Child Abuse & Neglect 38 (2014) 192–201

Contents lists available at ScienceDirect

Child Abuse & Neglect

Comfort drawing during investigative interviews: Evidence of the safety of a popular practice☆

Debra Ann Poole[a,*], Jason J. Dickinson[b]

[a] Department of Psychology, Central Michigan University, Mt. Pleasant, MI 48859, USA
[b] Department of Psychology, Montclair State University, Montclair, NJ 07043, USA

ARTICLE INFO

Article history:
Received 27 September 2012
Received in revised form 20 February 2013
Accepted 23 April 2013
Available online 14 June 2013

Keywords:
Children
Eyewitness testimony
Investigative interviews
Comfort drawing

ABSTRACT

Objective: This study evaluated the impact of comfort drawing (allowing children to draw during interviews) on the quality of children's eyewitness reports.
Methods: Children (N = 219, 5 to 12 years) who had participated in an earlier memory study returned 1 or 2 years later, experienced a new event, and described these events during phased, investigative-style interviews. Interviewers delivered the same prompts to children in the no drawing and drawing conditions but provided paper and markers in the drawing condition, invited these children to draw, and periodically asked if they would like to make another picture.
Results: Most children in the drawing condition were interested in using the materials, and measures of eyewitness performance were sensitive to differences in cognitive ability (i.e., age) and task difficulty (i.e., delay between the remote event and interview). Comfort drawing had no overall impact as evidenced by nonsignificant main effects of condition across 20 performance measures, although more of the younger children reported experienced touching in the drawing than no drawing condition.
Conclusions: The children successfully divided attention between voluntary drawing and conversations about past events. Importantly, comfort drawing did not impair the amount of information recalled, the accuracy of children's answers, or even the extent to which interviewers needed to prompt for answers. Due to the large number of analyses, the benefit of drawing for younger, touched children requires replication.
Practice Implications: Comfort drawing poses no documented risks for typically-developing school-aged children, but the practice remains untested for younger children and those with cognitive impairments.

Introduction

Following a series of high-profile abuse cases in the 1980s, findings from studies of children's eyewitness testimony prompted the development of best-practice standards that direct how professionals should interview children about suspected abuse (Poole & Dickinson, 2013). But despite widespread consensus on some issues, there is considerable variability in the specific recommendations that appear across interviewing protocols (cf. Anderson et al., 2010; Lamb, Hershkowitz, Orbach, & Esplin, 2008) and even more variability in the ways interviewers translate advice into practice. Some of this variability stems from the fact that researchers (and, therefore, best-practice standards) have not sufficiently addressed

☆ This research was supported by grants SES-0921227 (DP) and SES-SES-0921161 (JD) from the National Science Foundation.
* Corresponding author address: Department of Psychology, 231 Sloan Hall, Central Michigan University, Mt. Pleasant, MI 48859, USA.

0145-2134/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chiabu.2013.04.012

children's emotional needs, so some investigators feel justified in using untested strategies. The current study examined one of these strategies: comfort drawing, which is the practice of allowing children to draw during forensic interviews.

Although there are no data on how often interviewers provide drawing materials to children, comfort drawing appears in videotapes from communities across the United States and is routinely used during interstate child pornography investigations. Proponents argue that drawing facilitates children's testimony by creating a more supportive environment and diffusing the stress of an interview. For example, U.S. Federal Bureau of Investigation child/adolescent forensic interviewer Catherine S. Connell told us that coloring has always been instrumental in her interviews because it "seems to make the atmosphere less stressful when a victim and/or interviewer can look down at the paper and crayons as opposed to staring at each other waiting for a question or answer" (February 15, 2013, used with permission). On the other hand, there is concern that coloring will divert children's attention and deplete the cognitive resources that promote accurate accounts of events (e.g., U.K. Ministry of Justice, 2007).

We initiated the current study at the request of prosecutors who discovered that interviewers in their jurisdictions were allowing children to play with markers or crayons during sexual abuse interviews. Although there was research on the impact of directed drawing (i.e., asking children to draw target events; see Pipe & Salmon, 2009, for a review), we found no study that had evaluated the benefits and risks of using a familiar activity such as drawing simply to create a less stressful interview environment. At the time, we were conducting a multisite study that exposed children to a target event and, with the cooperation of their parents, misinformation about some event details (Poole & Dickinson, 2011). To respond to prosecutors' concerns, we invited families to return to our laboratories (approximately 1 or 2 years after their initial sessions), staged a new event to permit questioning about a recent event in addition to the long-ago (i.e., remote) event, and assigned each child to one of two versions of a phased interviewing protocol: with or without comfort drawing.

The literature on children's memory and attention explains why there are competing hypotheses about the impact of comfort drawing on children's testimony: Social support and distractions act in opposite directions, with interviewer support improving performance while a competing task usually (but not always) impairs it. Specifically, a supportive interviewing style has a positive impact on children's free recall performance and helps them resist misleading suggestions (as long as supportive and unsupportive styles are sufficiently different; Bottoms, Quas, & Davis, 2007; Hershkowitz et al., 2007), and these benefits are greater for children who are most physiologically reactive (Quas, Bauer, & Boyce, 2004; Quas & Lench, 2007). However, multitasking typically impairs children's and adults' performance on cognitive tasks (e.g., Anderson, Bucks, & Bayliss, 2011; but see Ciaramelli, Ghetti, & Borsotti, 2009). Therefore, it is possible that comfort drawing could improve the quality of children's autobiographical reports (e.g., if drawing was stress-reducing but not very distracting), impair performance (e.g., if stress levels were not strongly affected but drawing reduced attention to interviewers' questions), or have no impact (e.g., if drawing did not change the emotional tone of interviews and was so familiar that it did not tax cognitive resources).

