Journal of Experimental Psychology: Applied
2002, Vol. 8, No. 4, 233–246

Copyright 2002 by the American Psychological Association, Inc.
1076-898X/02/$5.00 DOI: 10.1037/1076-898X.8.4.233

# Measurement of Individual Differences in Children's Suggestibility Across Situations

Matthew H. Scullin, Tomoe Kanaya, and Stephen J. Ceci
Cornell University

The authors attempted to use scores on the Video Suggestibility Scale for Children (VSSC, M. H. Scullin & S. J. Ceci, 2001) to predict 50 preschool children's performance during a field study in which they were interviewed suggestively 4 times about both a true event and a suggested event. Among the 25 children over age 4 years 6 months, tendencies on the VSSC to respond affirmatively to suggestive questions ("yield"), change answers in response to negative feedback ("shift"), and the sum of these ("total suggestibility") were all related to lack of accuracy about the true event in the field study and to both accuracy and lack of accuracy about the suggested event. Results support a 2-factor model of suggestibility.

Since Alfred Binet's pioneering work on suggestibility nearly 100 years ago (Cunningham, 1988), psychological researchers have conducted numerous studies to identify external or "situational" variables that affect children's testimony (see Ceci & Bruck, 1993, 1995, for reviews). More recently, researchers have begun to examine the characteristics within children that cause some to be more accurate than others (for reviews see Bruck, Ceci, & Melnyk, 1997; Quas, Qin, Schaaf, & Goodman, 1997). According to these reviews, there is a growing consensus that children's suggestibility entails an interplay between individual characteristics and situational factors and that further research is needed to examine this relationship in greater detail.

There are at least four reasons for examining the interplay between developmental, situational, and individual characteristics in detail: First, young children have been found to be more suggestible than adults or older children, with preschool children being the most suggestible (Ceci & Bruck, 1993). However, even among very young children there is still a great deal of variability, with some younger children being less suggestible than some older children. It is unclear how much of this variability is related to relatively stable individual differences among children of the same age—for example, in general cognitive functioning—and how much may be due to maturational differences in specific cognitive attainments such as theory of mind (Templeton & Wilcox, 2000; Welch-Ross, 1999), source monitoring ability (Lindsay, Johnson, & Kwon, 1991), or executive control (Perner & Lang, 1999).

Second, there is a great deal of interest among lawyers, caseworkers, and psychologists for a means of determining the degree to which a particular child is vulnerable to leading questions and social pressure (Quas et al., 1997). Because of the large number of young children entangled in the juvenile, family, and criminal justice systems as a consequence of abuse and neglect proceedings, acrimonious custody disputes, and Persons in Need of Supervision actions (Ceci & Friedman, 2000), it is useful to examine the role of individual differences in suggestibility. This is because investigators are less interested in being told that young children are, on average, more suggestible than older children, and instead prefer to be told whether a particular child is suggestible. A valid measure of individual differences in suggestibility could help identify children who are prone to making erroneous reports following suggestions and provide important information about when special interviewing precautions should be taken.

Third, forensic studies have found that adults who make false confessions score higher on a psychometric measure of interrogative suggestibility than do control groups (Gudjonsson, 1992b). This finding suggests that it may be fruitful to use a similar approach to study differences in suggestibility of children.

Finally, the Video Suggestibility Scale for Children (VSSC), comparable in some psychometric respects to adult suggestibility scales (Gudjonsson, 1984, 1987), was recently developed (Scullin & Ceci, 2001). However, it is unclear whether the scale has construct validity. The study reported here is an attempt to establish the VSSC's construct validity by comparing participants' performance with a variety of measures derived from a separate study in which the same children were interviewed suggestively about both real and imaginary events.

In what follows, we briefly discuss the nature of suggestibility and how interviewer bias and suggestive interviewing techniques

Matthew H. Scullin, Tomoe Kanaya, and Stephen J. Ceci, Department of Human Development, Cornell University.

This research was supported in part by grants from the College of Human Ecology at Cornell University. We thank Angela Crossman, Livia Gilstrap, and Heather Warren for their contributions to this study, and Jersy Chen, Amar Gavhane, Stephanie Linkous, Jessica Passero, Nina Piccoli, and Lisa Tallon for their data collection and coding. We thank the teachers and parents at Ithaca Community Childcare Center and IACC in Ithaca, New York, Happy Way Childcare Center in Dryden, New York, and Newfield Day Care, Newfield, New York, for their participation in this research.

Correspondence concerning this article should be addressed to Matthew H. Scullin, who is now at the Department of Psychology, West Virginia University, Morgantown, West Virginia 26506. E-mail: mhscullin@mail.wvu.edu

affect the reliability of children's testimony. Then we discuss the investigation of individual differences in suggestibility.

## Defining Suggestibility

When an individual tries to recall an event, information received both before and after the event can affect how it is recalled. Against this constructivist backdrop, Ceci and Bruck (1993) conceptualized suggestibility as "the degree to which children's encoding, storage, retrieval, and reporting of events can be influenced by a range of social and psychological factors" (p. 404). This is a broader view of suggestibility than Gudjonsson and Clark's (1986) model of interrogative suggestibility, in which *suggestibility* is defined as "the extent to which, within a closed social interaction, people come to accept messages communicated during formal questioning, as the result of which their subsequent behavioural response is affected" (p. 84). The Gudjonsson and Clark model placed an emphasis on coping strategies a witness develops to deal with the uncertainty and expectations of an interrogation and the incorporation of postevent information into memory. Ceci and Bruck additionally emphasized the possibility that information the child receives preceding an event may affect the events' encoding, storage, and retrieval. Further, their conceptualization of suggestibility goes beyond coping strategies to include conscious acquiescence and lying as well as destructive updating of memory due to incorporation of misinformation into the memory trace.

