ABSTRACT

MELUSO, ANGELA C. Retrieval-Enhanced Suggestibility: A Developmental Examination. (Under the direction of Dr. Lynne Baker-Ward.)

Retrieval-enhanced suggestibility (RES), a recently documented phenomenon, occurs when an initial interview about an observed event increases adults' vulnerability to subsequently presented misinformation. This investigation was designed to extend the understanding of RES in two ways: by examining the presence of RES in younger and older preschoolers, and by determining whether participation in an initial interview prior to misinformation exposure interferes with source discrimination. At each age level, half of the children received inconsistent information about a previously observed event after participating in an initial cued recall test, whereas the remaining children were not tested prior to exposure to the inconsistent information. All participants were administered a delayed cued-recall test followed by an age-appropriate source discrimination task. As expected, older children were significantly better able to recall the original event than the younger children were; however, RES was not observed at either age level. The post-event narrative through which the misinformation was conveyed also included items that were consistent with the to-be-remembered event, and children at both age levels were better able to recall the consistently reported information than either the original items that were omitted from the narrative or presented in an altered form. Although the expected main effect of age on source discrimination was not observed, a 3-way interaction indicated that the initial test had different effects on source discrimination for misleading items at the two age levels. Younger preschoolers in an initial test condition provided significantly fewer correct source discriminations for misleading items than younger preschoolers in a delay-only condition.

The pattern was reversed for older preschoolers. Together, these results suggest that initial testing prior to exposure to misinformation may not have the same effects on suggestibility in preschoolers as in adults, although further research is necessary to explain the influence of initial testing on source discrimination in preschool children.

Retrieval-Enhanced Suggestibility: A Developmental Examination

by
Angela C. Meluso

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Psychology

Raleigh, North Carolina

2011

APPROVED BY:

Lynne Baker-Ward, Ph.D. Committee Chair	Jason Allaire, Ph.D.
Christopher Mayhorn, Ph.D.	Joan Pennell, Ph.D.

Thomas Hess, Ph.D.

DEDICATION

This dissertation is dedicated to my family. Grandma and Grandpa, thank you for making it possible for me to attend college in the first place, so many years ago. Without you, this dissertation would never have come to fruition. You are my angels. Mom and Bruce, thank you for your constant support and for giving me the backbone of strength upon which I have leaned continuously throughout the years, even when you did not realize it. Dad, thank you for all of your constant love and support. Julia, I could not have survived the distance and struggles of these past several years without you. You have brought me happiness in the hardest of times. I love you sister. Carla, the strength that you have demonstrated throughout these past several years has been inspiring and has given me the strength to persevere, despite challenges in life's darkest moments. I love you, little sister. Bellino, my handsome little man, I would be lost without you. You've been there with me through it all. Anthony, you have changed me for the better. You have given me the motivation to live in the moment while also remembering that greater moments have yet to come. I can't wait for the next step of my life, knowing that I get to take it with you by my side. Finally, to the rest of my family, you have all inspired me to pursue my goals in your own ways and for that, I am forever indebted. I love you all.

BIOGRAPHY

Angela Meluso attended Camden County Community College in New Jersey, graduating with an Associate degree in 2004. She then went on to attend Rutgers University in Camden for the completion of her Bachelor of Arts degree in 2006 and her Master of Arts degree in 2008, both in the discipline of Psychology. Angela's research at North Carolina State University has centered on eyewitness event memory and the ability to recall events accurately at various points in development. Angela currently resides in North Carolina and plans to relocate to the Northeast with her fiancé, Anthony, upon graduation.

ACKNOWLEDGEMENTS

Thank you to my dissertation committee members: Tom Hess, Shevaun Neupert, Chris Mayhorn and Joan Pennell for your guidance and support throughout my doctoral candidacy. Thank you to my previous advisor, Karen Thierry, who guided me throughout my Master's level education. Without your guidance, I likely would not have pursued a doctoral degree. You showed me that it is possible to work hard and achieve your goals, providing the strongest of examples through your own accomplishments. Finally, the sincerest of thanks are extended to my advisor and friend, Lynne Baker-Ward. If it were not for you, I would not have made it to this point. Throughout the trials and tribulations of graduate school, from initial acceptance into the program here at NC State, through the hardships, both personal and educational, you have been a constant source of strength and inspiration. I owe this dissertation to you.

Special thanks are also extended to the undergraduate students who helped me throughout all stages of my dissertation data collection: Kylie Paulina, Victoria Piccicuto and Cassia Ehrenbock. I could not have conducted this study without you. Thank you also to the M&ND members, both past and present, for your encouragement and support throughout the years. Without each other, we'd all be lost. Special thanks are extended to my friend and past colleague, Mary Styers. You've guided and supported me in so many ways throughout this process and I will be forever grateful. Finally, I would like to thank the department of Psychology at NC State University for providing funding in partial support of this project through the John Oliver Cook Dissertation Award, and especially for providing the educational background that ultimately lay the foundation for this project and for my career.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Retrieval-Enhanced Suggestibility: A Developmental Examination

A robust benefit of initial and subsequent testing on the retention of verbal material has consistently been found (e.g., Roediger & Karpicke, 2006). In an effort to identify methods of decreasing eyewitness suggestibility in forensic situations, researchers recently examined the possibility that this "testing effect" could also facilitate eyewitness memory (Chan, Thomas & Bulevich, 2009). Contrary to expectations, however, recent studies have revealed that this benefit does not extend to eyewitness memory when misinformation about an eyewitness event is suggested to adults during a retention interval between tests (Chan et al., 2009). In fact, initial testing under such conditions has actually been found to lead individuals to be *more* as opposed to *less* suggestible (e.g., Chan et al., 2009; Thomas, Bulevich, & Chan, 2010). In other words, initial testing prior to misinformation exposure increases the likelihood that individuals will report suggested information from a post-event source as having come from the original source (i.e., witnessed event). This phenomenon was originally labeled the "reversed-testing effect" and has most recently been termed as "retrieval-enhanced suggestibility" or "RES" (Thomas et al., 2010).

To this point, RES has only been examined among adult participants. As will be discussed below, the examination of RES among young, preschool children is important for addressing issues in eyewitness testimony. This study sought to contribute to this literature by examining the following research questions:

1. How does an initial interview affect preschool-aged (i.e., 3- to 5-year-old) children's subsequent recall and suggestibility? Specifically, do preschool children demonstrate retrieval-enhanced suggestibility, as do adults?

2. Does initial recall, prior to misinformation exposure, have an effect on preschool children's ability to discriminate the sources of their memories and if so, do older and younger preschoolers differ in terms of their ability to monitor the sources of their original and post-event memories as a result of such effects?

Background and Significance

When involved in situations as eyewitnesses, individuals of all ages must often later recall the events that they either witnessed or directly experienced after extended delays following those events, thereby making it particularly important that accurate information is retained in memory over time. Memory for the details of such events is unfortunately increasingly vulnerable to decay as time elapses between encoding and retrieval (e.g., Brainerd & Reyna, 1998). To further compound the effects of time on memory retention, exposure to post-event information from a variety of sources (e. g., other people, imagination) is inevitable and can have profound effects on reports of experienced events, especially if the individual does not efficiently monitor the sources from which this information is obtained (Johnson, Hashtroudi, & Lindsay, 1993). Post-event information may or may not be accurate, in that it may be consistent with what an individual remembers, may replace information that had never initially been encoded or had since been forgotten, or may be inconsistent with or different from what happened and from what one remembers as having happened (Baker-Ward, Ornstein, & Starnes, 2009). Exposure to post-event information becomes problematic when individuals incorporate inaccurate or suggested misinformation into their memory reports, and the possibility of such suggestibility becomes increasingly likely as time elapses between an event and recall (Brainerd & Reyna, 1998).

Misinformation effects are often thought to occur unintentionally, mainly because individuals have difficulty discriminating among information that they witnessed during an original event and information that they believe may have happened but that had not actually been encoded at the time of the event (e.g., information from a post-event source, imagination, etc.) (Garry, Loftus, & Brown, 1994). Since misinformation reports often reflect the witness' beliefs about what had happened rather than deliberate misrepresentations of the facts, it is imperative that the factors that contribute to the likelihood of reports of misinformation from other sources are as completely understood as possible.

One method that has been identified as useful in facilitating memory for encoded details over time is early memory recall, also referred to commonly as initial testing or initial recall (e.g., Melnyk & Bruck, 2004; Powell & Thompson, 1997; Warren & Lane, 1995). Such recall at the outset of a given retention interval is thought to insulate or preserve memory. In addition, initial recall increases the number of source cues available for future recall attempts because memories are reconsolidated according to a variety of contexts (Chan et al., 2009). To elaborate, initial recall insulates original memories and increases the likelihood that information about an event will remain preserved and available, as well as more strongly interconnected for recall at a later time (Brainerd & Ornstein, 1991; Roediger & Karpicke, 2006). Through reconsolidation, initial recall is thought to benefit the long-term retention of details of an experience because the to-be-recalled information, encoded in a variety of contexts, is consequently associated with additional source information that can serve as cues during later retrieval (Thomas et al., 2010). In other words, individuals asked to recall information about an event that they had previously recalled will have information

from the *original encoding context* and from the time they were asked *to first recall* that information. Consequently, they are able to think back to either or both of those recall attempts (i.e., sources) to retrieve information, thereby increasing the likelihood of remembering in comparison to an alternative circumstance in which information from only one source (i.e., the original encoding context) is available during recall. As mentioned earlier, the overall benefit of initial testing at the outset of a retention interval on subsequent recall has been referred to as the testing effect and has been found consistently in many studies with adults and children, which are generally reported in the verbal learning literature (Roediger & Karpicke, 2006).

Researchers interested in eyewitness memory have recently begun to make connections among event memory retention over time, suggestibility and the testing effect with adult samples (Chan et al., 2009). Chan and colleagues (2009) were the first to examine if conducting an initial recall test could serve as a potential method to increase resistance to misinformation, assessing whether the testing effect could be extended to circumstances in which false, post-event information is encountered between tests. Based on the testing effect, initial testing was expected to preserve memory over time, consequently increasing the likelihood that individuals would resist reporting false information as part of their original memories on a delayed test. Contrary to this prediction, however, Chan et al. (2009) found a reversal of the original testing effect, in that participants tested prior to misinformation exposure demonstrated *increased* suggestibility in comparison to those who did not receive an initial test. This counterintuitive phenomenon, now referred to as retrieval-enhanced

suggestibility (RES), has been replicated in several studies (e.g., Chan & Langley, 2010; Thomas, et al., 2010).

RES may be especially problematic in real world, eyewitness contexts in which witnesses and victims initially provide event reports (e.g., to first responders, police officers, the media) and are then asked to recount those same events in reports on subsequent occasions (Thomas et al., 2010). Because witnesses are often exposed to conflicting information between reports (e. g., from other witnesses, discussion of events with other non-witnesses, poor interviewing practices, etc.), it is of practical importance to understand how initial reporting affects susceptibility to subsequent exposure to inconsistent information and later event reports. Importantly, as the above-mentioned research questions indicated, little is known at this time regarding the potential existence of the retrieval-enhanced suggestibility phenomenon in early childhood.

Retrieval-Enhanced Suggestibility and Children

Although children must often provide reports of events in forensic contexts, there are no studies to date that have examined whether RES is a phenomenon that pertains to children as well as adults. In studies of suggestibility, however, preschool-age children have been consistently found to be more suggestible than adults in terms of event memory in general, although suggestibility decreases across the preschool and early elementary years (3- to 7-years-old). This decrease is thought to be largely a product of increases in source-monitoring ability across this period of cognitive development (Bright-Paul, Jarrold, & Wright, 2005; Ceci & Bruck, 1993).