The possibility that children might be able to navigate answering questions while drawing is suggested by research on another favorite activity: watching television. Even preschool children easily divide attention between ongoing play and programming, and they are skilled at monitoring soundtrack features that signal when something interesting will occur (Anderson & Levin, 1976; Anderson, Lorch, Field, & Sanders, 1981; Calvert, Huston, Watkins, & Wright, 1982). If this ability to multitask generalizes to other situations involving highly familiar tasks, then drawing might not interfere with children's ability to answer developmentally appropriate questions. After all, drawing is neither a particularly gripping activity (that is, it should not "lock in" children's attention) nor one that requires significant cognitive resources (as evidenced by the fact that children chat effortlessly while drawing). Conversely, young children find some question forms challenging (e.g., yes–no questions, Peterson, Dowden, & Tobin, 1999), so it is possible that even minor distractions could derail accuracy when interviewers ask about specific event features.

The lack of research on comfort drawing, along with the possibility that drawing could affect performance differently for different types of questions, led us to cast a wide net and code 20 variables that reflected children's performance throughout the interview. We included a measure of engagement in the interview before interviewers raised the first target topic (because early behavior could influence perceptions of children's maturity and credibility); the success of topic introduction for the remote event; and measures that tapped different memory skills, including free-recall (i.e., responses to open-ended invitations), recognition (i.e., responses to yes-no questions), and source-monitoring ability (i.e., the ability to report whether knowledge of an event originated from personal experience or another source). We also looked at children's ability to report a set of forensically relevant details and how frequently interviewers needed to prompt for answers in the no drawing and drawing conditions.

Because existing research did not allow us to hypothesize whether comfort drawing would benefit, hinder, or have no impact on eyewitness testimony, we recruited a large sample ($N = 219$) to increase our chances of detecting modest differences between conditions. This sample also allowed us to look separately at the performance of younger children and children with inhibited temperaments. To measure the latter variable, we asked parents to rate how behaviorally inhibited their children were using the Behavioral Inhibition Questionnaire. Behavioral inhibition to the unfamiliar is a well-researched construct associated with traits/behaviors that have important consequences for eyewitness performance, including anxiety, hesitancy to approach strangers, little spontaneous speech, and the need for more prompting to encourage responses (Bishop, Spence, & McDonald, 2003). Because the physiological pathways involved in reactions to highly stressful experiences are mirrored at lower levels of stress among inhibited children (Kagan, Reznick, & Snidman, 1987), behavioral inhibition scores allowed us to explore the effects of comfort drawing on a group of children who adjust slowly to novel situations.

Table 1
Mean characteristics of children in the no drawing and drawing conditions.

Characteristic	No drawing		Drawing		*t*	*p*
	(*n* = 108)		(*n* = 111)			
1-Year delay						
Age at time of event (years)	6.97	(1.71)	7.06	(1.71)	−.30	.77
Age at final interview (years)	8.34	(1.71)	8.43	(1.70)	−.31	.75
Delay between remote event and interview (years)	1.37	(0.06)	1.37	(0.06)	−.40	.69
Previous free recall production	162.16	(126.06)	168.23	(125.10)	−.27	.79
Previous source-monitoring score	.75	(.21)	.77	(0.19)	−.52	.60
2-Year delay						
Age at time of event (years)	6.92	(1.89)	7.15	(1.75)	−.62	.54
Age at final interview (years)	9.19	(1.87)	9.43	(1.75)	−.64	.52
Delay between remote event and interview (years)	2.27	(0.07)	2.28	(0.06)	−.57	.57
Previous free recall production	141.87	(105.43)	137.04	(105.11)	.22	.83
Previous source-monitoring score	.77	(0.15)	.78	(0.19)	−.10	.92
Total sample						
Age at time of event (years)	6.95	(1.79)	7.10	(1.72)	−.64	.52
Age at final interview (years)	8.71	(1.82)	8.85	(1.78)	−.57	.57
Delay between remote event and interview (years)	1.76	(0.46)	1.75	(0.45)	.22	.83
Previous free recall production	153.33	(117.44)	155.31	(117.74)	−.12	.90
Previous source-monitoring score	.76	(0.19)	.77	(0.19)	−.47	.64

Note: Standard deviations are in parentheses. Previous free recall production scores are Modified Word Counts in the earlier study (Poole & Dickinson, 2011); previous source-monitoring scores are the proportion of ten target events that were accurately reported in the earlier study as experienced or nonexperienced after source-monitoring instructions and questioning.

Method

Participants

Following approval from institutional review boards at the two research sites, families who had participated in the previous study received a thank-you letter with a form and return envelope to express interest in two new opportunities: filling out a questionnaire about their children's personalities (for $25) and returning to the laboratory for an additional session (for $35). Assistants then called families to describe the new projects.

The final sample included 219 children from 5 to 12 years of age (83.9% of the previous study's sample; 49.8% female). Just under half (*n* = 103) lived in small town/rural communities in the Midwest. Participants at this site were predominantly Caucasian (91.3%, with 1.0% African American, 3.9% Asian, and 3.9% "other") and 6.8% were Hispanic. The participants were economically diverse: 16.0% of parents who revealed family income reported incomes under $30,000 per year, and 46.0% reported incomes under $50,000 per year. Children at a second site (*n* = 116), in the New York metropolitan region, were also predominantly Caucasian (82.8%, with 6.0% African American, 3.4% Asian, and 7.8% "other") and 6.0% Hispanic; only 6.1% of these families reported incomes under $50,000 per year. All children were fluent in English, although across sites 6.4% of families said that another language was also spoken in the home. Earlier analyses found no overall differences in eyewitness performance across sites (Poole & Dickinson, 2011).

Initial study enrollees were randomly assigned to drawing or no drawing conditions, with subsequent enrollees assigned to balance a number of subject and performance variables across conditions. As described in Table 1, the two conditions were comparable in terms of children's ages at the time of the memory event and the final interview, the length of time between the target event and interview, and children's performance on two measures from the initial study: free recall productivity (i.e., number of words spoken, as measured by Modified Word Counts; Dickinson & Poole, 2000) and overall source-monitoring scores (i.e., the proportion of 10 target events that each child accurately reported as experienced or nonexperienced; see Poole & Dickinson, 2011). The two performance variables were selected a priori because the major goal of interviewing enhancements is to increase the amount of information children report (as reflected in words spoken during free recall) without sacrificing accuracy.

