane & Powell, 2002; Scullin, Kanaya, & Ceci, 2002). However, the suggestibility component of the VSSC may not have been as robust at predicting children's performance in the present study because the target interviews were not very suggestive. This interpretation is supported by the fact that the VSSC open-ended component, a non-suggestive portion of the VSSC, was significantly associated with children's performance in the present study, and more strongly related to children's accuracy than the

suggestive portion of the VSSC.

Although the language and memory variables significantly predicted the accuracy of children's responses at Time 2, these variables were less consistent at predicting shifts in children's responses between Time 1 and Time 2. Syntactic Understanding and CMS predicted both the magnitude of the shift in correct responses and the direction of children's shifts in correct responses. Children with higher SU scores tended to shift toward having fewer correct responses at Time 2 and children with higher CMS scores shifted toward having more correct responses at Time 2. The Forensic Language variable and VSSC Suggestibility score were associated with children's tendency to shift to correct answers in their closed-ended responses and RV scores were negatively associated with children's shifts to incorrect responses.

Thus, in the present study it does not appear that individual differences have a clear relationship with the consistency of children's testimony, particularly for closed-ended responses. This may be because children's responses between time points were highly correlated. This suggests that the magnitude and direction of change may not have differed much between children and time may have had more of an impact on shifts in children's responses than did individual differences.

These findings are also consistent with literature suggesting that there is no relationship between children's report consistency over time and the accuracy of their reports (Malloy & Quas, 2009; Quas, Davis, Goodman, & Myers, 2007). In the present study the language and memory variables were associated with accuracy, but the pattern of association with shifts was less clear and less robust. If the consistency of children's testimony relies more on context than on children's abilities, then it would follow that there would not be a strong relationship between consistency and variables predicting testimonial accuracy. However, more research is needed to

examine whether there are stable individual differences in children's abilities to provide consistent testimony. Future research could also examine whether consistency speaks more to a child's competency, or credibility, or, in line with the arguments made by Quas and colleagues (Malloy & Quas, 2009; Quas et al, 2007), perhaps consistency should not be considered when evaluating a child's legal testimony.

The present study's findings concerning competency cutoffs supported the first hypothesis, that empirical measures of language and memory would predict report accuracy, and partially supported the second hypothesis, that the empirical variables would predict children's report accuracy more robustly than would the forensic variables. Children who were deemed competent in this study differed from incompetent children on every empirical variable, but they also differed somewhat on the forensic variables. The children who could provide more accurate than inaccurate details in their open-ended reports tended to be older and higher on VI, SU, CMS, VSSC Open, and both forensic variables. They were also marginally higher on RV.

Children who provided more accurate than inaccurate answers in response to closed-ended questions were older and scored higher on all four language variables, including the forensic language assessment. They also scored higher on CMS and VSSC Open and they were less likely to yield to suggestive questions in the VSSC. Although both the forensic language and the forensic memory variable were associated with higher open-ended accuracy ratios, the effect sizes for the forensic variables were slightly smaller than the effect sizes for Age, VI, or CMS. The effect size for the relationship between forensic language and children's closed-ended accuracy ratios was smaller than the effect sizes for VI, SU, and CMS, and comparable to the effect sizes for Age, RV, and VSSC Open. These findings are largely compatible with the results

of the original regression analyses in which VI, CMS, and VSSC explained more variance in the accuracy of children's responses than the other language and memory measures.

Finally, children who were average or above average for the population on the three normed tests were more accurate than children who were below average. Children who were at or above the 50th percentile on VIQ and RV were more likely to be competent at answering closed-ended questions than children who were below the 50th percentile. Children at or above the 50th percentile on SU were more likely to be competent at answering both open- and closed-ended questions. However, courts may choose to shift the percentile cutoff for these measures to adjust acceptable errors. In the present study, by using the 50th percentile on each of the tests as the cutoff point, depending on the test used, between 17 and 42% of the sample would have been excluded from providing closed-ended testimony, despite providing more accurate than inaccurate details in their closed-ended responses. Between 3 and 11% would have been allowed to provide testimony and then provided either an equal number of correct and incorrect details, or more incorrect than correct details.