Under optimal conditions, children's memory can be highly accurate (Baker-Ward, Gordon, Ornstein, Larus, & Clubb, 1993; Goodman, Bottoms, Schwartz-Kenney, & Rudy, 1991). A single leading question by an otherwise neutral interviewer is not usually sufficient to produce an inaccurate response from a child witness. Warning a child that questions may be tricky and explaining when to say "I don't know" can reduce suggestibility, at least to some extent (Endres, Poggenpohl, & Erben, 1999; Warren, Hulse-Trotter, & Tubbs, 1991). However, the use of strong social incentives by an interviewer, such as telling a child what other children have provided the interviewer, or strongly praising or criticizing the child for an answer, results in highly elevated suggestibility scores. In one study, 3- to 6-year-old children responded affirmatively to a mean of 58% of eight leading questions within a single interview when given social incentives compared with only a mean of 17% of leading questions among children who were simply asked the leading questions (Garven, Wood, Malpass, & Shaw, 1998). This sort of feedback from an interviewer can occur as the result of a constellation of interviewer attitudes, beliefs, and behaviors that have been referred to as *interviewer bias*, characterized by the relentless pursuit of a single hypothesis by an interviewer who is uninterested in testing alternative hunches and consequently uses very strong or persistent suggestions (Bruck, Ceci, & Hembrooke, in press; Ceci & Bruck, 1995).

When a child provides information that is inconsistent with a biased interviewer's beliefs, the child's response is ignored and the interviewer may convey to the child that the response was incorrect. Studies have shown that large suggestibility effects can be found when children are confronted by a biased interviewer who barrages them with implicit and explicit suggestions, repetitions of questions, and the creation of expectations about what a proper answer should be (e.g., Garven et al., 1998). The strong effects of social and situational influences on suggestibility imply that individual difference measures of suggestibility will be imperfect measures across situations, as situational variance may sometimes override variance associated with individual differences.

## Measurement of Individual Differences in Children's Suggestibility

The pioneering research in the study of interrogative suggestibility was conducted by Gudjonsson (1984, 1987), who developed two parallel, interchangeable forms of the Gudjonsson Suggestibility Scale that have been used extensively in forensic research but primarily on adults and adolescents. In Gudjonsson's scales, a brief story is played on audiotape. Immediately afterward, the participant is asked to recall as much of the story as possible, with each of the salient points contributing to a Memory Recall score. The participant is then asked a series of 20 questions, 15 of which consist of varieties of inaccurate leading questions and 5 of which are "true," calling for an affirmative response to something that actually happened in the story. Each affirmative response to an inaccurate leading question is scored as a yield. After participants have answered all of the questions, they are given negative feedback about their performance (e.g., "You made some errors so I'm going to ask the questions again.") and the interviewer reads all the questions again. This leads to the second measure, Shift, which assesses the number of answers to the 15 inaccurate leading questions that were subsequently changed in response to the negative feedback. Total Suggestibility comprises Yield and Shift subscales. Factor analyses have shown that Yield and Shift items load on different factors, supporting Gudjonsson's view that there are two types of interrogative suggestibility that correspond to traitlike characteristics in the degree to which people give in to misleading questions (yield) and respond to negative feedback (shift) (Gudjonsson, 1984, 1992a).

When these scales have been used in developmental analyses, there is a decline in total suggestibility from adolescence to adulthood (reviewed in Gudjonsson, 1992b). This decline in suggestibility is not uniform, because although adolescents tend to shift more than adults, they do not yield more (Richardson, Gudjonsson, & Kelly, 1995). Using Gudjonsson's scale with even younger participants, Danielsdottir, Sigurgeirsdottir, Einarsdottir, and Haraldsson (1993) found a decline in yield and total suggestibility between the ages of 6 and 12. Warren et al. (1991) also found a decrease in suggestibility with age among 7-year-olds, 12-year-olds, and adults.

The finding of a decline in suggestibility with age in these studies is consistent with Ceci and Bruck's (1995) review of a wide variety of research that has consistently found a negative correlation between age and suggestibility. This decline is especially well-documented between the ages of 3 and 6. Although Gudjonsson's scale has been used with children as young as 6 years old, the use of an audiotaped story makes its use with younger children problematic. Notably, Warren et al. (1991) played the audiotaped story twice for the 7-year-old children in their study to ensure that they understood and followed it. However, reading a story twice to children has been found to greatly reduce suggestibility by increasing memory trace strength (Endres

et al., 1999). Therefore, to develop a fuller picture of children's suggestibility and investigate individual differences in suggestibility in preschool children, researchers have recently developed suggestibility scales specifically for young children (Endres et al., 1999; Candel, Merckelbach, & Muris, 2000; Scullin & Ceci, 2001). In this study, we discuss an attempt to validate one of these, the VSSC (Scullin & Ceci, 2001), which draws upon and extends Gudjonsson's research.

## The VSSC

In developing the VSSC, Scullin and Ceci (2001) retained Gudjonsson's concepts of memory recall, yield, shift, and total suggestibility. To extend the examination of individual differences to a younger age than had been previously addressed, a number of changes were introduced to make the scale more developmentally appropriate and consistent with forensic demands on children. A video (rather than an audiotape) was used to tell a story that would be encoded in a manner similar to that of an event that a child might actually witness or experience (i.e., involving both auditory and visual encoding). Scale questions were administered at intervals of 1 day to 1 week after the story (rather than immediately after the story, as is the case of Gudjonsson's scale) to better reflect the retrieval of memories some time after an event occurs, thus allowing realistic time for forgetting. Additionally, children were twice given mild negative feedback (e.g., "You missed a few of the questions. Let's go through them again and see if you can do better this time.") to keep the feedback salient, with half the questions repeated each time.