Procedure

Interviewers asked children about two events: a laboratory visit that had occurred approximately 1 or 2 years earlier (i.e., the remote event, see Poole & Dickinson, 2011) and a new event staged minutes before the final interview. Questions about the remote event were challenging not only because this event had occurred long ago but also because the children's parents had exposed them to false suggestions regarding some components of that event.

Remote (Mr. Science) event. Each initial laboratory visit began with a 15-minute target event during which "Mr. Science" secured ties on the back of the child's lab coat, set a timer, and presented four science demonstrations with encouragement for the child to handle the materials (for a description, see Poole & Lindsay, 2001). Individual children experienced one of

two demonstration sets: (a) pulley systems, floating paperclips, catching tops with and without prism glasses, and testing paper airplanes, or (b) the magic eyedropper, blowing up balloons with baking soda and vinegar, making a telephone, and lifting drawings with play putty.

Each child assigned to a touch condition also experienced two target touches. First, Mr. Science tried unsuccessfully to fit a small wrist band on the child, wrapped his fingers around the child's wrist to measure it, and then taped on a larger band that remained on the child until Mr. Science removed it at the end of the session. At that time, Mr. Science attempted to push a worn-out reward sticker on the child's shoulder but then gave the child a strip of stickers to take home instead. Immediately after these events, a female interviewer delivered a short baseline interview consisting of rapport building and three open-ended prompts that invited the child to talk about what had happened in the science room.

Several months later, each parent received a book in the mail with instructions to read the book to their child on 3 consecutive days before the next interview, to record those dates on the front cover, and to read twice in one day if they inadvertently missed a day. Each story described two science demonstrations the child had experienced and two nonexperienced demonstrations. Counterbalancing of multiple versions of the book and the two demonstration sets produced a within-subject design in which each of the eight science demonstrations appeared in four demonstration conditions across children: experienced (but not described in the book), experienced and described in the book, suggested (not experienced but described in the book), and control (neither experienced nor described in the book). The books for children who had not been touched also described the two target touch events, thereby creating suggested touch events for these children.

Each child then participated in a body-diagram-focused or standard interview delivered by one of 20 interviewers who conducted interviews in both conditions (see Poole & Dickinson, 2011, for a description of the interview protocols). Major findings were that body-diagram-focused interviews elicited more reports of experienced touching, suggested touching, and nontarget touching (spontaneous intrusions) than did standard interviewing, but a source-monitoring procedure at the end of interviews (which asked about experienced vs. suggested events) equalized performance across the two interviewing conditions.

Recent (Dog Lady) event. When families arrived for the follow-up interview, assistants first obtained parental consent followed by child assent (for children less than 7 years of age) or consent (for children 7 years of age or older). Each child then visited with "Bonnie," a woman wearing a red smock printed with cartoon dogs. Bonnie explained that people called her the "Dog Lady" because she loved dogs and that her boss let her pick which smock to wear for work. Bonnie then pointed out the different breeds of dogs and their names (which were printed on the fabric) and asked how many string tails the child could find on the smock. Next, Bonnie explained that she was supposed to take the child's picture, but after taking one picture with the child sitting down she slapped her head and admitted that she had done something wrong (because her boss had asked her to take pictures with children standing on an X that was taped to the floor). After taking a second picture, Bonnie let the child select one of three dog activity worksheets, showed the child a funny sheet that another child had made, and asked the child to tell her why the sheet was funny. (Everything was the wrong color and the other child had completed drawings of dogs by adding parts from other animals.) This event provided the opportunity for interviewers to probe for six forensically-relevant details: whether Bonnie had taken pictures of the child, how many pictures she had taken, whether she had done something wrong, whether she had given the child anything (scoring the child correct for mentioning the activity sheet), and whether she had shown the child anything (scoring the child correct for mentioning the apron and the other child's activity sheet).

Final interview. Seven interviewers conducted interviews in both conditions, with one session handled by a back-up interviewer. Instructions and questions/prompts were identical in the no drawing and drawing conditions except for the following: In the drawing condition, interviewers invited children to draw a picture while they talked (using paper and markers that were on the table) and for children who had stopped drawing, asked if they would like to make another picture at two predetermined points during the interview.

Each phased interview began when the interviewer introduced herself ("Hi, my name is. . ."), explained that her job was to listen to children and that she might write down some notes, and pointed out that the tape recorder would record what the child said, "So I need you to talk nice and loud, okay?" (introduction). She then asked questions about the child's age and grade (initial rapport-building) and continued rapport-building by asking the child to talk about what happens at school, "from the time you get to school until the time you come home" (practice interview). During the practice interview, interviewers prompted reluctant children to begin talking ("What do you do when you first get to school?") and helped children narrate events until the end of the school day with open-ended prompts (e.g., "And then what happens?").

The interviewer then explained the need to not guess, to tell the interviewer when the child didn't understand a question, to correct the interviewer when she said something wrong, and to tell the truth (ground rules phase). This section of the interview contained seven scorable questions (e.g., "For example, what is my dog's name?", "I'm sitting down right now. Is that true or not true?", "Will you tell me the truth today?").

Interviewers introduced conversation about the remote event by asking if the child remembered coming to the university a long time ago to do some science experiments and to talk to a lady. Two additional prompts were delivered if the child had difficulty latching onto the topic, including (if necessary) showing the laboratory coat and timer that were used in the remote event (topic introduction). Interviewers next delivered four open-ended prompts about the science experience (free narrative phase) and 10 yes-no questions about specific event components (e.g., "Did Mr. Science make a telephone?"),

followed by requests to describe whenever the child said "yes" (e.g., "Tell me about the telephone"; specific questioning phase).