With regard to the effects of repeated recall on the accuracy of children's event reports over time, findings have been mixed. Some studies have indicated that repeated recall can have detrimental effects, in that preschool-aged children may change their responses to interview questions and may experience memory alterations if an original memory trace is weak (e. g, Memon & Vartoukian, 1996). In contrast, other investigations report benefits of repeated interviewing, in that early interviews may help children to organize their memories and retain information over time (Melnyk & Bruck, 2004; Tizzard-Drover & Peterson, 2004). Despite the extensive literature examining the effects of repeated recall on children's memories (for review, see LaRooy, Lamb, & Pipe, 2008), it is nevertheless difficult to make comparisons across studies due to differences in designs in factors such as the age groups included, the timing of interviews in reference to the target event and to other interviews, the presence or absence of suggestive interviewing techniques, and other variables. Early, neutral (i.e., non-suggestive) interviews (i.e., within the 1-week time frame that is generally associated with consolidation of memory for an experience in early childhood) have also been found as beneficial to children's memory over time (Melnyk & Bruck, 2004). This facilitation is typically explained as the result of better consolidation resulting from retrieval early in the retention interval. Additionally, early retention and the accompanying enhanced memory consolidation is also thought to provide "inoculation against forgetting" (Brainerd & Ornstein, 1991; LaRooy et al., 2008; Melnyk & Bruck, 2004; Powell & Thompson, 1997; Warren & Lane, 1995).

In contrast, the results of some studies instead have indicated that initial, neutral interviewing is not always beneficial to increasing the amount or the accuracy of preschool-

aged children's memory reports over time (e.g., Baker-Ward, Gordon, Ornstein, Larus, & Clubb, 1993), sometimes even leading to decreases in accuracy (see LaRooy et al., 2008 for review). Initial *suggestive* interviews (e.g., consisting of leading questions or interviewer suggestions) have also been found to have inconsistent effects on recall, most often decreasing the amount and accuracy of information presented over time in preschool and kindergarten-aged children (e.g., LaRooy et al., 2008; Melnyk & Bruck, 2004). To further complicate matters, the repetition of questions within an interview (or even within an experimental session) can cause young children to assume that their initial responses were incorrect and hence to change their answers (Memon & Vartoukian, 1996).

It has remained unclear as to whether an examination of misinformation effects in tandem with initial and repeated interviews among preschool children would reveal the same counterintuitive phenomenon of retrieval-enhanced suggestibility as has been found with adults (Chan et al., 2009). Some researchers have proposed that misinformation acceptance is more automatic and as a result, more common and less preventable among younger in comparison to older children (i.e., 7-years-old and up) and adults (Holliday & Hayes, 2000). Due to the fact that preschool children are generally found to be more suggestible and consequently more likely to incorporate information from another source into their memory reports than older children and adults, it is possible that initial testing will be equally or even more detrimental to the accuracy of event reports with preschoolers than with adults. One possible reason why this result could occur is that young children often incorporate suggested information into their reports because of adult examiners' authority (Memon & Vartoukian, 1996). When information provided by an adult contradicts children's own event memory,

young children tend to acquiesce to suggestions and provide misinformation in their reports of original events because they often think that they must be wrong and the adult must be right (Beuscher & Roebers, 2005; LaRooy, et al., 2008; Memon & Vartoukian, 1996). Thus, RES may be especially likely for preschoolers because they oftentimes change their answers when questioned multiple times. It should be noted, however, that false reports in such situations are not necessarily indicative of actual changes in memory, but rather represent compliance with the adult interviewer's suggestions (Beuscher & Roebers, 2005).

Initial Interviews, Misinformation, and Source Monitoring

Suggestibility effects depend upon the strength of the memory traces for both the original and suggested information sources (Gobbo, 2000). The source-monitoring framework proposes that memories of experienced or witnessed events should consist of a good deal of perceptual and contextual source detail, whereas memories of imagined events or those that are merely heard about or suggested from another source should consist mainly of cognitive operations and processes that were engaged in during encoding (Johnson et al., 1993). Inefficient monitoring or evaluation of these characteristics during retrieval allows memories of misinformation to pass as memories of original event information (Echterhoff, Hirst ,& Hussy, 2005; Johnson et al., 1993). Even if an individual still has intact memory for both the original and misinformation, misattributing event details to the wrong source would result in misinformation effects.

Developmental research has established that the ability to monitor source develops across the preschool years (e.g., Drummey & Newcombe, 2002). Laboratory studies examining episodic memories with children have indicated that preschoolers (i.e., from 3-

and up through 6-years of age) are more likely than older children and adults to confuse memory sources, such as confusing events that they heard about with things that they saw (e.g., Thierry & Spence, 2002; Thierry, Spence, & Memon, 2001). Three- to 4-year-old children in particular have the most difficulty identifying and discriminating memory sources and often perform at only chance levels, thereby indicating they may have little or no awareness of discriminating memory features (e.g., Drummey & Newcombe, 2002). Preschool-aged children do in fact experience forgetting more rapidly than older children and adults, leaving them with weak or even non-existent source-specific information about an event over a short period of time (e.g., Brainerd & Reyna, 1998). This loss of source information over time increases the likelihood that young children will report suggested information because it is stronger than their own memory or because it fills gaps where their memory for specific information about an experienced event is lacking (Brainerd & Reyna, 1998). It should be noted, however, that it is unlikely that this effect will be observed in situations in which the original and suggested information are presented within the same experimental session because information loss during such a brief interval would be unexpected, even among very young children.

The development of source-monitoring ability across the preschool years could possibly shed light on the underlying mechanisms of RES. It has been found that testing prior to the presentation of misinformation about an event increases source-monitoring difficulty among adults (see Thomas et al., 2010). Meanwhile, recent studies with adults have indicated that warnings implying that memory sources should be carefully monitored have been useful in eliminating the detrimental relationship between initial testing and

suggestibility defined as RES. These findings suggest that RES may be driven by a cognitive mechanism, such as inefficient source monitoring, in adulthood. In terms of RES (i.e., increasing misinformation effects) within one experimental session with preschool children, suggestibility *may not* be enhanced by retrieval because preschoolers are already quite likely to report suggested information, regardless of whether tested initially or only once following a delay (e.g., Bright-Paul et al., 2005). From a developmental standpoint, if young children do demonstrate high levels of suggestibility under similar conditions as found in RES studies with adults, the mechanisms behind this effect may be different. Meanwhile, before the mechanisms can be examined, it must first be determined whether this same, counterintuitive phenomenon is present with young children as it is with adults. An examination of the effects of initial tests on preschoolers' source-monitoring ability is the first step in pursuing an understanding of these mechanisms for future research.

The study reported here involved 3-to 5-year-old children, allowing for an examination of the developmental shifts in suggestibility and source-monitoring capability where they have most consistently been found, between 3-years-of-age and prior to elementary school age (i.e., age 6) (e.g., Melnyk & Bruck, 2004; Thierry & Spence, 2002). More specifically, the children were categorized as younger preschoolers (3.00- to 4.49-year-olds) and older preschoolers (i.e., 4.50- to 5.99-year-olds) in line with previous research that have provided evidence for increases in source-monitoring abilities between these age groups (e.g., Karpinski & Scullin, 2009). The methodology replicated that used in previous studies with adults (see Chan et al., 2009) but instead incorporated age-appropriate materials. This research sought to examine whether preschool children, like adults, demonstrate retrieval-

enhanced suggestibility. Further, this study investigated whether younger (3.00- to 4.49-years-old) and older preschool children (4.50 to 5.99-years-old) who receive an initial test before misinformation exposure maintain explicit source information for an original event as adults do, in addition to an examination of the effects of an initial interview on source-monitoring accuracy. More specifically, this study examined whether preschoolers were more or less able to accurately remember event details and to accurately discriminate the sources of original and post-event information when they received an initial test, in comparison to when they did not receive an initial test prior, prior to exposure to post-event misinformation.

Hypotheses

This research examined the following hypotheses:

1) The presentation of an initial interview will result in higher levels of overall correct recall. In other words, a testing effect would be found (i.e., increased memory of event-consistent and control information) in groups of younger (3.00- to 4.49-year-old) and older (4.50-to 5.99-year-old) preschool children who received an initial test (initial test condition) in comparison to peers who were not administered an initial test (delay-only test condition).
2) An Age Group x Test Condition interaction will be observed in responses to misleading items in preschool aged children. Older preschoolers in the initial test condition are expected to demonstrate RES, as evidenced by increased misinformation responses to misleading Items in comparison to those in the delay-only test condition. Meanwhile, a different pattern is predicted with younger

preschoolers. Following from the suggestibility literature, younger preschoolers in general are expected to accept high amounts of misinformation (i.e., regardless of test condition), as a function of other developmental and social factors not specifically examined in this investigation but commonly found in studies with children of this (i.e., errors of acquiescence, changing answers, etc.). Accordingly, younger preschoolers in the initial test condition *may not* demonstrate *increased suggestibility* associated with the initial test in comparison to those in the delay-only test condition.

3) An Age Group x Test Condition interaction was predicted in terms of source-monitoring performance. Specifically, in addition to a main effect of age (i.e., older preschoolers are expected to provide more accurate source discriminations), it was predicted that the older preschoolers in the initial test condition would be better able to discriminate the sources of event information across item type, in comparison to those in the delay-only test condition. Test condition was not expected to affect performance of the children in the younger age group. This pattern was predicted because the initial test should provide additional source cues that may be used when asked to explicitly monitor the sources of event memories. Only older, but not younger preschoolers were expected to be developmentally able to use such source cues, due to the limitations in source-monitoring ability that children in the younger end of the preschool age group inherently encounter.

Method

Participants and Design

All children were enrolled in one of five recruited preschool facilities in the suburban communities surrounding Raleigh, North Carolina. Of these five preschools, three were operated by the same parent company. All were similar in terms of academic environments, curricula and available resources and each ensure that their teachers meet or exceed the state requirements to work in early childhood. More specifically, lead teachers at these schools hold a minimum of a Child Development Associate credential, with additional education levels ranging through 2- and 4-year degrees with multiple years of preschool experience, whereas assistant teachers each have relevant experience and are typically working towards degrees. Also, all teachers at these schools complete detailed training according to state-specific rating requirements during their first week of employment.

A total number of 69, 3- to 5-year-old preschool children were recruited for the present study. Because one child did not assent to participation, the final sample consisted of 68 children. None of the participants had begun kindergarten at the time of their participation. As confirmed by their teachers, all children were native speakers of English. Although information regarding parental education and family income was not obtained, based on the characteristics of their communities and their preschools, the majority of the children can be considered to be from middle-class backgrounds. The sample was predominantly European American (82.4%), with the following ethnic backgrounds also represented: Asian American (8.8%), Black or African American (4.4%), and Latina/Latino (1.5%). Ethnic background was not specified for 3% of the sample.

Sample sizes were estimated on the basis of a G*Power3^{TM} power analysis, indicating that at least 52 children were needed in total to adequately examine interactions with a moderate eta squared effect size of .30. Given the 2 (Age Group: younger, 3.00- to 4.49-years; older, 4.50- to 5.99-years) x 2 (Test Condition: delay-only, initial test) design, a minimum of 13 children was needed for each of the cells of the design. The order of test conditions was randomly assigned by first determining the number sessions with each age group that would take place that day (based upon attendance) and then by flipping a coin. Upon arrival at each preschool, the names of children in attendance that day whose parent(s) provided consent were obtained from the teachers. Within each age group, children were also randomly assigned by flipping a coin to one of the two test conditions. An effort was made to include all children in attendance on the day of data collection whose parents granted consent. This resulted in a larger number of younger (n = 41) than older preschoolers (n = 27) being represented in the sample due to a higher response rate from parents of younger children. The number of children assigned to the delay-only and initial test conditions were 19 and 22 respectively, at the younger age level, and 14 and 13 at the older age level. Table 1 presents the number of male and female children in each of the test conditions within each age group.