The findings from the present study complement the existing research on children's testimonial competency. I know of only two studies which have examined the language and memory components of the competency exam. The first focused on creating an exam with internal validity with a pilot sample of 17 children, but no attempt was made to measure predictive validity (Hansen, 1990). Further, the exam was not compared against current methods of competency assessment. The present findings address this gap by assessing predictive validity and by comparing the predictive ability of empirical measures with an approximation of the predictive power of current methods. In the present study efforts were also made to maintain validity and reliability for many of the pilot variables by selecting them from standard cognitive

tests for children.

The second study examined a number of child variables including language ability, but not general memory skills, as predictors of testimonial performance (Jacobson, 2002). The outcome variables for the study included factors relating to the narrative quality of children's responses as well as child demeanor on the stand and case outcome. Because court transcripts were examined instead of controlled staged-events, the accuracy of the children's reports could not be directly examined. In fact, the author's competency ratings had no relationship with case outcomes. Consistent with the current study and with the vast majority of existing literature, older children provided higher-quality testimony. However, unlike the present study, children's verbal skills were associated only with the confidence children demonstrated on the stand, and not with the quality of testimony. This discrepancy is not surprising given that both the verbal measures and outcome variables differed between studies. While this in-depth study of children's testimonial abilities provides insight into the qualities of children's in-court testimony and provides high ecological validity, it cannot address the issue of predictive validity.

The present work also complements the developing literature on the truth and lie components of the competency exam. The pioneering work on children's truth and lie-telling and testimonial competency has demonstrated that the types of questions currently asked to establish children's truth-telling competencies are not predictive of children's truth-telling behavior (London & Nunez, 2002; Talwar et al, 2002). These studies further suggest that discussing the importance of truth-telling with children, and especially, asking children to promise to be truthful, are sufficient in increase children's truth-telling behavior (London & Nunez, 2002; Talwar et al, 2002). Some initial work has also been done proposing a novel method of assessing truth-telling behavior, though as of yet, it has not been validated (Lyon & Saywitz, 1999, 2000).

When combined with these findings, the present work can be used to contribute to the policy and practice of testimonial competency determinations.

However, the present study is limited in ecological validity. Though portions of the staged events and some of the interview questions were designed to loosely approximate court cases, children were not exposed to the stress and trauma that would normally be associated with testifying in court and there is research evidence that suggests that memory for high-stress situations is different from memory for low-stress situations (Chen, Zeltzer, Craske, & Katz, 2000; Bauer, 2004). Further, there were very few children in the sample who were in the bottom quartiles of the standardized tests, and most children could provide more accurate than inaccurate information about the staged events. Therefore, there were few children from the population of interest- those with low testimonial competency. Future research should investigate how the language and memory variables from the present research function with actual child witnesses and should examine whether the predictive validity of the measures hold up with larger samples of children who are low on these measures. The latter research could be particularly useful to the majority of courts whose goal is to exclude only child witnesses with very low capacity to provide useful testimony.

Recommendations and Conclusions

The findings from this study were in line with the gist of competency law. Children with higher language and memory abilities were more accurate when providing testimony about a series of staged events. However, approximations of the types of questions currently used in courts to assess language and memory proficiency were not the most effective measures. The raw scores from the WPPSI-III verbal intelligence subtests, a non-leading event memory task, and the core tests from the Children's Memory Scale emerged as the most robust predictors of

accuracy. Based on these initial data it appears that when a child witness's competency is called into question the presiding judge should request that these tests be administered by a trained psychologist in order to determine competency. However, more research is needed to pinpoint which scores warrant excluding a child from testifying in court.