The final questions measured whether the child could distinguish between experienced events and events that had only been described in the misleading book. The interviewer began with the following instructions (source-monitoring phase):

> I'm going to ask a few more questions about the things you told me. Sometimes we know about something because we remember when it happened—we remember seeing and hearing it happen. But sometimes we know about things because someone told us about them. I'm interested in what you really remember happening with Mr. Science—things you remember seeing or hearing when you were with Mr. Science. For example, I'm going to say something. "Mr. Science flew across the room!" Did Mr. Science really fly across the room? (Wait for a response or prompt a response.) That's right, he didn't really fly across the room, so you were right to say "no." If I ask about something you remember doing with Mr. Science, then I want you to say "yes." Let's do some more.

Next, the interviewer revisited each yes-no question that had elicited a "yes" response and asked, "Did Mr. Science really. . .?"

The interviewer then transitioned into conversation about the Dog Lady event. She completed the interview with two free-narrative prompts (e.g., "Tell me everything that happened while you were with Bonnie. . .") and the following questions: "Did Bonnie take your picture?" (if the child had not mentioned this detail), "How many times did Bonnie take your picture?", "Bonnie just started working for us this year. . ..did she do anything wrong?", "Did Bonnie give you anything?" (followed by "Tell me about what she gave you."), and "Did Bonnie show you anything in the room?" (followed by "Tell me about what she showed you.").

Recording interactions with the drawing materials

To describe children's use of the drawing materials, a coder watched each interview in the comfort drawing condition and recorded one of three behaviors within each of 10 interview segments: no interaction with the drawing materials (or only held them), play behavior only (e.g., twirling markers, moving markers in and out of the box, or other interactions with the markers, box, or paper without marking on paper), or drawing behavior (i.e., any marking on paper). A second coder independently reviewed 23 videotapes (230 interview segments). Intercoder agreement on children's behavior was 95.2%; Cohen's $k = .85$.

Performance measures

Twenty performance measures captured the quantity and accuracy of the information children provided, as well as how often interviewers needed to prompt for answers. This list follows the overall progression of the interview but blocks together measures produced by the same coding process.

Early interview phases. 1. Presubstantive prompts (proportion correct): the proportion of the first 11 questions/prompts that children answered correctly, including the 3 introduction/rapport-building questions, initiation of the practice interview (about a day at school), and 7 questions during the ground rules discussion. Children were credited with correct responses if they provided the desired response, even if interviewers needed to prompt for answers or clarify questions. Individual responses were evaluated by two coders; agreement = 99.5%, $k = .89$. (For all measures, disagreements were resolved by discussion.)

Remote event. 2. Successful topic introduction (0 = unsuccessful, 1 = successful): the percentage of children who acknowledged remembering their previous visit to the laboratory (intercoder agreement = 99.1%, $k = .89$).

3. Free recall—length of narrative: the amount of information children provided to four open-ended questions about their earlier study participation (the Mr. Science event), coded as Modified Word Counts (MWCs). This measure of verbal productivity included accurate descriptions of experienced events and inaccuracies (e.g., errors in descriptions of experienced events and intrusions, such as descriptions of school science fairs).

MWCs are the number of words that remain after coders scrub narratives to delete uninformative words (e.g., talk that anyone would recognize as off topic, conjunctions, false starts, and repeated phrases), and these values correlate highly with discrete units of information (Dickinson & Poole, 2000). Throughout this manuscript, word counts refer to MWCs. Two coders processed answers to open-ended questions and narratives following yes–no questions; agreement on keeping or deleting individual words was 94.6%, $k = .88$.

4. Free recall—proportion of narrative describing experienced events: the number of words describing experienced events divided by the total number of words. This measure captured the proportion of talk that described participation in the previous study but did not dock children for inaccuracies unless they were clearly describing demonstrations that had not occurred. Counterbalancing assignments in the previous study determined which specific events were experienced, suggested, or intrusions. To parse narratives into these categories, two coders first recorded any mention of the eight science demonstrations, the two touch events (wristband and sticker), and intrusions (nontarget events). Overall agreement (number

of agreements divided by the number of events recorded by at least one assistant) was 90.1%; kappas for the presence/absence of individual event components in answers to individual questions ranged from .77 to 1.00.

5–6. Yes–no questions—science demonstrations (proportion correct): the proportion of questions about experienced (experienced only or experienced and described in the book) and nonexperienced (suggested or control) science demonstrations that children answered correctly.

7–8. Yes–no questions—touching allegation (no "yes" responses = 0, one or two "yes" responses = 1): the percentage of children who alleged touching, broken down by those who had experienced touching and those who only heard about those events from the misleading story. (In Poole & Dickinson, 2011, children who acknowledged the wrist band but did not explicitly describe touching were not coded as having made a touching allegation for that event. Because descriptions of touching were infrequent, however, for this study we coded any "yes" responses to questions about the wristband and sticker as an allegation of touching.)

9–12. Yes–no questions—length of narratives: the length of children's descriptions when interviewers prompted for information following "yes" responses to yes–no questions, computed separately for experienced events, nonexperienced events, experienced touching, and nonexperienced touching.

13. Yes–no questions—the accuracy of descriptions of experienced events (proportion of narrative that was accurate). This variable captures detail errors (such as misreporting the color or names of objects), script-based errors (i.e., describing a well-known demonstration rather than the experienced demonstration), and other false descriptions triggered by questions about experienced events. Two coders classified each word in children's narratives as accurate or inaccurate; proportion agreement = 94.2%, $k = .78$.

14–17. Source-monitoring of previously-accepted events (proportion correct): the proportion of events that children correctly categorized as experienced or nonexperienced in response to source-monitoring questions, reported separately for experienced science events, nonexperienced science events, experienced touching, and nonexperienced touching. Note that interviewers asked source-monitoring questions only about events children said had occurred. Therefore, these scores represent how often children reaffirmed experienced events and rejected nonexperienced events when interviewers asked them to reflect on their earlier "yes" responses.

Recent event. 18. Free recall—length of narrative: the length of narratives in response to the initial request to recall what happened and a request to tell more.