When administered to a sample of 51 3-year-olds, 98 4-year-olds, and 39 5-year-olds, the VSSC showed levels of internal consistency higher than those obtained for the Gudjonsson Suggestibility Scale (Cronbach's $\alpha = .85$ for yield and .75 for shift for the VSSC vs. .77 for yield and .67 for shift reported by Gudjonsson, 1984). Yield and Shift items loaded on well-defined separate factors when scale items were analyzed together (Scullin & Ceci, 2001). Following a promax (oblique) rotation, the Yield and Shift factors correlated at −.16. All Yield items loaded greater than .43 on the first factor, and all Shift items loaded more heavily on the second factor than the first, though only 10 of the 18 Shift items loaded greater than .40 on the second factor. Scullin and Ceci (2001) reported that there were no statistically significant differences on Memory Recall, Yield, or Shift between boys and girls, and varying the length of time between viewing the video and administering the scale between 1 day and 7 days did not result in statistically significant different scores on any of the scale measures. As children got older, each age group recalled significantly more details than the next younger group in memory recall (*M*s = 1.6, 2.5, and 3.5 details for 3-, 4-, and 5-year-olds, respectively; with Cohen's *d* for the difference between 3- and 4-year-olds = 0.47 and Cohen's *d* for the difference between 4- and 5-year-olds = 0.50). Children in the 3-year-old group had lower Shift scores (i.e., changed their answers less in response to negative feedback) than 4- and 5-year-olds (*M*s = 3.1, 4.7, and 4.7 for 3-, 4-, and 5-year-olds, respectively; with Cohen's $d = 0.51$ for both the difference between 3- and 4-year-olds, and 3- and 5-year-olds). Thus, although children's memory for the video increased between the ages of 3 and 5, their sensitivity to social pressures to change their answers also appeared to increase.

In summary, the Yield and Shift subscales tap into two different traitlike characteristics that correspond to a tendency to respond affirmatively to leading questions and a tendency to be socially sensitive to negative feedback (Gudjonsson, 1992b; Register & Kihlstrom, 1988). The Yield part of the scale contains both a cognitive component (how well the child remembers the video) and a social component (a need to provide affirmative responses to the interviewer in the absence of feedback). Shift also contains a cognitive component (children may be less likely to shift about a well-remembered detail), although it is thought to be largely reflective of a social component (some children are more sensitive to negative feedback). Because the Yield and Shift subscales are largely uncorrelated with each other but each has been positively correlated with suggestibility, Gudjonsson (1992b) has hypothesized that the subscales add incremental predictive validity to each other. The present study was designed to test this hypothesis and others described below by examining the construct validity of the VSSC measures and by evaluating the relationship between the subscales and quantitative measures of children's suggestibility. The latter were derived from an independent field study that was developed to examine interviewer bias in poorly conducted forensic interviews on children's reports of true and suggested events.

## Overview

This study had three main objectives: (a) to evaluate the construct validity of the four measures assessed through the VSSC subscales (Memory Recall, Yield, Shift, and Total Suggestibility) by examining whether children who scored high on Memory Recall also remembered an experienced true event in a field study in which children were suggestively interviewed about both true events and suggested events and to evaluate whether children who scored high on Yield, Shift, and Total Suggestibility also had elevated levels of suggestibility for both true and suggested events in the field study; (b) to determine whether the measures of yield and shift predict levels of suggestibility about both true events and suggested events independently of the child's age; and (c) to explore whether yield and shift predict whether children still assent to a suggested event when they are interviewed using forensic interview techniques that have been shown to increase the amount of accurate recall among second graders (McCauley & Fisher, 1995). We evaluated eight hypotheses, one relating to the nature of the field study manipulation and the other seven evaluating the relationship between the VSSC measures and suggestibility and inaccuracy in the field study, which are outlined in detail below.

### *Hypothesis 1: True Events Versus Suggested Events*

Each child was repeatedly interviewed about a true event (one they participated in) and a suggested event (one they were told they participated in). Because children are expected to pick up some accurate information about the true event because they actually participated in it, Hypothesis 1 states that children will be more accurate about the true event than the suggested event and more inaccurate about the suggested event than the true event.

### *Hypothesis 2: Age Differences in the Predictive Ability of the VSSC*

There are at least three developmental reasons to expect that predictive ability of the scale measures will be better for older children. In the initial study of the VSSC (Scullin & Ceci, 2001), there was a linear increase in children's performance on the Memory Recall scale for each year in age approaching Cohen's criteria for a large effect size (mean Cohen's $d = 0.49$ per year between the ages of 3 and 5). Because 3-year-olds were near floor level in this study ($M = 1.55$, $SD = 1.69$), lack of variability may contribute to lack of predictive ability for this measure for younger children.

Second, a recent meta-analysis of 178 studies that examined the development of children's theory of mind found that it was at the age of 4 years 6 months that children begin to perform significantly above chance in tasks that require an understanding that others can hold false beliefs (Wellman, Cross, & Watson, 2001). A key component of the Shift measure of the VSSC is the ability to reflect on one's own memory in light of another person's conflicting beliefs. This may require a level of metacognitive sophistication that is beyond the ability of children below the age of 4 years 6 months.

Finally, developing in tandem with theory of mind is children's executive function or self-control, the ability to suppress a response to bring about a desired outcome (Perner & Lang, 1999). Inability to suppress an initial response may affect the VSSC Yield subscale by preventing children from taking the time to search their memory for an answer and also affect the VSSC Shift subscale by preventing them from strategically evaluating their earlier responses. Thus, Hypothesis 2 is that younger children will respond to the VSSC in a systematically different way than older children, particularly during memory recall and shift, and this in turn will affect how well the scale correlates with the validation criteria. These age differences may be reflected in an improvement in how well the scale predicts accuracy and suggestibility for children after the age of 4 years 6 months.