These data also speak to the historically controversial question of whether age should be considered in competency decisions (Child Victims and Child Witnesses Rights Act, 1990; Myers, 1992, Myers, 1997). While the data replicated the well-established finding that older children tend to be more accurate (see Ceci & Bruck, 1993 for a review), some measures predicted more variance in children's performance than did age. In fact, age was excluded in the final steps of the primary regression analyses indicating that age did not explain any additional variance in children's responses above the variance explained by the language and memory variables. Additionally, Chi-Square tests indicated that even when children's language scores were standardized to remove the influence of age, children with higher language scores were more accurate. Therefore, the present research suggests that child age should not be considered when making competency decisions, which is in line with federal law, and the law of most state courts (Child Victims and Child Witnesses Rights Act, 1990; Myers, 1992, Myers, 1997).

Research in the past two decades has begun drawing distinctions between children's performance on traditional tests of testimonial competency and their performance in court. These distinctions are both counterintuitive and disturbing. While the present findings support competency law as it is written, we argue that the practice of competency determination is both unreliable and not valid. The effect of current practice may be that incompetent children are put on the stand and competent children miss their opportunity for justice. These effects are particularly dangerous in cases of child abuse where the only evidence of abuse usually lies in

children's testimony. Only in rare cases are other witnesses or physical evidence available (Ceci & Bruck, 1995). Unreliable and invalid competence exams could result in the prosecution of innocent people on the one hand and under-prosecution of child-abusers on the other hand. The present findings, when combined with the work on the truth and lie competent of the competency exam, and with future research, have the potential to inform developmental theory and to improve the legal system for child victims and witnesses.

APPENDIX A

Forensic Memory Interview

1) Tell me about breakfast this morning.

- Who (Who was there, who made breakfast, etc)
- What (What did you eat)
- When (When was breakfast, what did you do just before and after)
- Where (Where did you eat your breakfast)
- Why (Why did you eat what you ate, why did you eat at that time)

2) What happened on your last birthday?

- Who (Who was there, who planned it)
- What (What did you eat, what did you get for presents, what did you do)
- When (When is your birthday, what happened just before and after)
- Where (Where did you have your birthday celebration)
- Why (Why did you have your birthday where you did, why did you eat what you ate, why did you do what you did, why were the people there that were there)

3) What did you do yesterday?

- Who (Who was there, who made breakfast, etc)
- What (What did you eat)
- When (When was breakfast, what did you do just before and after)
- Where (Where did you eat your breakfast)
- Why (Why did you eat what you ate, why did you eat at that time)

4) What did you do last summer?

- Who (Who was there, who did you play with, which grown ups did you spend time with)
- What (What did you do)
- When (When did you do things listed above, what order did they happen in)
- Where (Where did you spend your summer)
- Why (Why did you spend the summer the way you did)

APPENDIX B

Staged Event

“Hi [child name]! I’m [confederate name]. I’m here to play some games with you while we’re waiting for [interviewer].”

- 1) “I brought this book for us to read together. Let’s do that first.”

[read the book *I Love you Just the Way you Are* (1998)]

- 2) “Okay, how about a game of ‘Simon Says’? Do you know how to play that game? I tell you to do silly things, but you only do them if I say ‘Simon Says’ first. Like, if I said ‘[child’s name] touch your nose’ would you do it? [That’s right! You wouldn’t!] or [Actually, you wouldn’t because I didn’t say ‘Simon Says’] What if I said ‘Simon says jump up and down’? [That’s right! You’d jump up and down since I said ‘Simon Says’] or [Actually, you would jump up and down because I said ‘Simon Says’]. Great, let’s try it!”

- a. “Simon Says put on this tee-shirt”
- b. “Touch your cheeks”
- c. “Simon says throw me the ball”
- d. “Stomp your feet”
- e. “Simon says take off the tee-shirt”
- f. “Clap your hands”
- g. “Simon says do a silly pose for the camera” [confederate pretends to take child’s photo with a play camera]

“Good job!”