19. Desired answers to focused questions (proportion correct): the proportion of the six event components that children accurately reported; intercoder agreement = 99.4%, $k = .99$.

Entire interview. 20. Answered without prompting (proportion of target questions): the proportion of 22 questions that children answered after first delivery of the question. (When children asked for clarification, interviewers' responses to this on-topic conversation were not considered prompts.) These included all questions until the first free narrative section and 10 questions about the Dog Lady event, starting with "We're going to talk about one more thing. Did you see Bonnie in the other room today?" Intercoder agreement for coding answers as not prompted, prompted, or other (e.g., codes for questions that were omitted because the child previously answered or denied the set-up question) was 98.7%, $k = .95$.

Behavioral inhibition scores

We measured inhibition to the unfamiliar with total scores from the Behavioral Inhibition Questionnaire (BIQ), a 30-item parental report measure. Bishop et al. (2003) provided data on the excellent psychometric properties of the BIQ, which support both an overall tendency toward behavioral inhibition and a degree of situational specificity (see also Kagan, Snidman, & Arcus, 1998). Cronbach's alpha (a measure of the extent to which items measure the same construct) from our sample was .96, which mirrored previous results (e.g., Kim et al., 2011).

Results

Two situations would prevent us from evaluating the impact of the drawing materials: (a) if children rarely interacted with those materials, or (b) performance measures were insensitive to the risks and benefits of comfort drawing due to floor effects (i.e., most children remembered very little) or ceiling effects (i.e., most children performed extremely well). To rule out these possibilities, we first analyzed the extent to which children assigned to the drawing condition interacted with the materials and then looked at whether our paradigm produced the significant effects of age and delay that would be expected based on prior memory studies. With these requirements satisfied, we explored the impact of interview condition, starting with moderated (or moderated logistic) regression analyses and then bracketing effect sizes with confidence intervals to better convey the distance between the two interviewing conditions.

Interactions with markers in the drawing condition

Most children took advantage of the opportunity to draw: In 110 video recorded sessions, only 4 children never interacted with the drawing materials (one 8-year-old, one 10-year-old, and two 11-year-olds), and the remaining children marked

on paper at some time during the majority of the 10 coded interview segments ($M = 8.31$, $SD = 2.52$). Younger children did draw during more interview segments than older children did, $r = -.23$, $p = .018$, but older children also frequently interacted with the drawing materials. For example, on average the 5-year-olds and 6-year-olds drew during 8.67 and 9.20 interview segments, respectively, compared to 7.81 for 9-year-olds and 8.12 for 10-year-olds.

Although some children were less interested in drawing than others, collectively the children retained interest in the drawing materials throughout the interview. For example, the percentages of children with some drawing behavior during the last three segments of the interview were 90%, 76%, and 87%, respectively. Keep in mind, however, that coders recorded drawing or marking behavior if children did so at any time during an interview segment, no matter how briefly. In fact, children frequently stopped drawing to listen to the interviewer and to answer questions.

Sensitivity of the measures to cognitive ability and task difficulty

Each of the 20 outcome measures was entered into a moderated regression (or logistic regression) analysis with age, delay between event and final interview (for the remote event only), and interviewing condition (no drawing and drawing) entered as the first step, followed by the 2-way and then 3-way interactions in the second and third steps. Twelve measures showed significant age trends: Compared to younger children, older children provided desired answers to a higher proportion of presubstantive prompts ($p < .001$), more often stated they remembered the science experience during topic introduction ($p = .003$), provided longer free narrative descriptions when asked about the remote event ($p < .001$) and produced a higher proportion of narrative describing the previous science experience ($p = .021$), were more accurate in response to yes–no questions about experienced science demonstrations ($p < .001$) and produced longer narratives about these events ($p = .001$), were more accurate in response to source-monitoring questions about experienced science events ($p = .001$) and nonexperienced science events ($p = .047$), and were less likely during source-monitoring to reaffirm a nonexperienced touching event ($p = .034$). Finally, the older children provided longer free narratives about the recent Dog Lady event ($p < .001$), gave desired responses to a higher proportion of the target details contained in this event ($p < .001$) and, throughout the experimental session, more often answered questions without prompting ($p < .001$).

Despite the fact that memories fade most quickly during the initial weeks after an event (e.g., White, 2001), 6 of the 17 measures pertaining to the remote event showed significant differences between the 1-year and 2-year delays. As expected, a longer delay was associated with less successful topic introduction ($p = .021$), shorter responses to open-ended questions ($p = .003$) with a lower proportion of narrative describing target events ($p = .004$), fewer accurate responses to yes–no questions about experienced science events ($p < .001$), fewer disclosures of experienced touching ($p = .024$), and lower accuracy on source-monitoring questions about experienced science events ($p = .004$). Only three measures were qualified by Age × Delay interactions (p values are for the interaction terms): a longer delay (and, therefore, older children) dampened age trends for the proportion of presubstantive prompts that children correctly answered ($p = .028$) and the proportion of questions they answered without prompting ($p = .009$), and there was no longer a significant age trend for the length of freely recalled narratives after a 2-year delay ($p = .004$).

Collectively, these results documented that our performance measures were sensitive to variation in cognitive ability and task difficulty. Importantly, children's level of performance on most measures was adequate to detect benefits or challenges attributable to comfort drawing (see Table 2).

The impact of comfort drawing

The entire sample. Table 2 describes children's performance in the no drawing and drawing conditions across the 20 performance measures. Main effects of interviewing condition from moderated regression (or logistic regression) analyses appear on the right side (βs and significance values); effect sizes are partial eta-squares for continuous measures (from analyses of covariance that controlled for age and delay) or log-odds (for logistic regression analyses). Confidence intervals for the differences between means (or percentages) and for effect sizes convey whether sample sizes were adequate to produce stable estimates.

Most striking is the similarity of performance across the two conditions, with no significant main effects despite the large number of outcome measures. Moreover, there was no consistent direction to nonsignificant differences that suggested an overall benefit from one interviewing condition or the other.