The Principal Investigator and three trained undergraduate research assistants took part in each session of data collection. The research assistants and the Principal Investigator were all female. Three researchers were European American and one was Asian American. The research assistants were advanced undergraduate Psychology majors at the time of data collection, which took place during the summer months prior to their senior year. As

described below, three of the four researchers administered the recall test during data collection and at least two researchers interacted with each participant.

Materials

To-be-remembered video. The target, to-be-remembered stimuli have been used in several previous studies of preschool-aged children's event and source memory (e.g., Poole & Lindsay, 1995; Thierry & Pipe, 2009; Thierry & Spence, 2002). The target event was presented via a video called "Mrs. Science's Experiments," which consists of six science experiments performed by an actor, referred to as "Mrs. Science" (See Appendix A). For instance, in one experiment, children watch Mrs. Science make a red balloon collect static electricity. Mrs. Science uses the static electricity from the balloon to pick up pieces of blue paper. Each experiment lasts no more than a 2- to 3-minutes.

Storybook. A storybook narrative recap of the events, which had been used in prior studies, was slightly modified for the present investigation by replacing some details with plausible alternatives (misinformation). In particular, the storybook recapped the information depicted on the video with descriptions of some details consistent with those shown and some replaced with plausible alternatives. A consistent item on the final test was a question that asked about a detail that was described in the storybook as it appeared in the video; for example, a red balloon was used in an experiment in the video and the storybook included the statement that, "Mrs. Science used a *red* balloon". A misleading item on the final test was a question that asked about a detail that was presented as misinformation in the storybook (i.e. in contradiction to the depiction on the video), such as: "Mrs. Science used a *blue* balloon". Control items on the final test were questions that asked about items not

specifically mentioned in the storybook and were therefore only encountered while viewing the video (see Appendix B for all items).

Initial cued-recall test. The initial cued-recall test assessed the memory of the children in the initial test condition for the information depicted in the video prior to a delay. The test consisted of 15 open-ended questions (Appendix B) about the video, printed individually on index cards. Prior to each interview, the questions were shuffled to avoid the possibility of confounds between specific items and order of presentation. Children were also required to indicate their confidence in each particular response (except "don't know responses") on an age-appropriate Likert scale represented by four squares on a piece of paper that increased consecutively in size (Wrzesien & Raya, 2010) (Appendix C). Children indicated their level of confidence by pointing to the corresponding square on the scale for each item. Interviewers first explained the procedure to the children by stating, that the little box meant that they were "not very sure", the next larger box meant that they were "a little sure", the second largest box meant that they were "pretty sure" and the largest box meant that they were "really sure". Thus, after each cued-recall response, children were asked, "Can you point to how sure you are?" and were directed to the scale in front of them. Each child practiced this at least twice until the interviewer concluded that the child understood the procedure.

Final cued-recall and source-monitoring test. All children received an experimenter-administered cued-recall/source-monitoring test that probed for information that was presented consistently and inconsistently (i.e., misinformation) in the recap, as well as control information, which was not mentioned in the recap. This assessment was the same

cued-recall test with the same questions (Appendix B) as the one that the children in the initial test condition had previously taken, with the addition of a source-monitoring probe following each item. Consistent with the procedure used in administering the initial test, the source-monitoring questions were printed individually on index cards and shuffled prior to each interview. Following each item response (except "don't know" responses), children were asked to indicate their confidence in their responses on the same age-appropriate Likert scale (Wrzesian & Raya, 2010) as used with the initial test.

After they rated the confidence of their answer to a cued-recall question, children were asked to discriminate the source of the information referenced in the question (i.e., video, storybook, both, or neither) via a source-box procedure (see Bright-Paul et al., 2005). To do so, children placed the index card for each question in one of four boxes that were labeled with the following color photographs of the source options: the video on a screen, the storybook, the video on a screen and the storybook together, and a blank paper representing a "neither" response. Children were instructed to place each question card in the box that corresponded to the source(s) of information that they had encountered pertaining to that question. This procedure was adopted because very few children within this preschool age group were expected to have the mature reading and cognitive skills necessary to use written or verbal labels, as opposed to the pictorial labels utilized in this study that corresponded to the four source options. Each child was given at least two practice trials in discriminating source, with additional practice provided as necessary until the interviewer concluded that the child understood the source-box procedure.

Procedure

Preschool and daycare facilities located within a 20-mile radius of North Carolina State University were identified through an online search of local schools in the area that enrolled children in the target age range during the summer, as well as the academic year. Phone calls were made by the Principal Investigator to determine interest in participation and to schedule data collection dates. The Principal Investigator then provided the scheduled schools with supplemental materials, including a written study description, consent forms, a consent form for the school and individual consent forms to be signed by the parent or guardian of each child in the targeted age group. Preschool directors were instructed to distribute consent forms to all children who fell within the specified age ranges (i.e., 3- to 5-years-old). All target-aged children whose parents or guardians provided consent were eligible to participate in the proposed study. On the date of data collection, teachers were instructed to confidentially indicate to a research assistant the names of students whose data they thought should be excluded from the analyses on the basis of their knowledge of those students' characteristics (i.e., those who had significant learning delays, those who did not speak English as their first language, etc.). Importantly, teachers were instructed NOT to indicate the reasons why they felt that a particular child's data should be excluded. The teachers did not identify any children whose data should be excluded.

Data collection. All participants took part in identical procedures with the exception of the initial test/interview, as described below, and all children observed the same to-be-remembered video. The procedures transpired in quiet rooms or secluded areas at each of the five participating schools. Participants were approached in their classroom by a research

assistant and were asked to give assent to participate in a few activities. Only one child declined to give their assent and was not pressured to participate. Children participated in groups of three, although on a very few occasions, the testing group consisted of two rather than three children.

Upon entering the room, children were seated in front of a widescreen laptop and were instructed as follows:

> Research Assistant 1: *"For the first part of our time together, you are going to watch a short video called 'Mrs. Science's Experiments'. Now, I want you to pay attention to the video and watch it really carefully, even more carefully than you would at home. Try to pay attention to what Mrs. Science does and uses in her experiments. You are going to be asked some questions about what you saw in the video later on."*

All children then viewed the video. Although levels of engagement were generally good, one of the research assistants monitored the children's behavior while they viewed the video and reminded them to watch the video closely when they appeared distracted.

Test condition manipulations. After watching the video, participants in the initial test conditions received a cued-recall test administered by a researcher with whom they had not previously interacted. After the child's additional assent for the next phase of the experiment was obtained, one of the three research assistants escorted the child to a separate area of the facility for the administration of the initial cued-recall interview, as described above. The children did not receive feedback regarding the accuracy of their answers; however, they were given positive encouragement that was not contingent upon their responses throughout the interview (e.g., "You are doing a really great job."). Upon completion of the initial interview, each of the children was given a sticker. Children in the delay-only conditions participated in an age-appropriate block-balancing game (Jenga^{TM}) for

5-minutes, a time determined through piloting to be approximately equal to the time that it took children in the initial test condition to complete the initial cued-recall test.

Delay interval. All children then moved to a separate area of the school or classroom to play Chutes and Ladders™, a simple board game that involves counting and following instructions (Ramani & Siegler, 2008) for the approximately 5 minutes that served as the "delay" period.

Storybook recap. Next, all children were seated and were told by Research Assistant 1 that they were about to hear a recap of the video that they watched earlier, as follows:

> Research Assistant 1: *"Now I'm going to read you a story that repeats what you saw in the video you watched called, 'Mrs. Science's Experiments'. I want you to listen carefully to the story and pay attention to what the story says Mrs. Science does and uses in her experiments."*

Research Assistant 1 then read the narrative recap of the video to the students from a storybook. The storybook included information that was consistent with the video, as well as information that was replaced with plausible alternatives (i.e., misinformation) in the narrative recap.

Final test. Immediately following the presentation of the recap, each child was approached individually by an experimenter with whom the child had not been previously interviewed and was asked for his or her assent to answer some questions about the video. For children in the initial test condition, a different research assistant administered the initial and final tests. Only general encouragement was again provided. Upon completion of the final interview, the children were given stickers and were debriefed in an age-appropriate manner, prior to being escorted back to their classrooms. Specifically, they were told that

they may have heard some things in the story that were wrong, that the interviewer just wanted to know what they remembered and that any answers they had provided were good answers. Children were also asked not to talk about the experiment with their friends until everyone had a turn to participate.

Data Coding

The children's responses were recorded manually on prepared data sheets for both the initial (Appendix D) and final test (Appendix E) at the time the assessments were administered; hence, no audio or video recordings were made. Interviewers were instructed to write children's cued-recall responses as accurately as possible. Interviewers recorded source discriminations by circling one of the four source options on the interview sheet, and confidence ratings were recorded numerically on a scale of 1 (smallest box) to 4 (largest box). The scoring was objective; therefore reliability was not obtained. However, because the children's immature language in some cases made it difficult to discern response content and accuracy, each interview was coded by one research assistant and then double-checked by another, with discrepancies resolved by the Principal Investigator. Also, performance on the initial and final tests was determined by calculating proportion scores. To calculate these scores, the total number of responses provided (excluding "don't know" responses) was included as the denominator in each calculation, as opposed to including the total number of potentially answerable test items as the denominator. This was done in an effort to extend generalizability to forensically relevant circumstances where a "don't know" response is considered as qualitatively different from a response. In such circumstances, a response could potentially lead to a true or false allegation, whereas a "don't know" would not.

Initial cued-recall test. Responses to the initial cued-recall test were not the main focus of the reported study, but were coded as follows. First, the overall number of responses children provided to items on the initial test was calculated (i.e., the number of times children provided any response, excluding "don't know"). Answers were then coded as correct when the participant provided the targeted correct detail according to the video and as incorrect when the participant reported a detail inconsistent with the video. "Don't know" responses were also scored. A total score was tallied for each of the response categories: correct, incorrect and don't know. Proportions of correct responses were calculated to analyze the data as opposed to using raw scores in the analyses because children did not always respond to all items. Proportions of correct responses were calculated by dividing the total number of correct responses by the total number of responses overall. "Don't know" responses were then calculated by dividing the total number of "don't know" responses that children provided by the total number of items on the cued-recall test (i.e., 15 items). Because the children were not exposed to the recap before the initial test, items were not categorized by item type.

Final cued-recall test. There were three item types in the final test, with each corresponding to way that the detail was presented in the storybook narrative recap. *Consistent items* were details that were the same in the video and the narrative; *misleading items* were details that were changed from the video to narrative; and *control items* were details specified in the video but were omitted from the narrative. The total number of responses that each child provided on the cued-recall portion of the final test for each item

type was calculated. The overall total number of "don't know" responses that each child provided was also calculated.

To determine accuracy of responses, all cued-recall responses on the final test were coded as follows: correct, misinformation (provision of the inconsistent detail), incorrect (an idiosyncratic error) or "don't know". Therefore, a total was tallied for: 1) correct responses to consistent, misleading and control items, 2) misinformation responses to misleading items, 3) incorrect responses to consistent, misleading and control items, and also 4) "don't know" responses to consistent, misleading and control items. Proportions of correct responses for each item type were calculated by dividing the number of correct cued-recall responses that children provided for each item type by the number of responses that children provided to each respective item type on the final test. Proportions of misinformation responses were calculated by dividing the total number of misinformation responses that children provided by the number of responses children provided overall to misleading items. Proportions of "don't know" responses were calculated by dividing the number of "don't know" responses that children provided by the total number of items on the cued-recall test (i.e., 15 items). Proportions of "don't know" responses were calculated for each item type as well, by dividing the number of "don't know" responses for each item type by the total number of items representing each item type (i.e., 5) on the final cued-recall test.