- 3) “Let’s look at some of the toys in here.”

- a. “This one looks like a fire truck, neat, you can drive it around and pretend to fight fires!”
 - i. Action: roll the truck around and make fire truck noises
- b. “And this one is a doll. Look at this, you can take her dress on and off and play dress up.”
 - i. Action: lift the doll’s dress up and down
- c. “Wow, look at this one, it looks like you can push this button and it will do something.”
 - i. Action: push button

“Hmm, I guess it’s broken, I better take it. Don’t tell anybody, okay?”

- ii. Wait for child to agree (or disagree). If the child does not agree say “please?”
- iii. If the child still doesn’t agree, skip ahead.

“I had a great time playing with you, bye [child’s name]!”

APPENDIX C

Staged Event Interview

“Hi again [child’s name]. So, what happened while I was out of the room just now?”

*follow up with long pauses, and if needed, open-ended prompts such as “what else”, “what else can you tell me”, “please tell me one more thing”, “how about one more thing?”

Closed-ended

“[Confederate name] told me about some of the things she did with you, so next I’m going to ask you some questions about that”

- 1) What was the name of the little bear in the book? [Ba/Bartholomew]
- 2) What was the name of the big bear in the book? [George]
- 3) What was wrong with the little bear’s porridge? [too lumpy]
- 4) Did the little bear have a bath? [yes]
- 5) Did the big bear help the little bear brush his teeth? [no]
- 6) Did little bear give big bear a kiss? [yes]
- 7) What did the little bear do when he got lost? [leading- he didn’t get lost]
- 8) Did the little bear talk to his friend Bill? [leading- there was no character named Bill]
- 9) What kind of clothing did you put on during Simon Says? [blue teeshirt]
- 10) What did you do with the ball? [threw it to experimenter]
- 11) What did you do with your feet? [confederate asked them to stomp, but didn’t say ‘Simon Says’]
- 12) Did you touch your cheeks? [didn’t say ‘Simon Says’]
- 13) Did someone ask you to stomp your feet? [yes]
- 14) Did someone ask you to take off some clothes and pose for a picture? [yes]
- 15) What color was the hat [confederate] asked you to put on? [leading- child wasn’t asked to put on a hat]
- 16) Did [confederate] ask you to touch your toes? [no]
- 17) What was the first toy you played with after ‘Simon Says’? [fire truck]
- 18) What was the doll wearing? [pink dress]
- 19) What happened when you pushed the button on the last toy? [nothing- it was broken]
- 20) Did someone make fire truck noises? [yes]
- 21) Did someone lift up the doll’s clothes? [yes]
- 22) Did someone ask you to keep a secret? [yes]
- 23) What did you do with the toy monkey that was in the room? [leading- there was no monkey]
- 24) Did [confederate] show you the jump rope? [no, leading- there was no mention of a jump rope]
- 25) Do you know what happened to the Tigger toy? Where did it go? [the confederate took it, but asked the child not to tell]

REFERENCES

- 1 Leach 199, 168 E.R. 202 (1779)
- 40 Seton Hall L. Rev. 1169 2010
- Beuhring, T., & Kee, D. W. (1987). Developmental relationships among metamemory, elaborative strategy use, and associative memory. *Journal of Experimental Child Psychology, 44*, 377–400.
- Beuscher, E. & Roebbers, C.M. (2005). Does a warning help children to more accurately remember an event, to resist misleading questions, and to identify unanswerable questions? *Experimental Psychology, 52*(3), 232-241.
- Brainerd, C. J. & Reyna, V. F. (2007). Explaining developmental reversals in false memories. *Psychological Science, 18*(5), 373-417.
- Brainerd, C. J. & Reyna, V. F.; Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin, 134*(3), 343-382.
- Bauer, P.J. (2004). Event Memory. In William Damon, Richard M. Lerner, Deanna Kuhn, & Robert S. Siegler (Eds.), *Handbook of Child Psychology*, (Volume 2, pp. 267-297). Hoboken, NJ: John Wiley & Sons, Inc.
- Bruck, M., Ceci, S. J., & Hembrooke, H. (1997). Children's reports of pleasant and unpleasant events. In J Don Read and D. Stephen Lindsay (Eds.), *Recollections of trauma: Scientific research and clinical practice* (pp. 199–219). New York: Plenum Press.
- Bruck, M. & Melnyk, L. (2004). Individual differences in children's suggestibility: A review and synthesis. *Applied Cognitive Psychology, 18*(8), Special issue: Individual and Developmental Differences in Suggestibility, 947-996.