Three main effects were qualified by interactions involving condition. First was a 3-way interaction (age, delay, and condition) for the length of narratives produced in response to yes–no questions about nonexperienced science events, $p = .003$. (Children with longer narratives about nonexperienced events were more likely to draw from general knowledge to fill in memory gaps.) To help visualize this interaction, we dichotomized age (under 8.78, the median for the entire sample, or equal to or over 8.78) and conducted separate t-tests to compare interviewing conditions at each combination of age group and delay. The interaction resulted from nonsignificant trends for the older age group to provide more false descriptions in the drawing condition after the 1-year delay ($p = .09$) but more in the no drawing condition after the 2-year delay ($p = .052$).

There was also an Age by Condition interaction for the percentage of children who reported an experienced touching event: More of the younger children accurately reported touching in the drawing (73.91%) than no drawing condition (41.38%), $p = .02$, whereas means were nonsignificantly reversed for the older children, 53.13% versus 70.83%, respectively, $p = .13$. We caution that only half of the sample received touches in the prior study, so sample sizes were small for this

Table 2
Mean performance of children in the no drawing and drawing conditions.

	No drawing			Drawing			95% CI Δ	β	p	η_p^2[Exp(B)]	90%
	n	M or %		*n*	M or %						CI effect size
Early interview phases											
Presubstantive prompts (proportion correct)	108	0.98	(.05)	111	0.98	(0.05)	[−0.01, +0.01]	−.00	.91	.000	[.00, .00]
Remote event ("Mr. Science")											
Successful topic introduction (%)	108	94.44%		111	96.40%		[−4.80, +9.02]	.35	.62	[1.42]	[+.45, +4.54]
Free recall											
Length of narrative	108	101.44	(85.29)	111	103.69	(96.82)	[−26.57, +22.08]	−.73	.95	.000	[.00, .00]
Proportion of narrative describing experienced events	104	0.93	(0.21)	111	0.96	(0.14)	[−.07, +02]	.02	.35	.004	[.00, +.03]
Yes–no questions											
Experienced science events (proportion correct)	108	0.75	(0.28)	111	0.75	(0.26)	[−.07, +08]	−.02	.56	.002	[.00, +.02]
Length of narratives about experienced science events	101	20.59	(11.83)	108	22.92	(13.82)	[−5.84, +1.20]	2.28	.19	.008	[.00, +.04]
Nonexperienced science events (proportion correct)	108	0.76	(0.27)	111	0.71	(0.31)	[−.02, +13]	−.05	.19	.008	[.00, +.04]
Length of narratives about nonexperienced science events	60	14.00	(13.88)	66	15.52	(18.67)	[−7.37, +4.32]	1.24	.68	.001	[.00, +.03]
Allegation of experienced touching ("yes" to one or more events)	53	54.72%		55	61.82%		[−12.41, +25.96]	.25	.53	[1.29]	[+.66, +2.50]
Length of narratives about experienced touching	29	14.40	(8.31)	34	11.69	(9.34)	[−1.78, +7.19]	−2.61	.26	.021	[.00, +.11]
Allegation of nonexperienced touching ("yes" to one or more events)	55	30.91%		56	37.50%		[−12.07, +24.62]	.29	.47	[1.34]	[+.69, +2.61]
Length of narratives about non-experienced touching	17	15.41	(15.86)	21	13.71	(9.62)	[−6.76, +10.15]	−1.44	.74	.003	[.00, +.09]
Accuracy of descriptions of experienced events (proportion of narrative that was accurate)	100	0.87	(0.19)	109	0.88	(0.16)	[−.06, +04]	.01	.66	.001	[.00, +.02]
Source-monitoring											
Experienced science events (proportion correct)	101	0.93	(0.16)	108	0.94	(0.17)	[−.05, +04]	.01	.82	.000	[.00, +.01]
Nonexperienced science events (children who said "yes" earlier—proportion correct)	60	0.20	(0.38)	66	0.23	(0.40)	[−.17, +11]	.02	.73	.001	[.00, +.03]
Allegation of experienced touching ("yes" to one or more)	29	93.10%		34	82.35%		[−9.38, +29.17]	−1.01	.26	[.36]	[+.08, +1.61]
Allegation of nonexperienced touching ("yes" to one or more = 1)	17	76.47%		21	85.71%		[−18.67, +37.95]	1.03	.32	[2.79]	[+.52, +15.02]
Recent event ("The Dog Lady")											
Free recall: Length of narrative	108	85.97	(59.66)	111	88.17	(70.46)	[−19.61, +15.21]	.04	.996	.000	[.00, .00]
Desired answers to focused questions (proportion correct)	108	0.78	(0.18)	111	0.79	(0.17)	[−.06, +03]	.01	.64	.001	[.00, +.02]
Entire interview											
Answered without prompting (proportion of target questions)	108	0.96	(0.05)	111	0.96	(0.06)	[−.02, +01]	.00	.67	.001	[.00, +.02]

Note: Standard deviations are in parentheses.

subgroup of younger children ($n_{\text{no drawing}} = 29$, $n_{\text{drawing}} = 23$), resulting in a broad confidence interval for the mean difference, [+2.69%, +55.42%]. The percentage of children who falsely reported nonexperienced touching also produced a 3-way interaction, $p = .04$, but follow-up tests were not significant and trends across age groups and delays failed to paint a consistent picture. What is important given the previous finding is that 32.14% of young children reported suggested touching in the no drawing condition compared to 33.33% in the drawing condition. Thus yes–no questions are problematic for young children, but this problem does not appear to be exacerbated by drawing.

The higher disclosure rate among younger children in the drawing condition led us to explore whether drawing might have produced a "yes" bias. For the younger, touched children, there was no significant difference between interviewing conditions in the proportion of "yes" responses to the eight yes–no questions about science demonstrations ($M_{\text{no drawing}} = .50$, $M_{\text{drawing}} = .48$), $p = .85$, nor was there a significant difference for the sample as a whole ($M_{\text{no drawing}} = .50$, $M_{\text{drawing}} = .52$), $p = .56$.