### *Hypothesis 3: Accuracy About a True Event*

The Memory Recall scale of the VSSC is a measure of a child's ability to verbalize accurately about an experienced birthday party video event, so we expect this measure to correlate with verbal recall about a different true event in response to both open-ended and probe questions. Hypothesis 3 states that the VSSC Memory Recall scale will be related to accuracy about the true event. We had no specific hypothesis about the magnitude or directionality of the relationships between yield, shift, and total suggestibility with accuracy about the true event.

### *Hypothesis 4: Inaccuracy About a True Event*

The Yield, Shift, and Total Suggestibility scales are measures of suggestibility and inaccuracy about a true event, so Hypothesis 4 asserts that higher scores on these scales are positively related to inaccuracy about the true field event.

### *Hypothesis 5: Accuracy About a Suggested Event*

The ability to accurately describe a nonexperienced event to an interviewer is an important element of suggestibility, and some children have been shown to generate believable, detailed stories about nonexperienced events with minimal prompting (Ceci, Loftus, Leichtman, & Bruck, 1994; Ceci, Huffman, Smith, & Loftus, 1994). Thus, Hypothesis 5 is that the VSSC Yield, Shift, and Total Suggestibility measures are positively related to a child's ability to provide accurate information about a nonexperienced event.

### *Hypothesis 6: Inaccuracy About a Suggested Event*

Over the course of the interviews, children also received some inaccurate information about the suggested event from the interviewer. Because the child was not at this event, the child's imagination could also be a source for inaccuracy about the suggested event. Hypothesis 6 is that yield, shift, and total suggestibility are positively related to inaccuracy about the suggested event.

### *Hypothesis 7: Incremental Predictive Ability of Yield and Shift Above Age*

Children's memory for events gets better as they get older and their suggestibility decreases (Ceci & Bruck, 1993). We thus sought out to evaluate whether the Yield and Shift scales are a better predictor of a child's suggestibility than simply knowing the child's age. Hypothesis 7 is that Yield and Shift scores add incrementally to the predictive power of age in months, specifically in predicting summary indices of children's suggestibility for the true and suggested events.

### *Hypothesis 8: Predicting Assent to a Suggested Event in a Forensic Interview*

In the final interview of the field study, we interviewed children with techniques derived from the Cognitive Interview for children, which uses an interviewing strategy in which the interviewer attempts to help the child reinstate the context of an event (Fisher & Geiselman, 1992; Köhnken, 1993). The Cognitive Interview has been shown to help 2nd-grade children recall details of an event without increasing the proportion of inaccurate information (McCauley & Fisher, 1995). When the Cognitive Interview was administered to 8- to 9-year-old witnesses who viewed a videotaped event and were then asked leading questions about the event (Memon, Holley, Wark, & Bull, 1996), they found that to some extent the Cognitive Interview "inoculated" the children from the deleterious effects of prior suggestive questions. An issue we sought to address is whether the VSSC, as a measure of individual differences in suggestibility, can predict which children will continue to report that they participated in a suggested event when a forensic interviewing technique designed for use with children is implemented. Hypothesis 8 states that the VSSC can predict who assents to the suggested event when a forensic interviewing technique is used to elicit free recall after the child has undergone repeated suggestive interviews.

## Method

### *Participants*

Sixty-six children who completed a prior repeated-interviews suggestibility study were recruited as participants. Because of various circumstances (time constraints, children's illnesses, etc.), several children were not able to complete all portions of the study, reducing the sample size to 25 children in our younger group (age 4 years 6 months and below; 10 boys and 15 girls, mean age = 48.76 months, *SD* = 4.40), and 25 children in our older group (over age 4 years 6 months; 16 boys and 9 girls, mean age = 60.40 months, *SD* = 4.70). All participants were recruited from four different day-care centers and preschools, and informed consent was obtained from a parent or guardian of each child prior to the child's participation in the project.

### *Procedure*

#### *Phase 1: Repeated Interview Study*

Children's reports of both a true event in which they participated and a suggested event in which they were falsely told they participated were generated through the use of repeated suggestive interviews. In a fully counterbalanced design, 25 of the children were randomly placed in the monkey-true event condition, consisting of actually helping a stranger find her son's (stuffed) monkey in a location outside the classroom. For the children in this condition, the suggested event consisted of falsely suggesting that they helped a stranger carry some Play-Doh down the hallway to another room. The stranger tripped, it was suggested, and hurt her ankle, which was bandaged by a nurse who happened to be in the hallway at the time. The other 25 children were assigned to the Play-Doh true event condition, consisting of participation in the Play-Doh event. For this group, the suggested event was the false suggestion that they participated in the lost monkey event. Children were interviewed about both the true event and the suggested event in counterbalanced order on four different occasions, each occurring approximately 1 week apart.

In sum, each of the 50 children received four weekly interviews. During each of the first three of these weekly interviews, they were queried about true and false events. To control for event-specific responses, the true event for one group was the false event for the other group.

The same interviewer conducted the first three interviews, during which children were given a brief description by the interviewer of both of the events prior to questioning. In the first interview, this description included mostly accurate information, with three pieces of misinformation per event (e.g., the lady with the monkey might be incorrectly described as having black hair, wearing a red sweatshirt, and giving children a present). Following this description, children were simply asked if each event had happened to them ("Did you ever help a lady who dropped some Play-Doh and then fell and hurt herself?" and "Did you ever help a lady find a monkey?"), and if so, to provide some details. The second and third interviews also began with descriptions of the events that provided children with additional true and false details and a chronology of the events ("The lady [with the Play-Doh] spilled some orange juice too and it made a big mess."). For both the true event and the suggested event for each child, some of the information was accurate and some was inaccurate, but because the child was not at the suggested event even the accurate information may be regarded as misleading. Suggestive techniques used by the interviewers in the first three interviews included mentions of peer conformity (e.g., "Sally and Carol told me you were one of the kids who helped the nurse."); selective reinforcement (e.g., saying "Wow, you are so helpful" when the child provided desired information); and leading questions (e.g., "Was Erin the name of the lady with the monkey?"). The questioning part of the second and third interview began with free recall, followed by a variety of scripted probe questions, some of which imparted accurate information (about the true event), such as the monkey lady's name, and others imparted inaccurate information (e.g., about the monkey lady's appearance). The probe questions contained a mix of correct "yes" (for the true event) questions and correct "no" questions, along with open-ended probe questions.