- Cashmore, J. & Bussey, K. (1996). Judicial perceptions of child witness competence. *Law and Human Behavior*, 20(3), 313-334.
- Ceci, S.J., & Bruck, M. (1993). The suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, 113, 403-439.
- Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony*. Washington, DC: American Psychological Association.
- Ceci, S.J., Crotteau, M.L., Smith, E., and Loftus, E.F. (1994) Repeatedly thinking about non-events: Source misattributions among preschoolers. *Consciousness and Cognition*, 3, 388-407.
- Ceci, S.J., Fitneva, S.A., & Williams, W.M. (2010). Representational constraints on the development of memory and metamemory: A developmental–representational theory. *Psychological Review*, 117(2), 464-495.
- Ceci, S.J., Papierno, P.B., & Kulkofsky, S. C. (2007). Representational constraints on children’s memory and suggestibility. *Psychological Science*, 18(6), 503-509.
- Chen, E., Zeltzer, L.K., Craske, M.G., & Katz, E.R. (2000). Children’s memories for painful cancer treatment procedures: Implications for distress. *Child Development*, 71(4), 933-947.
- Child Victims and Child Witnesses Rights Act of 1990, 18 U.S.C. 3509 (2008).
- Cohen, M. 1997. Children’s Memory Scale. San Antonio: Harcourt Assessment.
- Conte, J.R., Sorenson, E., Fogarty, L., & Rosa, J. (1991). Evaluation children’s reports of sexual abuse: Results from a survey of professionals. *American Journal of Orthopsychiatry*, 78, 428-437.
- DeCasper, A.J. & Spence, M.J. (1986). Prenatal maternal speech influences newborns’

- perceptions of speech sounds. *Infant Behavior and Development*, 9, 133-150.
- Eisen, M. L., Qin, J. J., Goodman, G. S., & Davis, S. L. (2002). Memory and suggestibility in maltreated children: Age, stress arousal, dissociation, and psychopathology. *Journal of Experimental Child Psychology*, 83, 167–212.
- Finkelhor, D., & Jones, L. (2006). Why have child maltreatment and child victimization declined? *Journal of Social Issues*, 62(4), 685-716.
- Fivush, R., Haden, C., & Adam, S. (1995). Structure and coherence of preschoolers' personal narratives over time: Implications for childhood amnesia. *Journal of Experimental Child Psychology*, 60(1), 32-56.
- Fivush, R., Hudson, J., & Nelson, K. (1984). Children's long-term memory for a novel event: An exploratory study. *Merrill-Palmer Quarterly*, 30(3), 303-316.
- Fritzley, V. H. and K. Lee (2003). Do young children always say yes to yes-no questions? A metadepvelopmental study of the affirmation bias. *Child Development*, 74(5), 1297-1313.
- Gershman, B. L., (2001). Child Witnesses and Procedural Fairness. *Pace Law Faculty Publications*. Paper 127. <http://digitalcommons.pace.edu/lawfaculty/127>.
- Goodman, G. S., Quas, J. A., Batterman-Faunce, J. M., Riddlesberger, M., & Kuhn, J. (1994). Predictors of accurate and inaccurate memories of traumatic events experienced in childhood. *Consciousness and Cognition*, 3, 269-294.
- Gordon, B. N., Ornstein, P. A., Nida, R. E., Follmer, A., Crenshaw, M. C., & Albert, G. (1993). Does the use of dolls facilitate children's memory of visits to the doctor? *Applied Cognitive Psychology*, 7, 459–474.