Children 5–7 years. Professionals may still wonder if differences between conditions are greater for particular subgroups of children. Because some supporters and critics of comfort drawing are most interested in the developmental needs of young children, we reran analyses for children ages 5–7 years, which was a cut-off that produced reasonable sample sizes ($n_{\text{no drawing}} = 44$, $n_{\text{drawing}} = 40$). Again, performance was highly similar across conditions with one exception: Consistent with the previous follow-up result (and involving mostly the same children), fewer young children said "yes" to questions about experienced touching in the no drawing (34.78%) than drawing (73.68%) condition, $\beta = -2.55$, $p = .02$.

Uninhibited and inhibited children. It is also possible that temperament rather than age determines who benefits from emotionally supportive environments (see Quas & Lench, 2007). To investigate, we selected (a priori) seven performance measures to capture the overall quantity and accuracy of the information children provided about the remote event: the length of their freely-recalled narratives; the proportion of those narratives that described experienced events; the accuracy of responses to yes–no questions about experienced science demonstrations, nonexperienced science demonstrations, experienced touching, and nonexperienced touching; and the proportion of narrative in response to yes–no questions about experienced events that was accurate (i.e., the quantity of errors in reports of experienced events). We reran analyses for these measures with behavioral inhibition added as a predictor (centering predictor variables). Because behavioral inhibition scores varied widely (from 34 to 177, $M = 91.99$, $SD = 30.39$, median = 88.50) and did not correlate significantly with age, $r = .007$, $p = .92$, these analyses targeted a different group of children than did analyses of children 5–7 years of age.

As expected, behavioral inhibition was associated with eyewitness performance: Inhibited children talked less in response to initial open-ended questions, $p = .001$, and their descriptions of experienced events during yes–no questioning were more accurate, $p = .006$. However, none of the 2-way interactions between behavioral inhibition and interviewing condition reached significance, indicating that performance was similar across no drawing and drawing conditions regardless of whether children tended to be bold or inhibited.

Discussion

This study investigated whether helping children relax with a familiar, voluntary activity during interviews influences the quality of their autobiographical reports. We maximized our ability to detect differences between interview conditions by enrolling a large number of children who had participated in a previous study, balancing conditions for their prior performance, and asking a set of challenging questions (about an event that occurred 1 or 2 years earlier) along with questions about a recent event. Despite the fact that our procedures detected performance differences associated with age and, for the remote event, the amount of time that had transpired since the event, there was no overall impact of comfort drawing across 20 measures of eyewitness performance. This means that for children as a whole, comfort drawing did not significantly increase (or decrease) the amount of information they recalled, the accuracy of their answers, or even the extent to which interviewers needed to prompt for answers. Throughout the age range studied (5–12 years), children easily combined drawing and answering questions as evidenced by trivial effect sizes despite interest in the drawing materials.

Although drawing offered no advantages or disadvantages to inhibited children relative to their uninhibited peers, younger children in the drawing condition disclosed experienced touching more often than did younger children in the no drawing condition, and this result was not due to an overall tendency to say "yes" to yes–no questions. This finding requires replication, however, because we conducted a large set of analyses (so an isolated finding could be due to chance alone) and the sample size was small for this subset of children. If future studies confirm this result, then providing drawing materials to children would be one of the easiest ways to create a more emotionally supportive interviewing environment.

Studies of performance on other tasks help explain why comfort drawing had little impact on children's testimonial quality. For example, earlier we mentioned that young children regulate attention between television shows and ongoing play by monitoring sound-track features that signal when something interesting will happen. Importantly, their attention and memory for content is tightly linked to comprehensibility, such that they look (Anderson et al., 1981) and remember (Calvert et al., 1982) when salient auditory features signal content they understand. Similarly, we noticed that children in the present study typically drew in spurts and often stopped drawing to listen to and answer questions. Most likely, comfort drawing had little impact on performance because it was a voluntary activity that children could start and stop at will.

There are, however, several caveats to our results. First, few of our participants were under 6 years of age, whereas concerning interviews often involve younger children. Therefore, it is important to caution that comfort drawing is still an

untested practice for preschool and atypically-developing children. The literatures on cognitive development and children's eyewitness testimony have identified special limitations among these groups, including impaired ability to update working memory and poorer executive control (Carretti, Belacchi, & Cornoldi, 2009; Garon, Bryson, & Smith, 2008, see also Brown, Lewis, Lamb, & Stephens, 2012), and these characteristics could alter the impact of letting children draw during interviews. Therefore, we encourage additional research targeting these populations to provide interviewers with more guidance about when it is appropriate to customize interviews by adding comfort drawing.

Finally, although many findings from research involving staged events have generalized to interviews about naturally-experienced events (e.g., differences in the amount and accuracy of information elicited with various question types; Fivush, Peterson, & Schwarzmueller, 2002; Lamb et al., 2008), there is an important difference between our procedures and abuse interviews: The children in our study had no emotional reasons for withholding information, unlike many victims of sexual abuse. For reluctant abuse victims, drawing could encourage disclosures beyond the younger children who benefitted in our study (by reducing anxiety and feelings of embarrassment) or, potentially, impede disclosures (by providing a diversionary tactic in which children can refocus conversation to their drawings). The later possibility does not negate the current findings, however, because interviewers who choose to provide drawing materials can simply remove these items from children who prefer to talk about their drawings. By showing that voluntary drawing did not hinder school-aged children's abilities in a challenging memory task, our study alleviates concern about the cognitive consequences of this potentially-distracting activity.

References

Anderson, D. R., & Levin, S. R. (1976). Young children's attention to Sesame Street. *Child Development*, *47*, 806–811. http://dx.doi.org/10.2307/1128198

Anderson, D. R., Lorch, E. P., Field, D. E., & Sanders, J. (1981). The effects of TV program comprehensibility on preschool children's visual attention to television. *Child Development*, *52*, 151–157. http://dx.doi.org/10.2307/1129224

Anderson, J., Ellefson, J., Lashley, J., Miller, A. L., Olinger, S., Russell, A., Stauffer, J., & Weigman, J. (2010). The CornerHouse forensic interview protocol: RATAC®. Thomas M. Cooley. *Journal of Practical and Clinical Law*, *12*, 193–331.