The fourth interview was conducted by a different interviewer who questioned the children about the events using the Cognitive Interview's context reinstatement techniques. These entail the interviewer emphasizing to the child that she (the interviewer) had not been at the event and did not know what had happened. Next, she asked the child to visualize the scene of the event (e.g., "Picture yourself back in the playground on the day when the lady asked you to help her find her monkey. Can you see the place now?"). Following the visualization request, the child was asked for free recall. If the child did not assent to the event in his or her free recall, the interviewers discontinued the interview. If the child assented to the event, the interviewer proceeded with some scripted open-ended questions about the event. The interviewer attempted to create a picture in the child's mind of an aspect of the scene and then asked some probe questions about that aspect (e.g., "Picture the lady with the monkey." "What was the lady wearing?" "What did the lady say?" "What did the lady do?"). As discussed later, these aspects of context reinstatement from the Cognitive Interview are not without some suggestiveness, because the enjoinder to create mental pictures could lead to source confusions. Nevertheless, the findings of McCauley and Fisher (1995) indicate that they lead to enhanced accuracy.

We transcribed all four interviews with each child from videotapes with the assistance of interviewer notes. The monkey event and Play-Doh events were collated such that all four monkey interviews and all four Play-Doh interviews for each child could be read and coded separately. This resulted in 50 sets each of the monkey and Play-Doh interview transcripts. Because each child only participated in one of the events, within each set of 50 transcripts half the transcripts were about a true event and half were about a false or suggested event. Two independent, trained undergraduate coders who were blind to the child's gender, event condition, and how the child had been interviewed about both events coded all 100 transcripts. Interrater reliability ranged from .89 to .99 for all measures. We resolved discrepancies through discussion after all of the transcripts had been coded. Children's utterances were coded on the basis of accuracy, inaccuracy, and how often the children denied something the interviewer said, using checklists prepared for this purpose. Definitions of all dependent measures are reported below.

#### *Phase 1: Coding Scheme and Measures*

*Total accurate and inaccurate utterances.* The major unit of analysis of children's statements was the "utterance idea," or *utterance* for short. All statements beyond a simple "yes," "no," or "I don't know" response were coded at the utterance level. At the most general level, an utterance contains a verb and is bounded by a pause. An utterance could have an implicit verb if it was in response to a question. For example, if a child said "blue with white stripes" in response to the question "What color was the lady's shirt?" this was counted as an utterance. If an utterance was repeated or paraphrased, it was only counted once. For example, when a child said at different points within an interview, "She fell down on the bridge," "She fell down," and "Then she fell down," it was counted as a single utterance idea.

Utterances were coded as accurate when the child provided correct information to the interviewer and inaccurate when the child provided mistaken or incorrect information to the interviewer. Because the coders were blind about whether an interview they were coding was about a true event or a suggested event, children who mentioned accurate information about their suggested event (e.g., either by guessing or by repeating things that had been mentioned by the interviewer) would receive points for accuracy. When some of the child's response was correct but other ele-

ments incorrect or unclear, it was coded as mixed (e.g., in the statement "Ricky found the monkey under the monkey bars," when Ricky really did find the monkey, but it was in a toy car rather than under the monkey bars). When information was on-topic but not possible to verify, it was coded as unclear (e.g., "I looked for the monkey under the slide," when there is no way of knowing whether this occurred). Finally, an utterance was coded as a denial when the child denied that the event took place or denied the premise of an interviewer's question (e.g., "I wasn't there." or "There was no monkey."). These denial utterances were not evaluated for accuracy because the coder did not know whether the child had been at the event. Even if a child denied being at an event in free recall in Interviews 2 and 3, the interviewer asked the child probe questions about the event, and we evaluated children's responses to these questions.

For this article, we examined only accurate and inaccurate utterances, as they provide the clearest indicators of a child's accuracy. Because children tended to provide relatively little information in response to free-recall questions, we analyzed these separately from utterances made in response to probe questions. All children were asked free-recall questions in all four interviews. Probe questions were asked of all children in the second and third interviews, but in the fourth interview probe questions were only asked of children who first assented to having participated in an event during their free recall. To control for the number of questions asked, we only looked at responses to probe questions in Interviews 2 and 3. For the true event, we created a composite total-inaccurate-utterances index consisting of the sum of freely recalled inaccurate utterances in Interviews 1–4 as well as probed inaccurate utterances in Interviews 2 and 3. For the suggested event, we summed free-recall total accurate utterances and free-recall total inaccurate utterances in Interviews 1–4, plus probe-question accurate and probe-question inaccurate utterances in Interviews 2 and 3 for our composite inaccurate utterance index. This was done because even utterances coded as accurate by raters for the suggested event could actually be considered inaccurate because the child was not at the event, thus requiring the formation of this composite.