- Greenhoot, A.F., Ornstein, P.A., Gordon, B.N., & Baker-Ward, L. (1999). Acting out the details of a pediatric check-up: the effects of interview condition on individual differences in children's memory reports. *Child Development, 70*, 363-380.
- Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences, 5*, 303-314.
- Gudjonsson, G. H. (1987). A parallel form of the Gudjonsson Suggestibility Scale. *British Journal of Clinical Psychology, 26*, 215-221.
- Greenhoot, A.F., Ornstein, P.A., Gordon, B.N., & Baker-Ward, L. (1999). Acting out the details of a pediatric check-up: The impact of interview condition and behavioral style on children's memory reports. *Child Development, 70*(2), 363-380.
- Hale, S. (1990). A global developmental trend in cognitive processing speed. *Child Development, 61*(3), 653-663.
- Hall, S.R., Sales, B.D. *Courtroom modifications for child witnesses: Law and science in forensic evaluations*. Washington, D.C.: American Psychological Association; 2008.
- Hansen, D.A. (1990). The development of an instrument to test the testimonial competence of children: A pilot test of the Hansen competency test for children (HCTC). *Dissertation Abstracts International, 50*(8-B), 3697.
- Haugaard, J.J., Repucci, N.D., Laird, J. & Nauful, T. (1991). Children's definitions of the truth and their competency as witnesses in legal proceedings. *Law and Human Behavior, 15*(3), 253-271.
- Holliday, R.E., Douglas, K.M., & Hayes, B.K. (1999). Children's eyewitness suggestibility: Memory trace strength revisited. *Cognitive Development, 14*(3), 443-462.

- Jacobson, L.G. (2002). Psychological predictors of children's competency as witnesses in sexual abuse trials. *ETD Collection for Wayne State University*. Paper AAI3047558.
- Kail, R.V. (1986). Sources of age differences in speed of processing. *Child Development*, 57(4), 969-987.
- Kail, R.V. (2007). Longitudinal evidence that increases in processing speed and working memory enhance children's reasoning. *Psychological Science*, 18(4), 312-313.
- Kail, R. V. & Ferrer, E. (2007). Processing Speed in Childhood and Adolescence: Longitudinal Models for Examining Developmental Change. *Child Development*, 78, 1760-1770.
- Kohlberg, L. (1963). The development of children's orientations toward a moral order: I. Sequence in the development of moral thought. *Vita Humana*, 6(1-2), 11-33.
- Kulkofsky, S. (2010). The effects of verbal labels and vocabulary skill on memory and suggestibility. *Journal of Applied Developmental Psychology*, 31(6), 460-466.
- Kulkofsky, S. & Klemfuss, J.Z. (2008). What the stories children tell can tell about their memories: Narrative skill and young children's suggestibility. *Developmental Psychology*, 44, 1442-1456.
- Kulkofsky, S., Wang, Q., & Ceci, S.J. (2008). Do better stories make better memories? Narrative quality and memory accuracy in preschool children. *Applied Cognitive Psychology*, 22(1), 21-38.
- Leichtman, M.D. & Ceci, S.J. (1995). The effects of stereotypes and suggestions on preschoolers' reports. *Developmental Psychology*, 31(4), 568-578.
- Leippe, M.R., Manion, A., & Romanczyk, A. (1992). Eyewitness persuasion: How and how well do factfinders judge the accuracy of adults' and children's memory reports? *Journal of Personality and Social Psychology*, 63, 181-187.