Source discrimination. Note that because the final test consisted of cued-recall items as opposed to forced-choice recognition (i.e., yes/no) items, children were asked to discriminate source after providing ANY response other than "don't know" (regardless of response accuracy). Consequently, it was in fact possible for children to have provided a

misinformation or incorrect response and still have the opportunity to discriminate the source(s) of information about that particular question to the correct source. Note, however, that source discrimination was considered independently of correct cued-recall. Therefore, source discriminations were examined for each cued-recall item to which children provided a response. To score each source discrimination response, it was determined whether the child identified the supplying source(s) of information about that particular question correctly. Totals were calculated that represented the number of times each child provided the correct source(s) for each of the three item types. Total source discrimination scores for each item type were calculated for each child by dividing the total number of correct source attributions made to each item type by the number of responses provided for each respective item type.

Results

Preliminary analyses were conducted to address the possibility that children differed across age groups and test conditions in terms of gender and ethnicity distribution. An initial chi square analysis revealed that despite the higher number of younger versus older preschoolers, the distribution of younger and older preschoolers did not differ significantly across delay-only (younger: $n = 19$, older: $n = 14$) and initial test conditions (younger: $n = 22$, older: $n = 13$), $X^2 (1) = .20$, $p = .66$. Ethnicity distributions were independent of age, $X^2(5) = 9.30$, $p = .10$, and test conditions, $X^2 = 3.95$, $p = .56$.

To insure that children in the different test conditions within each age group were equivalent with regard to months in age, t-tests were conducted. No differences were found within the younger age group between the delay-only ($M = 45.84$, $SD = 5.03$) and initial test conditions ($M = 46.59$, $SD = 3.56$) in months in age, $t(39) = -.56$, $p = .58$, CI 95% [-3.47,

1.97]. Similarly, within the older age group, an independent samples t-test indicated that children did not differ across delay-only (M = 60.51, SD = 5.24) and initial test conditions (M = 63.71, SD = 5.64), $t(25) = -1.53$, $p = .14$, CI 95% [-7.51, 1.11] in months in age. In addition, preliminary regression analyses conducted to examine whether months in age was correlated within age group with any of the outcome variables were non-significant.

Initial Test Performance

No hypotheses were based on performance on the initial test. However, to examine whether expected age group differences were present in recall at the initial test, an independent samples t-test was performed to examine age group differences in the proportion of correct responses among children in the initial test condition only. Levene's test for equality of variances was non-significant, $F(33) = .18$, $p = .68$, and no significant differences between age groups were observed in the proportion of correct responses provided by children on the initial test, $t(33) = -1.35$, $p = .19$, CI 95% [-.28, .06], although the means were in the expected direction (Table 2). Further analyses were conducted to examine whether there were age group differences in the proportion of "don't know" responses on the initial test. Levene's test for equality of variance was again found to be non-significant, $F(33) = 2.17$, $p = .15$, and no significant age differences were present, $t(33) = 1.33$, $p = .19$, CI 95% [-.05, .26] (Table 2).

Final Test Performance

Several analyses regarding performance on the final test revealed violations of equality of variance and/or sphericity, as reported below. Arcsine transformations were calculated to adjust for these violations where appropriate; however the transformations

exacerbated the existing violations that were present using untransformed data. Therefore, all results are reported using untransformed data.

Cued-recall performance. To test Hypothesis 1 (see p. 11), the expectation that the presentation of an initial test will result in higher overall levels of correct recall responses on the final cued-recall test, a mixed ANOVA was used to examine age group and test condition differences in correct responses to each item type on the final cued-recall test. Specifically, a 2 (Age Group: younger, older) x 2 (Test Condition: delay-only, initial) x 3 (Item Type: consistent, misleading, control) mixed ANOVA was used to examine differences across groups in terms of the proportion of correct responses on the final cued-recall test, with test condition and age group as between-subjects factors and item type as a within-subjects factor. Levene's test of equality of variances was non-significant across item types (consistent: $p = .44$, misleading: $p = .15$, control: $p = .28$) and Box's M was found to be non-significant as well, $p = .91$. Meanwhile, Mauchly's test of sphericity was significant, $p = .02$; therefore, Greenhouse-Geisser epsilons are reported accordingly when relevant. The expected main effect of age group was revealed across test conditions and item type, with older preschoolers providing more correct responses to all item types (i.e., consistent, misleading, control), $F(1, 64) = 9.42$, $p = .00$, partial eta squared = .13 (Table 3). The mixed ANOVA failed to reveal a main effect of test condition for correct responses, as had originally been predicted, $F(1, 64) = .28$, $p = .60$, partial eta squared = .01 (Table 3).

A Test Condition x Item Type interaction was also predicted, in that initial test participants were expected to provide significantly more correct responses to consistent items than to control items (i.e., a partial testing effect), whereas delay-only participants were not.

In addition, a three-way, Age Group x Test Condition x Item Type interaction was also possible, with older preschoolers only in the initial test condition possibly demonstrating a testing effect for control items, as well as for consistent items. This pattern was not predicted for the younger preschoolers. The mixed ANOVA revealed neither the above-mentioned Test Condition x Item Type interaction, $F(1.79, 114.51) = .15$ $p = .84$, partial eta squared = .00, nor the 3-way Age Group x Test Condition x Item Type interaction, $F(1.79, 114.51) = .60$, $p = .53$, partial eta squared = .01. Meanwhile, a main effect of item type was revealed, $F(1.79, 114.51) = 16.74$, $p = .00$, partial eta squared = .21. As expected, the main effect of item type indicated that across age group and test conditions, differences were significant between the means for the three item types, in that children provided more correct responses to consistent items than to control and misleading items and more correct responses to control items than to misleading items (Table 3).

Misinformation responses. To test Hypothesis 2 (see p. 11), the prediction that RES would be observed among older but not younger preschoolers, a 2 (Age Group: younger, older) x 2 (Test Condition: delay-only, initial) ANOVA was used to examine differences across groups in terms of misinformation responses. Children only provided misinformation responses to misleading items on the final cued-recall test (and never to consistent and control items), thus the only item type used in the analysis was the misleading item type. Test condition and age group were between-subjects factors and item type (i.e., misleading) was a within-subjects factor.

Levene's test of equality of variances was non-significant, $F(3, 64) = .08$, $p = .97$, however the ANOVA failed to reveal significant effects. The ANOVA failed to reveal any significant effects of age group, $F(1,64) = 1.74$, $p = .19$, partial eta squared = .03, or test condition, $F(1, 64) = .04$, $p = .84$, partial eta squared = .00, and there was no Age Group x Test Condition interaction, $F(1, 64) = .05$, $p = .83$, partial eta squared = .00 (Table 4). Hence, both younger and older preschoolers in this study failed to demonstrate retrieval-enhanced suggestibility.

To further examine the effects of age and test condition on misleading item performance, an additional analysis was conducted to determine whether these variables had an influence on incorrect responses to misleading items. In other words, group differences in the provision of ANY incorrect response to a misleading item, regardless of whether or not the response conveyed the actual misinformation detail from the story, were examined. Levene's test of equality of variances was non-significant, $F(3, 64) = .23$, $p = .88$, and the 2 (Age Group: younger, older) x 2 (Test Condition: delay-only, initial) analysis of variance revealed a main effect of age group, with younger children providing a significantly greater proportion of incorrect responses to misleading items than older children across test conditions, $F(1,64) = 7.71$, $p = .01$, partial eta squared = .11 (Table 4). Neither a main effect of test condition, $F(1, 64) = .09$, $p = .77$, partial eta squared = .00, nor an Age Group x Test Condition interaction, $F(1, 64) = .20$, $p = .65$, partial eta squared = .00, was found for incorrect responses to misleading items.

Confidence ratings. To examine differences across age groups and item type in confidence ratings for items on the cued-recall portion of the final test, two analyses were

conducted. First, a 2 (Age Group: younger, older) x 2 (Test Condition: delay-only, initial) univariate ANOVA was conducted to examine age group and test condition differences in total confidence to cued-recall items on the final test. Levene's test of equality of variances was not violated, $p = .36$, but the ANOVA failed to reveal significant effects, thus children did not differ across age groups, $F(1, 64) = 1.98$, $p = .17$, partial eta squared = .03, or test condition, $F(1, 64) = .01$, $p = .96$, partial eta squared = .00. The ANOVA failed to reveal a significant Age Group x Test Condition interaction as well, $F(1, 64) = .04$, $p = .84$, partial eta squared = .00.

To examine differences across item types in terms of confidence to cued-recall responses on the final test, a 2 (Age Group: younger, older) x 2 (Test Condition: delay-only, initial) x 2 (Item Type: consistent, misleading, control) mixed ANOVA was conducted, with age group and test condition as between-subjects factors and item type as a within-subjects factor. Box's M was non-significant, $p = .46$, as was Mauchly's test of sphericity, $p = .99$. Levene's test of homogeneity of variances was non-significant for each of the item types as well (consistent: $p = .58$, misleading: $p = .57$, control: $p = .78$). The mixed ANOVA failed to reveal any significant effects of age group, $F(1. 64) = 2.27$, $p = .14$, partial eta squared = .03, or test condition, $F(1, 64) = .00$, $p = .98$, partial eta squared = .00. However, there was a main effect of item type, $F(2, 128) = 6.69$, $p = .00$, partial eta squared = .10 (Table 5). Specifically, pairwise comparisons indicated that children reported higher confidence in responses to consistent item responses than to misleading item responses, $p = .01$, and higher confidence in control item responses than to misleading item responses, $p = .00$. Meanwhile, children did not differ in their confidence to consistent and control items, $p = 1.00$. Hence,

children's confidence in responses to misleading items, regardless of their accuracy, was in fact affected by exposure to misleading information in the storybook recap.

Source discrimination. To test Hypothesis 3 (see p. 11), the prediction that an initial test increases source discrimination performance among older but not younger children, a mixed ANOVA was used to examine age group and test condition differences in source-monitoring discrimination for consistent, misleading and control items on the source-monitoring portion of the final test. Specifically, a 2 (Age Group: younger, older) x 2 (Test Condition: delay-only, initial) x 3 (Item Type: consistent, control, misleading) mixed ANOVA was used to examine source discrimination accuracy, with age group and test condition as the between-subjects factors and item type as the within subjects factor. Box's M was found to be non-significant, $p = .61$; however Mauchly's test of sphericity was significantly violated, $p = .05$, therefore Greenhouse-Geisser epsilons are reported when relevant. Levene's test of equality of variances was not violated for consistent, $p = .68$, or control items, $p = .39$, but was significantly violated for misleading items, $p = .03$. The mixed ANOVA failed to reveal a main effect of age group, $F(1, 64) = .73$, $p = .40$, partial eta squared = .01, or test condition, $F(1, 64) = .46$, $p = .50$, partial eta squared = .01. The interaction between age group and test condition was also found to be non-significant, $F(1, 64) = .44$, $p = .51$, partial eta squared = .01. Additionally, the mixed ANOVA failed to reveal a main effect of item type for correct source discriminations, $F(1.83, 117.21) = 2.32$, $p = .10$, partial eta squared = .04.