*Accurate and inaccurate interviewer-supplied details.* All of the accurate and inaccurate information mentioned to the child by the interviewer in all the interviews was listed in a checklist. If the child mentioned any of this information in an utterance, it was checked off. Children only received credit for providing information if they actively did so. Therefore, simple affirmative responses (e.g., head nods or saying "yes") to yes–no questions were not checked to provide a conservative test for suggestibility effects. Because coder-rated accurate interviewer-supplied details about the suggested event could be considered inaccurate because the child was not at the event, a composite interviewer-supplied details index was created that summed the mentions of accurate and inaccurate interviewer-supplied details for the suggested event. For the true event, this index was simply a repetition of the mentions of inaccurate interviewer-supplied details measures. The number of interviewer-supplied details could be higher than the number of utterances reported here if the child mentioned multiple details in an utterance or mentioned these details in a mixed utterance (an utterance that contains some true and some false details).

*Inaccurate responses to five yes–no leading questions.* In the second and third interview, mixed among open-ended and leading questions that correctly (in the true event) called for an affirmative answer were five leading questions that incorrectly called for an affirmative answer (e.g., "Did you help the lady put a leash on the monkey?"). These questions provided the most direct test of the construct validity of the Yield subscale because it would be expected that children who have an affirmative response bias when answering the leading yes–no VSSC questions would also tend to answer this sort of question affirmatively in the repeated-interview situation.

*Free-recall assent rates.* Whether a child assented to participating in an event entailed at minimum stating "yes" when asked if he or she was there, even if no further details were provided. Interrater reliability measured by kappa for this measure was .76.

### *Phase 2: The VSSC Measures*

Approximately 2–4 weeks after Phase 1, children were shown a short, 5-min video entitled "Billy's Birthday Party" individually or in small groups. One to 4 days after the video was shown, a new interviewer, who was unaware of the children's performance in the repeated interview phase, asked the children what they remembered about the video in an open-ended, nonsuggestive manner. When children stop responding to open-ended probes such as "What else happened?" directed probe questions were asked, such as "What happened when the children opened the presents?" The number of events or people in the video who were correctly described in either free recall or in response to further open-ended questions was coded as a measure of memory recall. There are 68 names of the characters and key events used by Scullin and Ceci (2001) for coding memory recall on a gist basis. Each correctly remembered salient point was awarded 1 point resulting in a hypothetical range of 0–68 memory recall points.

Immediately after eliciting memory recall, 18 VSSC questions were administered, 14 of which were leading. Yield scores were derived by summing the number of times children responded affirmatively to 1 of 14 leading questions, permitting a range of 0–14. Some of the leading questions had presuppositions in them. For instance, the inaccurate leading question "Did the kids break a balloon while they were hitting them around?" presupposed that the kids were hitting the balloons around. Answers were scored as correct (nonyielding) if they correctly denied the main portion of the question (i.e., responded "no") but were not evaluated on whether they also denied the presupposition on those occasions when they did this. However, some children did accurately deny the incorrect presuppositions without addressing the main part of the question, typically in the form of "I don't think we saw that part of the video," and these were also scored as correct (nonyielding). "I don't know" was also scored as a correct, nonyielding answer. Twice during the administration of the scale, the child was told, "You made a few mistakes. Let's go through them again and see if you can do better this time." Each change in response from the initial questioning was given a Shift score of 1, permitting a range of 0–18. Total Suggestibility is a composite suggestibility score created by summing the child's Yield and Shift scores, thereby ranging from 0 to 32.

## Results

### *Hypothesis 1: True Versus False Event Manipulation*

Table 1 presents the means and standard deviations for each of the Phase 1 variables. To address Hypothesis 1; that is, children will be more accurate about the true event than the suggested event and more inaccurate about the suggested event than the true event, we calculated difference scores for each child between their recall of the true event versus the suggested event. As the dependent *t* statistics show, the true event versus suggested event manipulation was successful, with all difference scores except for younger children's affirmative responses to the five leading yes–no questions being significant and in the direction predicted.

### *Hypothesis 2: Age Differences in Suggestibility*

Hypothesis 2 is that there are age differences in the Memory Recall and Shift measures that will systematically affect the predictive ability of the scales. Table 2 presents the means, standard deviations, and number of observations for each of the Phase 2

Table 1
*Descriptive Statistics of Phase 1 Variables by Age Group*

| Variable | True event | | Suggested event | | Difference | | |
|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *d* |
| **Accuracy indices** | | | | | | | |
| Free-recall accurate utterances | | | | | | | |
| Older | 7.20 | 4.88 | 1.32 | 1.35 | 5.88† | 0.90 | 1.64 |
| Younger | 2.80 | 2.04 | 1.88 | 1.42 | 0.92† | 0.48 | 0.52 |
| Probe question accurate utterances | | | | | | | |
| Older | 9.20 | 5.14 | 3.08 | 2.60 | 6.12† | 0.99 | 1.50 |
| Younger | 5.80 | 3.57 | 3.12 | 1.96 | 2.68† | 0.77 | 0.93 |
| Accurate interviewer-supplied details | | | | | | | |
| Older | 12.32 | 4.33 | 5.60 | 3.18 | 6.72† | 0.88 | 1.77 |
| Younger | 10.20 | 3.52 | 6.16 | 2.49 | 4.04† | 0.73 | 1.33 |
| **Inaccuracy indices** | | | | | | | |
| Free-recall inaccurate utterances | | | | | | | |
| Older | 1.80 | 2.36 | 4.00 | 5.93 | −2.20† | 0.92 | −0.49 |
| Younger | 2.68 | 3.00 | 5.48 | 6.85 | −2.80† | 1.03 | −0.53 |
| Probe question inaccurate utterances | | | | | | | |
| Older | 7.84 | 5.18 | 12.84 | 9.16 | −5.00† | 1.32 | −0.67 |
| Younger | 9.16 | 6.45 | 13.08 | 8.47 | −3.92† | 1.51 | −0.52 |
| Inaccurate interviewer-supplied details | | | | | | | |
| Older | 2.72 | 2.30 | 5.56 | 3.00 | −2.84† | 0.59 | −1.06 |
| Younger | 1.92 | 2.52 | 5.08 | 2.55 | −3.16† | 0.47 | −1.24 |
| Inaccurate responses to 5 yes–no leading questions | | | | | | | |
| Older | 2.36 | 1.43 | 3.20 | 1.58 | −0.84† | 0.28 | −0.56 |
| Younger | 3.16 | 1.60 | 3.56 | 1.64 | −0.40 | 0.20 | −0.25 |
| **Composite inaccuracy indices** | | | | | | | |
| Total inaccurate utterances | | | | | | | |
| Older | 9.64 | 6.80 | 21.24 | 16.91 | −11.60† | 12.79 | −0.90 |
| Younger | 11.84 | 7.92 | 23.56 | 15.69 | −11.72† | 12.23 | −0.94 |
| Total inaccurate interviewer-supplied details | | | | | | | |
| Older | 2.72 | 2.30 | 11.16 | 5.48 | −8.44† | 4.82 | −2.01 |
| Younger | 1.92 | 2.52 | 11.24 | 4.28 | −9.32† | 3.84 | −2.65 |