Anderson, M., Bucks, R. S., & Bayliss, S. D. S. (2011). Effect of age on dual-task performance in children and adults. *Memory & Cognition*, *39*, 1241–1252. http://dx.doi.org/10.3758/s13421-011-0099-7

Bishop, G., Spence, S. H., & McDonald, C. (2003). Can parents and teachers provide a reliable and valid report of behavioral inhibition? *Child Development*, *74*, 1899–1917. http://dx.doi.org/10.1046/j. 1467-8624.2003.00645.x

Bottoms, B. L., Quas, J. A., & Davis, S. L. (2007). The influence of the interviewer-provided social support on children's suggestibility, memory, and disclosures. In M.-E. Pipe, M. E. Lamb, Y. Orbach, & A.-C. Cederborg (Eds.), *Child sexual abuse: Disclosure, delay, and denial* (pp. 135–157). New York, NY: Routledge.

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

Calvert, S. L., Huston, A. C., Watkins, B. A., & Wright, J. C. (1982). The relation between selective attention to television forms and children's comprehension of content. *Child Development*, *53*, 601–610. http://dx.doi.org/10.2307/1129371

Carretti, B., Belacchi, C., & Cornoldi, C. (2009). Difficulties in working memory updating in individuals with intellectual disability. *Journal of Intellectual Disability Research*, *54*, 337–345. http://dx.doi.org/10.1111/j. 1365-2788.2010.01267.x

Ciaramelli, E., Ghetti, S., & Borsotti, M. (2009). Divided attention during retrieval suppresses false recognition in confabulation. *Cortex*, *45*, 141–153. http://dx.doi.org/10.1016/j.cortex.2007.10.006

Dickinson, J. J., & Poole, D. A. (2000). Efficient coding of eyewitness narratives: A comparison of syntactic unit and word count procedures. *Behavior Research Methods, Instruments, and Computers*, *32*, 537–545. http://dx.doi.org/10.3758/BF03200826

Fivush, R., Peterson, C., & Schwarzmueller, A. (2002). Questions and answers: The credibility of child witnesses in the context of specific questioning techniques. In M. Eisen, J. A. Quas, & G. S. Goodman (Eds.), *Memory and suggestibility in the forensic interview* (pp. 331–354). Mahwah, NJ: Erlbaum.

Garon, M., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*, 31–60. http://dx.doi.org/10.1037/0033-2909.134.1.31

Hershkowitz, I., Orbach, Y., Lamb, M. E., Sternberg, K. J., Pipe, M.-E., & Horowitz, D. (2007). Suspected victims of abuse who do not make allegations: An analysis of their interactions with forensic interviewers. In M.-E. Pipe, M. E. Lamb, Y. Orbach, & A.-C. Cederborg (Eds.), *Child sexual abuse: Disclosure, delay, and denial* (pp. 97–113). New York, NY: Routledge.

Kagan, J., Reznick, S., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development*, *58*, 1459–1473. http://dx.doi.org/10.2307/1130685

Kagan, J., Snidman, N., & Arcus, D. (1998). Childhood derivatives of high and low reactivity in infancy. *Child Development*, *69*, 1483–1493. http://dx.doi.org/10.2307/1132126

Kim, J., Klein, D. N., Olino, T. M., Dyson, M. W., Dougherty, L. R., & Durbin, C. E. (2011). Psychometric properties of the Behavioral Inhibition Questionnaire in preschool children. *Journal of Personality Assessment*, *93*, 545–555. http://dx.doi.org/10.1080/00223891.2011.608756

Lamb, M. E., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2008). *Tell me what happened: Structured investigative interviews of child victims and witnesses*. West Sussex, England: Wiley.

Peterson, C., Dowden, C., & Tobin, J. (1999). Interviewing preschoolers: Comparisons of yes/no and wh- questions. *Law and Human Behavior*, *23*, 539–555. http://dx.doi.org/10.1023/A:1022396112719

Pipe, M.-E., & Salmon, K. (2009). Dolls, drawing, body diagrams, and other props: The role of props in investigative interviews. In K. Kuehnle, & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 365–395). Hoboken, NJ: Wiley.

Poole, D. A., & Dickinson, J. J. (2011). Evidence supporting restrictions on uses of body diagrams in forensic interviews. *Child Abuse & Neglect*, *35*, 654–669. http://dx.doi.org/10.1016/j.chiabu.2011.05.004

Poole, D. A., & Dickinson, J. J. (2013). Investigative interviews of children. In R. Holliday, & T. Marche (Eds.), *Child forensic psychology: Victim and eyewitness memory* (pp. 157–178). Hampshire, England: Palgrave Macmillan.

Poole, D. A., & Lindsay, D. S. (2001). Children's eyewitness reports after exposure to misinformation from parents. *Journal of Experimental Psychology: Applied*, *7*, 27–50. http://dx.doi.org/10.1037/1076-898X.7.1.27

Quas, J. A., Bauer, A. B., & Boyce, W. T. (2004). Physiological reactivity, social support, and memory in early childhood. *Child Development*, *75*, 797–814. http://dx.doi.org/10.1111/j. 1467-8624.2004.00707.x

Quas, J. A., & Lench, H. C. (2007). Arousal at encoding, arousal at retrieval, interviewer support, and children's memory for a mild stressor. *Applied Cognitive Psychology*, *21*, 289–305. http://dx.doi.org/10.1002/acp.1279

U.K. Ministry of Justice. (2007). *Achieving best evidence in criminal proceedings: Guidance on interviewing victims and witnesses, and using special measures*. London, England: Author.

White, K. G. (2001). Forgetting functions. *Animal Learning & Behavior*, *29*, 193–207. http://dx.doi.org/10.3758/BF03192887