A 3-way Age Group x Test Condition x Item Type interaction indicated that performance across item type differed as a function of age group and test condition in

unexpected ways, $F(2, 117.21) = 3.29$, $p = .045$, partial eta squared = .05 (Table 6). Inspection of a plot of the interaction indicated where the Age Group x Test Condition x Item Type difference was likely to be present. In the younger age group, it appeared as though children in the delay-only test condition provided fewer correct source discriminations for consistent items than initial test condition children, whereas it appeared as though younger children in the initial test condition provided significantly fewer correct source discriminations for misleading items, but did not differ in terms of performance across consistent and control items (Figure 1). In the older age group, it appeared as though children in the delay-only test condition provided the least correct source discriminations for misleading items, but did not differ in terms of performance between consistent and control items. Older children in the initial test condition did not seem to differ across item types (Figure 2).

To examine whether the patterns demonstrated by the plotted interactions were statistically significant, separate 2 (Age Group: younger, older) x 2 (Test Condition: delay-only, initial) univariate ANOVAs for were conducted for consistent and misleading items. An analysis was not conducted for control items, as there did not appear to be an interaction involving control items. For consistent items, Levene's test of homogeneity of variance was not violated, $p = .68$, meanwhile the ANOVA failed to reveal a significant Age Group x Test Condition interaction, $F(1, 64) = 2.47$, $p = .12$, partial eta squared = .04 (Table 7). Further, for misleading items, Levene's test of homogeneity of variances was violated, $p = .03$; however the Age Group x Test Condition interaction was significant, $F(1, 64) = 5.50$, $p = .02$, partial eta squared = .08 (Figure 3). Therefore, the 3-way Age Group x Test Condition x

Item Type interaction is attributable to a significant 2-way interaction between age group and test condition for misleading items but not for consistent items. The significant 2-way interaction revealed that younger children in the initial test condition provided significantly fewer correct source discriminations for misleading items than younger children in the delay-only condition, whereas the pattern was different in the older age group. Specifically, older children in the delay-only test condition provided significantly fewer correct source discriminations for misleading items than older children in the initial test condition (Table 7).

Discussion

This research provided an examination of the effects of initial recall prior to exposure to misinformation from a storybook source on preschoolers' memory for the details of a video-recorded event. In addition, the study examined the effects of initial recall prior to misinformation exposure on preschoolers' ability to discriminate the sources of original and post-event information. This investigation was prompted because previous examinations of retrieval-enhanced suggestibility had previously been limited to adult participants. The following sections will first outline the major findings of this investigation in terms of performance on the initial test, final cued-recall test, and source-monitoring discrimination. Next, these results will be discussed in the context of the predictions, with potential explanations for significance and null findings, concluding with a discussion of the limitations of the research and implications for future research.

The present study did not reveal RES, in that significant effects of test condition were not revealed, within or across age groups, with regard to the amount of misinformation provided. Therefore, an initial interview did not increase the amount of misinformation that

younger or older children provided on the final test. Confidence ratings across age groups in the present study provide some additional evidence that children's original memories were negatively affected by the storybook misinformation. Particularly, confidence in memories for information that was misrepresented in the storybook (i.e., misleading items) was significantly lower than for information not mentioned in the storybook (i.e., control) and information represented in a manner consistent with the original details.

Turning to overall source discrimination, effects of age group and test condition were not found and children across age groups demonstrated similar levels of source discrimination when collapsing across item type. However, children differed across age group, test condition and item type in the proportion of correct source discriminations. Specifically, in response to misleading items, younger children in the delay-only condition provided more correct source discriminations than those in the initial test condition. Older children demonstrated a reversal of this pattern by providing more correct source discriminations in the initial test condition than in the delay-only test condition. This pattern was partially in line with expectations, as it was originally predicted that the initial test would increase the amount of source cues that were available to the children at final recall, meanwhile only the older preschoolers were expected to be developmentally capable of using those source cues efficiently. The initial test increased the amount of correct source discriminations for misleading items in comparison to a delay-only test with the older preschool group; meanwhile, overall effects of test condition across item type were not found, contrary to the original predictions.

Initial Cued-Recall Performance

Researchers generally agree that age-related increases in this type of memory ensue across the preschool years, specifically between the ages of 2 and 6 years (e.g., Raj & Bell, 2010). It should be noted that the means were in the expected direction on the initial test, in that older children provided more correct responses and fewer "don't know" responses than younger children, albeit non-significantly. Since age group differences were significant on the final test, it is possible that the initial test occurred early enough to limit differences from appearing immediately after viewing the video. Thus, the absence of the age effect may be due to the fact that testing occurred immediately, thereby limiting information loss even among the younger children. Additionally, only half of the sample received an initial test, thereby limiting the power of detecting age group differences.

Do Preschoolers Demonstrate a "Testing Effect"?

Although the main purpose of the present study was to examine the existence of retrieval-enhanced suggestibility in preschoolers, the procedures also provided the opportunity to explore the original "testing effect" (Chan, McDermott & Roediger, 2006). Initial testing is thought to result in reconsolidation of tested material, thereby providing additional source cues for use during subsequent attempts to recall both the initially tested information, as well as non-tested information from the time of original encoding (Chan et al., 2006). Recall of initially tested information potentially facilitates recall of non-tested information by activating associated memories that were not explicitly tested initially (Chan et al., 2006). Based on research with adults, the testing effect would have been demonstrated had the children in the initial test condition demonstrated better recall for consistent and

control items in comparison to the children in the delay-only test conditions. However, in contrast to expectations, the present study did not reveal a testing effect for either consistent or control items.

The results do add to the existing literature, in that this is the first study to the authors' knowledge to examine the effects of an immediate cued-recall interview, *prior* to misinformation exposure, on preschoolers' subsequent recall performance across a short delay. This investigation presents findings consistent with studies that have failed to find increases in accuracy associated with early, neutral interviews (e.g., Baker-Ward et al., 1993; LaRooy et al., 2008). Meanwhile, this study contributes to this literature in that it is the first to examine the effects of such interviews on exposure to subsequently presented misinformation that is presented after the initial interview. In the present study, preschoolers in both age groups and test conditions provided more correct responses to consistent and control items than to misleading items, indicating they were less able to accurately recall information that was misrepresented in the storybook. This is consistent with previously well-established research that characterizes preschoolers as highly vulnerable to the influence of contradictory post-event information (e.g., Ceci & Bruck, 1993; Ceci, Huffman & Smith, 1994).

It should be noted, however, that a testing effect has been found in adult samples with delays as brief as 30 minutes (Chan et al., 2009; Chan & LaPaglia, 2011). Clearly, durations of time are not subjectively equivalent for individuals of greatly varying ages. On the one hand, the 10-minute delay used in this investigation with very young participants could be argued to be comparable to a longer interval as experienced by adult participants.

But, on the other hand, it is possible that the delay between initial and final interview in the present study was too brief to demonstrate the testing effect. Consistent with this possibility, a series of studies with adults recently revealed that the effects of repeated testing on memory for correct information are in fact stronger after extended delays, such 1-week in comparison to 30-minutes (Chan & LaPaglia, 2011). Benefits of early interviews have been found with children over lengthier delays than those utilized in the present study (i.e., at least 24 hours), suggesting that the testing effect could possibly become evident over similar time periods with children (e.g., Drohan-Jennings, Roberts & Powell, 2011; LaRooy, Pipe & Murray, 2005; LaRooy et al., 2008). Also, the effects of initial testing may become apparent in samples of children older than those in the present study, as test effects may be found at more advanced developmental levels. Previous studies examining memory for word lists have provided evidence for test effects in children ranging from at least 7- to 13-years-of age (e.g., Bouwmeester & Verkoeijen, 2011), although the existence of testing effects has yet to be examined with older children in terms of episodic memory.

Do Preschoolers Demonstrate a "*Reversed*-Testing Effect"?

Several predicted patterns regarding RES as characterizing children's event reports were not supported in the present investigation. First, age group differences were expected in misinformation responses to misleading items, overall. That is, younger children were expected to be unlikely to demonstrate a testing effect for two reasons. The baseline amount of misinformation provided (i.e., in the delay-only condition) was expected to be fairly high, as has been found consistently in previous research with this age group (e.g., Ackil & Zaragoza, 1995; Bright-Paul et al., 2005; Ceci et al., 1994). Also, younger children have

consistently demonstrated limitations in source-monitoring ability, and this capacity is thought to largely contribute to increased suggestibility (Johnson et al., 1993; Thierry & Spence, 2002; 2004). Unexpectedly, the younger children in this study did not report significantly more misinformation than the older children. This pattern of results conflicts with previous findings, which have revealed a memory advantage for older preschoolers in terms of recalling more information than younger preschoolers (e.g., Warren, Hulse-Trotter & Hubbs, 1991). Older preschoolers did, however, provide more *correct* responses on the final test, in line with established evidence revealing better performance on cued-recall memory tasks than younger children in general (e.g., Warren & Lane, 1995). Younger children also seemed to have more difficulty in general in responding to misleading items, overall, as evidenced by a higher number of *incorrect* responses to misleading items than older children, an additional frequently observed pattern (e.g., Ackil & Zaragoza, 1995; Ceci & Bruck, 1993; Karpinski & Scullin, 2009).

Second, contrary to predictions, children in the initial test condition *did not* provide more misinformation to misleading items than those in the delay-only test condition. Two potential possibilities may account for the lack of significance across test conditions. The present study may have failed to detect RES, even though it is present in preschoolers; or RES was not detected because it does not become problematic until later in development. It is possible that RES was not detected, even though it is present in preschoolers, for reasons mirroring those that might account for the non-significant effects of test condition on correct recall of consistent and control items, namely the short delay interval. The strength of the RES effect, as well as the original testing effect, has also been found as more pronounced

across lengthier delays in adult samples, even when correct recall increases as well (see Chan & LaPaglia, 2011). An increased delay may reveal RES in preschoolers, as suggestibility has sometimes been found to increase across multiple interviews in this age group as well (Melnyk & Bruck, 2004 for review; LaRooy et al., 2008). Alternatively, it is possible that testing effects, now in the context of RES, would become evident across multiple initial interviews as RES has been found to become pronounced in adult samples when multiple initial interviews occur prior to misinformation exposure (Chan & LaPaglia, 2011). Children do oftentimes become less accurate and more suggestible across multiple interviews with the consistency of their answers often fluctuating across interviews (Bruck & Melnyk, 2004; LaRooy et al., 2008). Multiple interviews prior to misinformation from a post-event source may therefore reveal the RES effect that had been originally predicted in older preschoolers.

The present study may also have failed to reveal RES among the preschool participants because the effect may be apparent only in more developmentally advanced individuals. Perhaps adults demonstrate RES, whereas preschoolers may not, because adults are developmentally able to hold onto memory traces longer in general (Brainerd & Reyna, 1998; Brainerd & Reyna, 2004). Verbatim, source-specific details such as those probed about in the present study are said to decay more rapidly in young children than in adults (Brainerd & Reyna, 1998). Memories for heard information consists of details that are much less perceptually salient than memories for witnessed information, making heard details much less memorable at the outset (Johnson et al., 1993). The relative strength of the encoded memories for the video and storybook misinformation probably differed considerably, affecting the retention of information from both sources. Preschoolers have in

fact been found to recall less story information than video information (Thierry & Pipe, 2009; Thierry & Spence, 2004), thereby suggesting that they may not have been capable of encoding and/or remembering the storybook (misinformation) enough to evidence RES.

Does Initial Testing Before Misinformation Influence Source Discrimination?