*Note.* The difference statistic *p* values are based on paired *t*-test values. Cohen's *d* statistic was calculated using pooled between-group standard deviations to prevent overestimation of the effect size, following the recommendations of Dunlop, Cortina, Vaslow, and Burke (1996).
† $p < .05$, one-tailed.

individual difference variables in the study derived from the VSSC by age group. For the VSSC Memory Recall measure, Scullin and Ceci (2001) found an average Cohen's *d* of 0.49 for each year in age from 3 to 5. Our power to detect a *d* score of this size or larger with this sample was 0.54, and in this sample the *d* score between older and younger children for free recall was 1.16, perhaps because our sample spanned the entire 3–5 age range even though the mean ages were only a year apart for the younger and older groups.

Table 2
*Descriptive Statistics of Phase 2 Variables by Age Group*

| Variable | Older ($n = 25$) | | Younger ($n = 25$) | |
|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* |
| Age in months | 60.40 | 4.70 | 48.76 | 4.40 |
| VSSC Memory Recall | 3.18 | 2.07 | 1.18 | 1.30 |
| VSSC Yield | 5.40 | 3.50 | 7.48 | 3.40 |
| VSSC Shift | 4.88 | 2.70 | 4.60 | 2.80 |
| VSSC Total Suggestibility | 10.28 | 4.35 | 12.08 | 4.11 |

*Note.* VSSC = Video Suggestibility Scale for Children.

We hypothesized that younger children would show a different pattern than older children in terms of how they shift their responses. An examination of the scatterplots of yield versus shift found this to be the case. Among the older group, there was no systematic pattern visible in the scatterplot, and there was no correlation between yield and shift for the older group, $r(25) = -.03$, *ns.* The younger group showed a striking negative linear relationship between Yield and Shift scores, with two children who were outliers in terms of having both very low Yield and very low Shift scores. The correlation between Yield and Shift scores without the two outliers was, $r(23) = -.66$, $p < .05$; with the two outliers, the correlation was, $r(25) = -.30$, *ns.* Younger children with low Yield scores tended to shift their "no" or "I don't know" responses to affirmative responses (and became less accurate and

more acquiescent), whereas high yielders would have low Shift scores because their affirmation bias would limit the number of additional "no" to "yes" changes they could make. Although the children could gain Shift points by changing a "yes" response to a "no" response, younger children tended not to do that. Among older children, there was no systematic relationship between yield and shift, suggesting that they used a different response strategy than younger children after receiving negative feedback.

Table 3 provides the Pearson correlations between the Phase 2 individual difference measures derived from the VSSC and the Phase 1 (repeated interview) variables for both the true event and the suggested event. For clarity and conciseness, we have presented the data in a manner that highlights the validity coefficients between the Phase 2 VSSC measures and the Phase 1 repeated interview variables. An N of 25 has a power of .80 for detecting a correlation of .49 at $\alpha < .05$ (one-tailed), just below the threshold of Cohen's (1988) definition of a large effect size for correlations (.50). This is around the value at which correlations between educational assessment measures and their criteria tend to cluster (Cohen, 1988). We viewed finding correlations of .40 or higher as being meaningfully high enough for the scale to be useful for research and practical purposes, and with an N of 25, our power was .67 for detecting correlations of this magnitude or higher. Statistical significance for correlations at $\alpha < .05$ (one-tailed) is .337, which is slightly above Cohen's (1988) definition of a medium-sized effect (0.30). Because this study was exploratory and extremely labor-intensive in nature, statistical power was sacrificed by considering the age groups separately, as a combined N of 50 would have provided a power of .80 to detect a correlation of .35 or greater.

Hypothesis 2—that there would be age differences in the predictive ability of the scale measures—received some support in Table 3, as there were 32 significant correlations in the predicted direction between the VSSC measures and the outcome criteria for older children, and 28 of these were over .40. In contrast, there was only one significant correlation in the predicted direction for younger children. It is important to note that because of our limited power to detect medium-sized effects (from $r = .30$ to $r = .49$), we cannot rule out that the scale has predictive ability in this range for younger children. However, it would require a substantially larger sample size to detect an effect toward the lower end of this range. For example, if the true relationship between a VSSC and a suggestibility criterion is $r = .30$ among younger children in the population, a sample size of 68 would be necessary to have a power of .80 of detecting this with a one-tailed alpha level of $p < .05$ (Cohen, 1988).

The larger correlations for the suggestibility measures among the older children than among the younger children may indicate that the VSSC measures show a closer relationship to the construct of suggestibility in forensic situations among the older children. However, one correlation being significant and another not being significant does not mean that the first correlation is significantly larger than the second, and thus we cannot draw strong conclusions about whether the scale has validity for use with younger children from Table 3 because of the small sample size of each of these groups.