- Leippe, M.R. & Romanczyk, A. (1989). Reactions to child (versus adult) eyewitnesses: The influence of jurors' preconceptions and witness behavior. *Law and Human Behavior*, 13(2), 103-132.
- Leippe, M.R., Romanczyk, A., & Manion, A.P. (1991). Eyewitness memory for a touching experience: Accuracy differences between child and adult witnesses. *Journal of Applied Psychology*, 76(3), 367-379.
- Loftus, E. F. (1992). When a lie becomes memory's truth: Memory distortion after exposure to misinformation. *Current Directions in Psychological Science*, 1(4), 121-123.
- London, K. & Nunez, N. (2002). Examining the efficacy of truth/lie discussions in predicting and increasing the veracity of children's reports. *Journal of Experimental Child Psychology*, 83(2), 131-147.
- Lyon, T.D. & Saywitz, K.J. (1999). Young maltreated children's competence to take the oath. *Applied Developmental Science*, 3(1), 16-27.
- Lyon, T.D., & Saywitz, K.J. (2000). Qualifying children to take the oath: Materials for interviewing professionals.
- Malloy, L.C., & Quas, J.A. (2009). Children's Suggestibility: Areas of consensus and controversy. In Kathryn Kuehnle & Mary Connell (Eds.), *The Evaluation of Child Sexual Abuse Allegations* (pp. 267-297). Hoboken, NJ: John Wiley & Sons, Inc.
- McGough, L. (1994). *Child witnesses: Fragile voices in the American legal system*. New Haven, CT: Yale University Press.
- Memon, A., Holliday, R., & Hill, C. (2006). Pre-event stereotypes and misinformation effects in young children. *Memory*, 14(1), 104-114.
- Miller, V. (1998). *I love you just the way you are*. London, England: Walker Books Ltd.

- Myers, J.E.B. (1986-1987). The testimonial competence of children. *Journal of Family Law*, 25(2), 287-356.
- Myers, J.E.B. (1992). *Legal issues in child abuse and neglect*. Newbury Park, CA: Sage Publications.
- Myers, J.E.B. (1997). *Evidence in child abuse and neglect cases*. New York, NY: John Wiley & Sons.
- Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: A social cultural developmental model. *Psychological Review*, 111, 486–511.
- Newcombe, R., & Reese, E. (2004). Evaluations and orientations in mother-child narratives as a function of attachment security: A longitudinal investigation. *International Journal of Behavioral Development*, 28, 230-245.
- Newcomer, P.L. and Hammill, D.D. 2008. Test of Language Development—Primary, 4th edition (TOLD-P:4). Austin, TX: Pro-ed.
- Ornstein, P. A., Merritt, K. A., Baker Ward, L., Furtado, E., Gordon, B. N., & Principe, G. (1998). Children's knowledge, expectation, and long-term retention. *Applied Cognitive Psychology*, 12, 387-405. doi:10.1002/(SICI)1099-0720(199808)12:43.0.CO;2-5
- Pezdek, K. & Roe, C. (1995). The effect of memory trace strength on suggestibility. *Journal of Experimental Child Psychology*, 60(1), Special issue: Early memory, 116-128.
- Piaget, J. (1932). *The moral judgment of the child*. New York: Harcourt, Brace.
- Pillemer, D.B., Picariello, M.L., & Pruett, J.C. (1994). Very long-term memories of a salient preschool event. *Applied Cognitive Psychology*, 8, 95-106.
- Pipe, M.E. & Wilson, J.C. (1994). Cues and secrets: Influences on children's event reports. *Developmental Psychology*, 30, 515-525.