The ability to monitor source is based largely on an ability to discern discriminating details that tie, or bind, memory details to one's own experience or to another source (Johnson et al., 1993; Sluzenski et al., 2004). Following from previous research revealing improvements in source-monitoring performance between 4 and 6 years of age (e.g., Drummey & Newcombe, 2002; Thierry et al., 2001; Thierry & Pipe, 2009), children in the present study were expected to differ across age groups and test conditions in source discrimination. Children did not, however, differ significantly across age groups or test condition when examining overall source discrimination, collapsed across item type. However, despite violations of homogeneity and sphericity, a significant 3-way interaction revealed that children differed across age group, test condition and item type in correct source discriminations for misleading items. Among younger children, those in the delay-only test condition provided significantly *more* correct source discriminations for misleading items than those in the initial test condition. Among older children, those in the delay-only test condition provided significantly *fewer* correct source discriminations for misleading items than those in the initial test condition. These interactions indicate that initial recall prior to hearing misinformation differentially affects younger and older preschoolers' source discrimination. Although in this study, initial recall appears to hinder younger preschoolers' ability to discriminate the source of misleading items, whereas initial recall appears to

enhance older preschoolers' ability to do so, it is unlikely that this pattern is quite so straightforward and a parsimonious explanation for this pattern has not been identified.

Turning to older children, those in the initial test condition provided significantly more correct source discriminations for misleading items than those in the delay-only test condition. Although this interpretation is speculative, it is possible that older children used the additional source cues afforded by the initial test to discriminate the source of misleading items. Perhaps the initial test helped to bind video information in memory in the older age group, resulting in stronger memories with more discriminating features with which to identify each of the possible sources of information (Drummey & Newcombe, 2002; Sluzenski et al., 2004). It is therefore possible that older preschoolers were better able to recall source, overall, when they received an initial test, as evidenced by similar levels of discrimination across item type. Thus, this study may provide an indication that initial recall facilitates older preschoolers' ability to discriminate the source of post-event misinformation and if so, could result in better memory over time.

Limitations and Implications

Other study-related factors could have contributed to the non-significant findings in the present study as well. For instance, the present study utilized cued-recall items, whereas many of the cited source-monitoring studies with preschoolers utilize forced-choice recognition questions (e.g., Thierry, 2009). Although less accurate in response to forced-choice questions, it is considerably more difficult for preschoolers to respond to open-ended or cued-recall questions (Warren & Lane, 1995). Also, most source monitoring studies cited in this report have incorporated tasks that require children to first indicate that they remember

an item before attributing source (e.g., Thierry & Spence, 2002; 2004). Children in the present study were required to attribute the source(s) of information, regardless of recall accuracy, which may have been challenging in that children were to recall *where* information came from, even if they could not recall the video information first. Young children also often experience difficulty monitoring simultaneous sources (e.g., Ceci et al., 1994); therefore choosing among the four source options could have been potentially challenging in the present study. This is, however, unlikely due to the utilization of the source-box procedure that had been previously created to compensate for this inherent difficulty (Bright-Paul et al., 2005). Four source options were presented because merely presenting children with only two source options (i.e., the original and misleading sources) could inaccurately inflate discrimination scores by limiting the ability to capture when children would not attribute their memories to *either* source.

Finally, potential limitations could exist in drawing comparisons across children considered as preschoolers in this study and those considered as preschoolers in other source-monitoring studies, in that the developmental levels of the older participants could vary depending upon age and exposure to formal schooling. Many studies cited here have considered preschoolers as children 6-years of age and under, however, in the state in which this study was conducted, children reach school age if they are five years of age by August 31st of that calendar year. Meanwhile, the educational experiences of preschoolers and school-aged children (i.e., structured school days/lessons, etc.) likely contribute to important cognitive changes that could have confound with the outcome variables in this study.

Therefore, preschool-age was considered according to local standards by limiting inclusion to pre-kindergarten children under 6-years of age only.

Conclusions

To draw conclusions pertaining to RES as a developmental phenomenon, as well as pertaining to the effects of initial testing on discrimination of original and post-event sources in this age group would be premature. The questions addressed by the reported research nevertheless establish important attempts in making contributions to our understanding of the susceptibility to misinformation across development. Although speculative at this point, this study potentially provides evidence that RES, a seemingly problematic phenomenon with adults, may not be an issue for preschool-aged children. The present study also potentially provides some evidence that initial testing has differential effects on older and younger preschoolers' discrimination of a misinformation source. Future studies should validate the findings reported here and examine whether they hold under circumstances incorporating lengthier delays and multiple interviews. Interviewing children repeatedly in forensic circumstances is the norm as opposed to the exception (LaRooy et al., 2008); therefore researchers must establish more conclusive evidence regarding the effects of multiple interviews at various time points on the accuracy and consistency of children's memory reports.

References

Ackil, J. K., & Zaragoza, M. S. (1995). Developmental differences in eyewitness suggestibility and memory for source. *Journal of Experimental Child Psychology, 60*, 57-83. doi:10.1006/jecp.1995.1031

Ayers, M. S., Reyder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin and Review, 5*, 1-21. doi: 10.3758/BF03209454

Baker-Ward, L, Gordon, B. N., Ornstein, P. A., Larus, D. N., Clubb, B. A. (1993). Children's long-term retention of a pediatric examination. *Child Development, 64*, 1519-1533. doi: 10.1111/j.1467-8624.1993.tb02968

Baker-Ward, L., Ornstein, P. A., & Starnes, L. P. (2009). Children's understanding and remembering of stressful experiences. *In J. A. Quas & R. Fivush (Eds). Emotion and Memory in Development: Biological, Cognitive, & Social Considerations.* Oxford University Press: NY. Baker-Ward2009 doi: 10.1093/acprof:oso/9780195326932.003.0002

Beuscher, E., & Roebers, C. M. (2005). Does a warning help children more accurately remember an event, to resist misleading questions, and to identify unanswerable questions? *Experimental Psychology, 5*2, 232-241. doi: 10.1027/1618-3169.52.3.232

Boumeester, S., & Verkoeijen, P. P. J. L. (2011). Why do some children *benefit more from testing than others?* Gist trace processing to explain the testing effect. *Journal of Memory and Language, 65*, 32-41. doi:10.1016/j.jml.2011.02.005

Brainerd, C. J., & Ornstein, P. A. (1991). Children's memory for witnessed events: The developmental backdrop. In. J. Doris (Eds.) *The suggestibility of children's recollections.* Washington, DC: American Psychological Association. doi: 10.1037/10097-002

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

Brainerd, C. J. & Reyna, V. F. (2004). Fuzzy-trace theory and memory development. *Developmental Review, 24*, 396-439. doi:10.1016/j.dr.2004.08.005

Bright-Paul, A., Jarrold, C., & Wright, D. B. (2005). Age-appropriate cues facilitate source-monitoring and reduce suggestibility in 3- to 7-year-olds. *Cognitive Development, 20*, 1-18. doi: 10.1016/j.cogdev.2004.06.001

Ceci, S. J., & Bruck, M. (1993). Suggestibility of child witnesses: A historical review and synthesis. *Psychological Bulletin, 113*, 403-439. doi: 10.1037//0033-2909.113.3.403

Ceci, S. J., Huffman, M. L. C., Smith, E.. & Loftus, E. F. (1994). Repeatedly thinking about a non-event: Source misattributions among preschoolers. *Consciousness & Cognition, 3*, 388-407.

Chambers, K. L., & Zaragoza, M. S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition, 29*, 1120-1129. doi: 10.3758/BF03206381

Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological Science, 20*, 66-73. doi: 10.1111/j.1467-9280.2008.02245.x

Chan, J. C. K., & LaPaglia, J. A. (2011). The dark side of testing memory: repeated retrieval can enhance eyewitness suggestibility. *Journal of Experimental Psychology: Applied*. doi: 10.1037/a0025147

Chan, J. C. K., & Langley, M. M. (2010). Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition. Online first publication*, 1-8. doi: 10.1037/a0021204

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135*, 553-571.

Drohan-Jennings, D. M., Roberts, K. P., & Powell, M. B. (2010). Mental context reinstatement increases resistance to false suggestions after children have experienced a repeated event. *Psychiatry, Psychology and Law, 17*, 594-606. doi: 10.1080/13218711003739110

Drummey, A. B., & Newcombe, N. S. (2002). Developmental changes in source memory. *Developmental Psychology, 5*, 502-513. doi: 10.1111/1467-7687.00243

Echterhoff, G., Hirst, W., & Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cognition, 33*, 770-782. doi: 10.3758/BF03193073

Garry, M., Loftus, E. F., & Brown, S. F. (1994). Memory: A river runs through it. *Consciousness & Cognition, 3*, 438-451. doi: 10.1006/ccog.1994.1025

Gobbo, C. (2000). Assessing the effects of misinformation on children's recall: How and when makes a difference. *Applied Cognitive Psychology, 14*, 163-182. doi: 10.1002/(SICI)1099-0720(200003/04)14:2<163::AID-ACP630>3.3.CO;2-8

Hekkanen, S. T., & McEvoy, C. (2002). False memories and source-monitoring problems: Criterion differences. *Applied Cognitive Psychology, 16*, 73-85. doi: 10.1002/acp.753

Holliday, R. E., & Hayes, B. K. (2002). Automatic and intentional processes in children's recognition memory: The reversed misinformation effect. *Applied Cognitive Psychology, 16*, 1-16. doi: 10.1002/acp.789

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28. doi: 10.1037/0033-2909.114.1.3

Karpinski, A. C., & Scullin, M. H. (2009). Suggestibility under pressure: Theory of mind, executive function and suggestibility in preschoolers. *Journal of Applied Developmental Psychology, 30*, 749-763. doi:10.1016/j.appdev.2009.05.004

LaRooy, D., Lamb, M. E., & Pipe, M.-E. (2008). Repeated interviewing: A critical evaluation of the risks and potential benefits. In K. Kuehnle & M. Connell (Eds.) *Child Sexual Abuse: Research, Evaluation, and Testimony for the Courts.* Wiley.

LaRooy, D., Pipe, M.-E., & Murray, J. E. (2005). Reminiscence and hypermnesia in children's eyewitness memory. *Journal of Experimental Child Psychology, 90*, 235-254. doi:10.1016/j.jecp.2004.11.002

Melnyk, L., & Bruck, M. (2004). Timing moderates the effects of repeated suggestive interviewing on children's eyewitness memory. *Applied Cognitive Psychology, 18*, 613-631. doi: 10.1002/acp.1013

Memon, A., & Vartoukian, R. (1996). The effects of repeated questioning on young children's eyewitness testimony. *British Journal of Psychology, 87*, 403-415.

Poole, D. A., & Lindsay, D. S. (1995). Interviewing preschoolers: Effects of nonsuggestive techniques, parental coaching, and leading questions on reports of nonexperienced events. *Journal of Experimental Child Psychology, 62*, 129-154. doi: 10.1006/jecp.1995.1035

Powell, M. B., & Thompson, D. M. (1997). The effect of an intervening interview on children's ability to remember one occurrence of a repeated event. *Legal and Criminological Psychology, 2,* 247-262.

Raj, V., & Bell, M. A. (2010). Cognitive processes in supporting episodic memory formation in childhood: The role of source memory, binding, and executive functioning. *Developmental Review, 30*, 384-402. doi:10.1016/j.dr.2011.02.001

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development, 79*, 375-394. doi: 10.1111/j.14678624.2007.01131.x

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255. doi: 10.1111/j.1467-9280.2006.01693.x

Sluzenski, J., Newcombe, N. S., & Ottinger, W. (2004). Changes in reality monitoring and episodic memory in early childhood. *Developmental Science, 7*, 225-245. doi: 10.1111/j.1467-7687.2004.00341.x

Thierry, K. L. (2009). Practice retrieving source enhances young children's discrimination of live and story events. *Journal of Applied Developmental Psychology, 30*, 552-562. doi:10.1016/j.appdev.2008.12.030

Thierry, K. L., & Pipe, M.-E. (2009). The susceptibility of young preschoolers to source similarity effects: Confusing story or video events with reality. *Journal of Experimental Child Psychology, 102*, 392-407. doi: 10.1016/j.jecp.2008.12.005

Thierry, K. L., & Spence, M. J. (2002). Source-monitoring training facilitates preschoolers' eyewitness memory performance. *Developmental Psychology, 38*, 428-437. doi: 10.1037//0012-1649.38.3.428

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

Thierry, K. L., Spence, M. J., and Memon, A. (2001). Before misinformation is encountered: Source monitoring decreases child witness suggestibility. *Journal of Cognition and Development*, *2*, 1-26. doi: 10.1207/S15327647JCD0201_1

Thomas, A. K., Bulevich, J. B., & Chan, J. C. K. (2010). Testing promotes eyewitness accuracy with a warning: Implications for retrieval enhanced suggestibility. *Journal of Memory and Language, 63*, 149-157. doi: 10.1016/j.jml.2010.04.004

Warren, A. R., Hulse-Trotter, K., & Tubbs, E. C. (1991). Inducing resistance to suggestibility in children. *Law and Human Behavior, 15*, 273-285.

Warren, A. R., & Lane, P. (1995). The effects of timing and type of questioning on eyewitness accuracy and suggestibility. In M. Zaragoza, J.R. Graham, G.C.N. Hall,

R. Hirschman, Y. S. Ben-Porath (Eds.), *Memory and testimony in the child witness* (pp. 44- 60). Thousand Oaks CA: Sage.

Wrzesien, M., & Raya, M. A. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers & Education, 55*, 178-187. doi: 10.1016/j.compedu.2010.01.003

Table 1

Gender distribution across age group and test condition

		Delay-only	Initial Test	*N*
Younger				
	Males	8	10	18
	Females	11	12	23
	Total	19	22	41
Older				
	Males	9	6	15
	Females	5	7	12
	Total	14	13	27
Total				68

Table 2

Proportion of responses on the initial test as a function of age group (SDs in parentheses)

	Age group	
Response	Younger	Older
Correct	.51 (.22)	.62 (.26)
Don't Know	.55 (.20)	.45 (.25)

Note. No significant effects were found for correct responses, $t(33) = -1.35$, $p = .19$, CI 95% [-.28, .06] or "don't know" responses, $t(33) = 1.33$, p = .19, CI 95% [-.05, .26].

Table 3

Correct responses on the final cued-recall test as a function of age group, test condition and item type (SDs in parentheses)

	Age Group					
	Younger			Older*		
Item type**	Delay-only	Initial Test	Younger Total	Delay-only	Initial Test	Older Total
Consistent	.64 (.26)	.55 (.29)	.59 (.28)	.74 (.21)	.78 (.25)	.76 (.23)
Misleading	.44 (.21)	.44 (.25)	.44 (.23)	.54 (.21)	.53 (.30)	.53 (.25)
Control	.54 (.18)	.46 (.24)	.50 (.22)	.65 (.20)	.65 (.16)	.65 (.17)
Total	.54 (.16)	.49 (.21)	.51 (.19)*	.64 (.13)	.64 (.21)	.64 (.17)

* Main effect of age group, older children providing more correct responses than younger children, $F(1, 64) = 7.46$, $p = .01$, partial et squared = .10.

**Main effect of item type, proportions of correct responses differed significantly across all 3 item types (Greenhouse-Geisser epsilons reported), $F(1.79, 114.51) = 16.74$, $p = .00$, partial eta squared = .21.

Table 4

Misinformation and incorrect responses to misleading items on the final cued-recall test as a function of age group and test condition (SDs in parentheses)

	Younger		Older	
	Delay-only	Initial	Delay-only	Initial
Misinformation	.16 (.17)	.14 (.19)	.21 (.20)	.21 (.23)
Incorrect	.40 (.21)*	.44 (.24)*	.27 (.22)	.26 (.26)

*Main effect of age group for incorrect responses to misleading items, younger children providing more incorrect responses than younger children, $F(1,64) = 7.71$, $p = .01$, partial eta squared = .11.

Table 5

Confidence in responses to cued-recall items on final test as a function of age group and test condition (SDs in parentheses)

	Age Group				
	Younger		Older		
Item type*	Delay-only	Initial Test	Delay-only	Initial Test	Total
Consistent	.72 (.17)	.72 (.24)	.81 (.21)	.79 (.24)	.75 (.21)
Misleading	.65 (.19)	.70 (.20)	.70 (.21)	.68 (.24)	.69 (.21)
Control	.73 (.19)	.71 (.18)	.81 (.18)	.81 (.20)	.76 (.19)
Total	.65 (.12)	.66 (.18)	.72 (.16)	.68 (.14)	.68 (.18)

*Main effect of item type, $F(2,128) = 6.69$, $p = .00$, partial eta squared = .10, with children reporting higher confidence in responses to consistent items than to misleading items, $p = .01$, and higher confidence in control items than to misleading items, $p = .00$.

Table 6

Source discrimination scores as a function of Age group, test condition and item type (SDs in parentheses)

	Age Group x Test Condition				
	Younger		Older		
Item type	Delay-only	Initial	Delay-only	Initial	Total
Consistent	.21 (.29)	.35 (.27)	.41 (.33)	.32 (.27)	.32 (.29)
Misleading	.34 (.33)*	.16 (.19)*	.16 (.19)*	.29 (.28)*	.24 (.26)
Control	.38 (.25)	.30 (.20)	.36 (.30)	.33 (.27)	.34 (.25)
Total	.32 (.13)	.27 (.11)	.33 (.11)	.32 (.12)	.31 (.12)

*Age Group x Test Condition x Item Type interaction, indicating that performance across item type differed as a function of age group and test condition (Greenhouse-Geisser epsilons reported), $F(2, 117.21) = 3.29$, $p = .045$, partial eta squared = .05.

Table 7

Correct source discriminations for consistent and misleading items only as a function of age group and test condition (SDs in parentheses)

	Younger		Older	
	Delay-only	Initial	Delay-only	Initial
Consistent	.21 (.29)	.35 (.27)	.41 (.33)	.32 (.27)
Misleading	.34 (.33)*	.16 (.19)*	.16 (.19)*	.29 (.28)*

*Significant Age Group x Test Condition interaction for misleading items (Levene's test of homogeneity of variances was violated, $p = .03$), $F(1, 64) = 5.50$, $p = .02$, partial eta squared = .08.

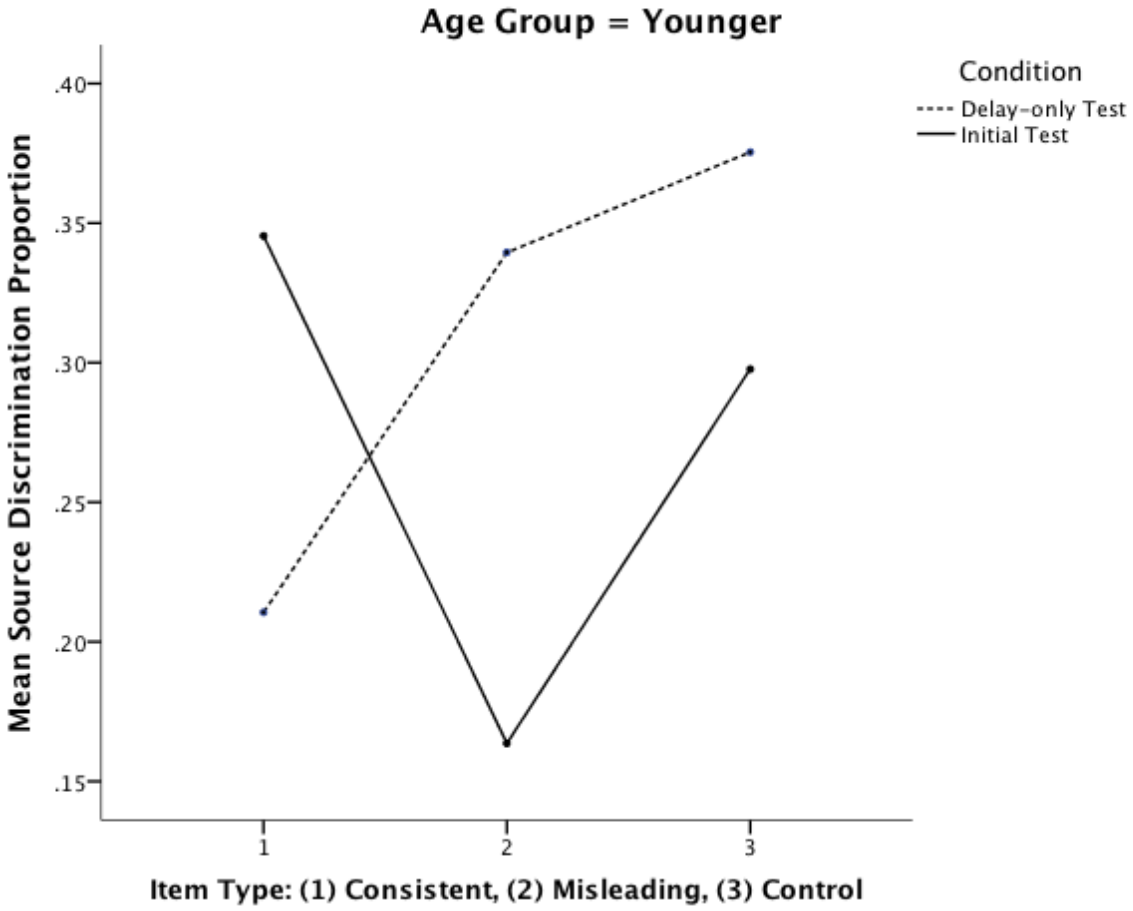


Figure 1. Source discrimination performance across item type as a function of test condition in the younger preschool age group.