- Quas, J.A., Davis, E.L., Goodman, G.S., & Myers, J.E.B. (2007). Repeated questions, deception, and children's True and false reports of body touch, *Child Maltreatment*, 12, 60-67.
- Reese, E., & Brown, N. (2000). Reminiscing and recounting in the preschool years. *Applied Cognitive Psychology*, 14, 1-17.
- Reese, E., & Fivush, R. (1993). Parental styles of talking about the past. *Developmental Psychology*, 29, 596-606.
- Rosenthal, R. (2002). Suggestibility, reliability, and the legal process. *Developmental Review*, 22, 334-369.
- Saywitz, K. J., Goodman, G.S. (1996). Interviewing children in and out of court: Current research and practice implications. In: *The APSAC handbook on child maltreatment*. Briere, John; Berliner, Lucy; Bulkley, Josephine A.; Jenny, Carole; Reid, Theresa; Thousand Oaks, CA, US: Sage Publications, Inc, 297-318.
- Saywitz, K.J., Goodman, G.S., & Lyon, T.D. (2002). Interviewing children in and out of court: Current Research and Practice Implications, In: *The APSAC handbook on child maltreatment*. John E.B. Myers, Lucy Berliner, John Brier, C. Terry Hendrix, Carole Jenny, & Theresa A. Reid (Eds.); Thousand Oaks, CA, US: Sage Publications, Inc, 349-378.
- Saywitz, K.J., Goodman, G.S., Nicholas, E., & Moan, S.F. (1991). Children's memories of a physical examination involving genital touch: Implications for reports of child sexual abuse. *Journal of Consulting and Clinical Psychology*, 59(5), 682-691.
- Saywitz, K.J., Snyder, L. (1993). Improving children's testimony with preparation. In: *Child victims, child witnesses: Understanding and improving testimony*. Goodman, Gail S.; Bottoms, Bette L.; New York, NY, US: Guilford Press, 117-146.

- Schneider, W., & Bjorklund, D.F. (1998). Memory. In W. Damon, R.S. Siegler, & D. Kuhn (Eds.), *Handbook of child psychology* (Vol. 2, pp. 467-521). New York: Wiley.
- Scullin, M., & Ceci, S. (2001). A suggestibility scale for children. *Personality and Individual Differences*, 30, 843–856.
- Scullin, M., & Hembrooke, H. (1998, March). Development and refinement of a suggestibility scale for children. Paper presented at the Biennial Conference of the American Psychology–Law Society, Redondo Beach, CA.
- Scullin, M. H., Kanaya, T., & Ceci, S.J. (2002). Measurement of individual differences in children’s suggestibility across situations. *Journal of Experimental Psychology: Applied*, 8(4), 233-246.
- Shields, P.J. & Rovee-Collier, C. (1992). Long-term memory for context-specific category information at six months. *Child Development*, 63(2), 245-259.
- Siegler, R.S. (1996). *Emerging minds: The process of change in children’s thinking*. New York: Oxford University Press.
- Stern, W. (1910). Abstracts of lectures on the psychology of testimony and on the study of individuality. *American Journal of Psychology*, 21, 270-282.
- Talwar, V., Lee, K., Bala, N., & Lindsay, R.C.L. (2002). Children’s conceptual knowledge of lying and its relation to their actual behaviors: Implications for court competence examinations. *Law and Human Behavior*, 26(4).
- U.S. Department of Health and Human Services. (2009). *Child Maltreatment*. Washington, DC: Author. Retrieved from <http://www.acf.hhs.gov/programs/cb/pubs/cm09/cm09.pdf#page=31>

- Walkenfeld, F. F. (2000). Reminder and language effects on preschoolers' memory reports: Do words speak louder than actions? Unpublished doctoral dissertation, City University of New York Graduate School.
- Walker, A.G. (1993). Questioning young children in court: A linguistic case study. *Law and Human Behavior*, 17(1), 59-81.
- Walker, A.G. (1985). From oral to written: The "verbatim" transcription of legal proceedings. University Microfilms International, Ann Arbor, Mich.
- Wheeler v. United States, 159 U.S. 523 (1895).
- Wechsler D. 2002. Wechsler Preschool and Primary Scale of Intelligence, 3rd edition (WPPSI-III). San Antonio, TX: Psychological Corporation.