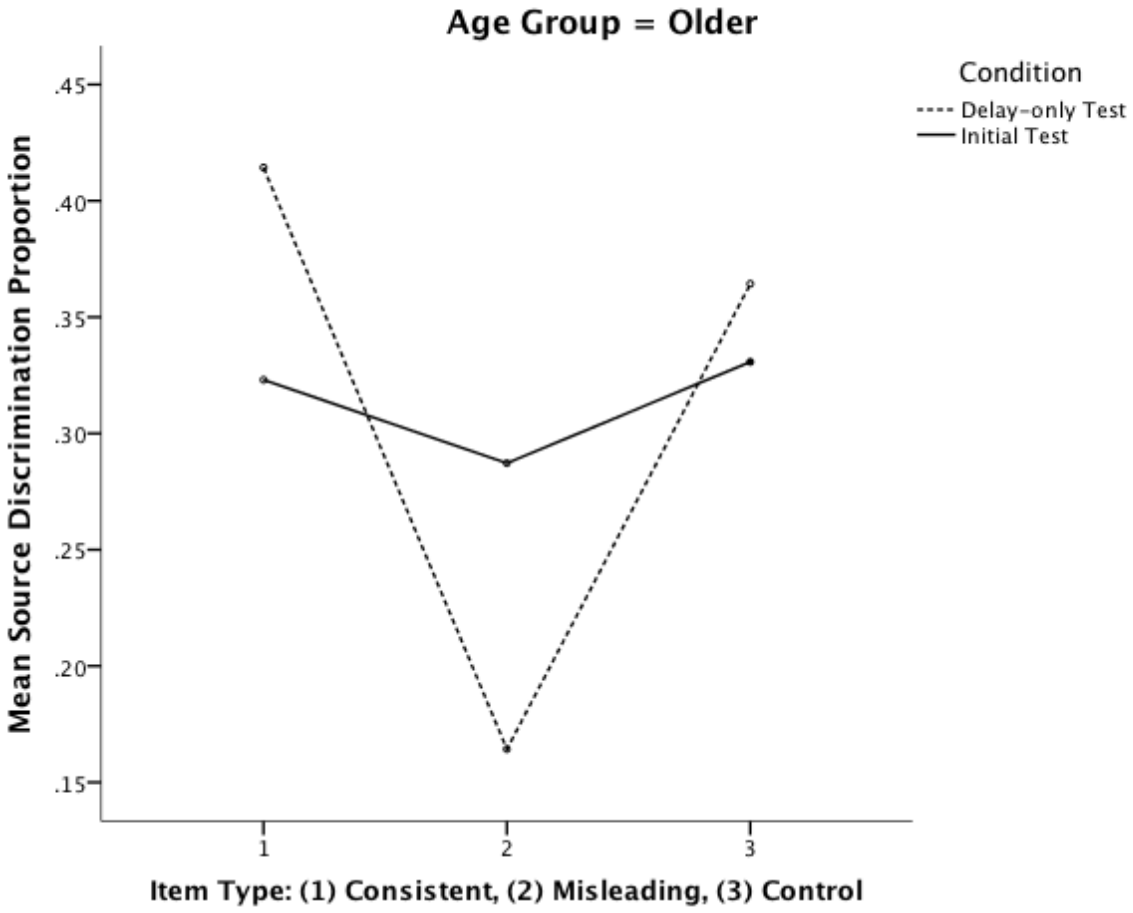


Figure 2. Source discrimination performance across item type as a function of test condition in the older preschool age group.

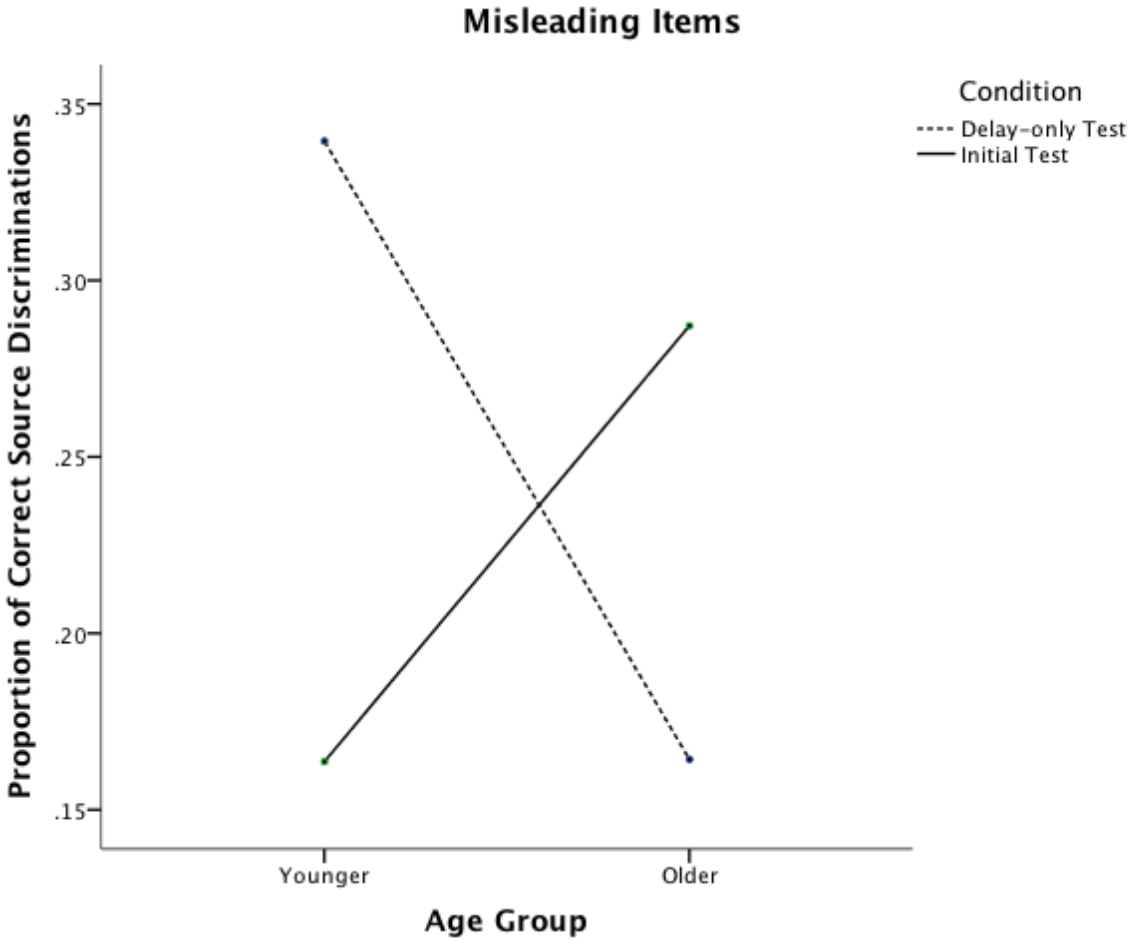


Figure 3. Two-way, Age Group x Test Condition interaction in the proportion of correct source discriminations for misleading items.

APPENDICES

Appendix A

Components in "Mrs. Science's" Experiments

1. Magnet Test	Mrs. Science: a. Drops 1st paperclip inside plastic cup of water b. Slides magnet outside cup to get paperclip out c. Puts 2nd paperclip on top of ruler d. Slides magnet underneath ruler to make paperclip "climb" the ruler
2. Magic Coin	Mrs. Science: a. Tapes coin to bottom of striped bowl b. Backs away from bowl until she cannot see coin anymore c. Pours water into bowl (using pitcher) d. Keeps pouring water until she can see coin
3. Rubber Band Guitar	Mrs. Science: a. Connects tube to shoebox to "make guitar" b. Places one rubber band around box c. Places second rubber band around box d. "Plays" guitar
4. Paper Plunge	Mrs. Science: a. Scrunches up piece of paper b. Places scrunched-up paper into bottom of empty glass c. Turns glass upside down and dip straight into bowl of water d. Takes out the glass to see that the paper remains dry
5. Magic Balloons	Mrs. Science: a. Rubs balloon against sweater b. Holds balloon over torn up pieces of construction paper c. Watches as paper sticks to balloon d. Rubs balloon against sweater
6. Picture Mixer	Mrs. Science: a. Rubs balloon against sweater b. Holds balloon over torn up pieces of construction paper c. Watches as paper sticks to balloon d. Rubs balloon against sweater

Appendix B

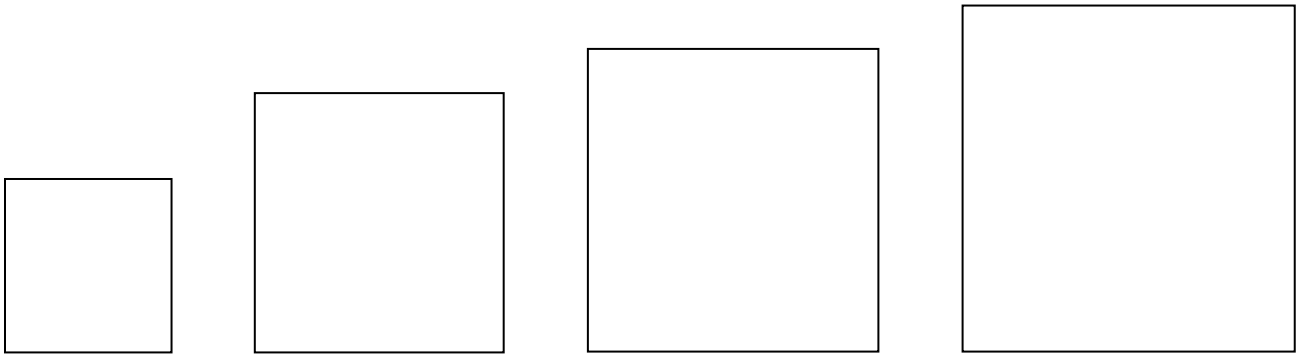
Cued-Recall Questions, Item Type and Correct Responses

No.	Question[1]	Item Type[2]	Correct Response	Misled Response[3]
1.	What did Mrs. Science rub a balloon against to make static electricity?	Consistent	Black Sweater	N/A
2.	Did any balloons break when she did the experiments?	Misleading	No	Yes
3.	What color was Mrs. Science's lab coat?	Control	White	N/A
4.	How many glasses did Mrs. Science use to get the paperclip out?	Control	One glass	N/A
5.	What happened when Mrs. Science tried to get the paperclip out?	Misleading	It worked	It didn't work
6.	What did Mrs. Science use to make the paperclip climb the ruler?	Consistent	Magnet	N/A
7.	What did Mrs. Science tape to the bottom of a bowl?	Consistent	Coin	N/A
8.	What did Mrs. Science pour into the bowl?	Consistent	Water	N/A
9.	What did Mrs. Science use to crush a paper into the bottom of a glass?	Control	Her hand	N/A
10.	What happened when Mrs. Science put the glass with the paper into the bowl?	Misleading	It stayed dry	It got all wet
11.	What did Mrs. Science do with the guitar she made?	Misleading	Strummed the bands	Played Twinkle Twinkle
12.	What did she use to play the guitar?	Control	Her hands	N/A
13.	Where did Mrs. Science get the pictures of the bird and the cage?	Misleading	She drew them	Her camera
14.	Did Mrs. Science spin the pencil fast or slow?	Control	Fast	N/A
15.	What did Mrs. Science do with the pencil after she put the pictures on top of it?	Consistent	Rubbed it with her hands	N/A

[1]Questions presented in random order.
[2]Consistent = same in video and narrative, Misleading = replaced with plausible alternative in narrative, Control = not mentioned in narrative.
[3]Misleading items only.

Appendix C

Confidence Rating Scale



Wrezsien & Raya (2010)

Appendix D

Initial Test Coding Sheet

Question #	Answer	Response: (C) Correct (IN) Incorrect (DK) Don't know	Confidence Rating
1.	Black sweater		
2.	No, 0 balloons broke		
3.	Blue		
4.	One glass		
5.	It worked/she got the clip out of the glass.		
6.	Magnet		
7.	Coin/money		
8.	Water		
9.	Her hand		
10.	The paper stayed dry		
11.	She played it		
12.	Her hand, fingers		
13.	She drew the pictures		
14.	Really fast		
15.	Rubbed the pencil between her hands		

Appendix E

Final Test Coding Sheet

Question #	Answer C: Correct M: Misinformation	Response C M IN DK	Conf. Rating	Source Answer V= video S = Story B = Both N = Neither	Source Response C IN
1. Consistent	C: Black sweater			B	
2. Misleading	C: No, 0 balloons broke M: Yes, balloons broke			S	
3. Control	C: Blue			V	
4. Control	C: One glass			V	
5. Misleading	C: It worked/she got the clip out of the glass M: It didn't work			S	
6. Consistent	C: Magnet			B	
7. Consistent	C: Coin/money			B	
8. Consistent	C: Water			B	
9. Control	C: Her hand			V	
10. Misleading	C: Paper stayed dry M: Paper got wet.			S	
11. Misleading	C: Played guitar M: Played Twinkle Twinkle			S	
12. Control	C: Her fingers			V	
13. Misleading	C: She drew the pictures M: From her camera			S	
14. Control	C: Really Fast			V	
15. Consistent	C: Rubbed the pencil between her hands			B	