Table 3
*Correlations of Phase 1 Variables × Phase 2 Variables by Age Group*

| Variable | True event | | | | | Suggested event | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MO | MR | Y | S | TS | MO | MR | Y | S | TS |
| Phase 1 accuracy indices | | | | | | | | | | |
| Free-recall accurate utterances | .10 | **.42**† | .12 | .45* | .38 | −.09 | .33 | **.34**† | **.32** | **.47**† |
| | .07 | **.02** | −.09 | −.05 | −.12 | −.29 | −.03 | **.26** | **.05** | **.29** |
| Probe question accurate utterances | .12 | **.38**† | .29 | .17 | .34 | .14 | .09 | **.55**† | **.17** | **.55**† |
| | .11 | **−.14** | −.10 | .28 | .10 | .09 | .23 | **.07** | **−.20** | **.07** |
| Accurate interviewer-supplied details | .24 | **.31** | .18 | .23 | .29 | .00 | .15 | **.62**† | **.17** | **.61**† |
| | −.09 | **.16** | −.12 | .10 | −.05 | −.10 | .13 | **.18** | **−.23** | **.01** |
| Phase 1 inaccuracy indices | | | | | | | | | | |
| Free-recall inaccurate utterances | −.25 | .14 | **.25** | **.58**† | **.56**† | −.31 | .31 | **.29** | **.42**† | **.49**† |
| | −.23 | .28 | **−.02** | **−.37** | **−.27** | −.11 | .19 | **−.25** | **−.19** | **−.37** |
| Probe question inaccurate utterances | −.20 | −.14 | **.53**† | **.52**† | **.75**† | −.18 | .16 | **.54**† | **.43**† | **.70**† |
| | −.38 | .15 | **.11** | **−.01** | **.10** | −.18 | .25 | **−.08** | **.08** | **−.03** |
| Inaccurate interviewer-supplied details | .03 | −.03 | **.32** | **−.20** | **.14** | .11 | .19 | **.30** | **.24** | **.39**† |
| | −.09 | .31 | **−.01** | **.00** | **−.01** | −.14 | .33 | **.14** | **−.29** | **−.06** |
| Inaccurate responses to 5 yes–no accurate "no" questions | .03 | −.07 | **.40**† | **.43**† | **.59**† | .11 | −.16 | **.56**† | **.19** | **.57**† |
| | −.06 | .01 | **.37**† | **−.11** | **.28** | .01 | .19 | **.28** | **−.31** | **.06** |
| Phase 1 composite inaccuracy indices | | | | | | | | | | |
| Total inaccurate utterances | −.24 | −.06 | **.49**† | **.60**† | **.76**† | −.19 | .24 | **.51**† | **.43**† | **.67**† |
| | −.39 | .23 | **.08** | **−.15** | **−.02** | −.16 | .25 | **−.12** | **−.06** | **−.16** |
| Total inaccurate interviewer-supplied details | .03 | −.03 | **.32** | **−.20** | **.14** | .06 | .19 | **.52**† | **.23** | **.57**† |
| | −.09 | .31 | **−.01** | **.00** | **−.01** | −.14 | .26 | **.18** | **−.30** | **−.03** |

*Note.* For clarity and conciseness, we present the data in a manner that highlights the validity coefficients between the Phase 2 VSSC measures and the Phase 1 repeated interview variables. Numbers in bold are hypothesized to be significantly positive. For each variable, older children ($n = 25$) are represented by the top-line values and younger children ($n = 25$) are represented by the bottom-line values. MO = age in months; MR = Video Suggestibility for Children (VSSC) Memory Recall; Y = VSSC Yield; S = VSSC Shift; TS = VSSC Total Suggestibility.
* $p < .05$, two-tailed. † **$p < .05$, one-tailed.**

### *Hypothesis 3: Accuracy About the True Event*

As can be seen in Table 3, the hypothesis that memory recall would be related to accuracy about the true event was partially supported given our low power and conservative criteria for statistical significance. Higher scores on Memory Recall were related to the total number of accurate utterances made in response to free-recall and probe questions among older children.

### *Hypothesis 4: Inaccuracy About the True Event*

Hypothesis 4, suggesting that yield, shift, and total suggestibility would be related to suggestibility and lack of accuracy about the true event, received partial support among older children. Older children's Shift and Total Suggestibility scores were related to three out of the four Phase 1 inaccuracy indices, including the number of inaccurate utterances made in response to both free-recall and probe questions and inaccurate responses to five yes–no leading questions. In addition to shift and total suggestibility, yield was also significantly related to inaccurate utterances in response to probe questions and the five yes–no leading questions. Yield, shift, and total suggestibility were also all significantly related to the composite-total-inaccurate utterances measure, which summarizes the free-recall and probe inaccurate utterances. Among younger children, the only significant relationships were between the Yield measure and the number of inaccurate responses to five yes–no leading questions, whose response format parallels that of the Yield subscale. None of the VSSC suggestibility measures were significantly related to children's mentions of inaccurate interviewer-supplied details for either age group. Because mentions of interviewer-supplied details about a true event is a frequently studied type of children's and adult's suggestibility (e.g., Loftus, 1979; Templeton & Wilcox, 2000; Zaragoza & Mitchell, 1996), the lack of a strong effect size relationship between the VSSC measures and this type of suggestibility reveals a potentially important limitation of the scale measures but does not preclude the existence of a meaningful medium effect size given our limited power.

### *Hypothesis 5: Accuracy About the Suggested Event*

Our hypothesis that VSSC Yield, Shift, and Total Suggestibility scores would be related to children's ability to make accurate utterances and mention accurate interviewer-supplied details about the nonexperienced event received some support, but only among older children. Older children with higher Yield scores and higher Total Suggestibility scores made more accurate utterances in response to both free-recall and probe questions and made more mentions of accurate interviewer-supplied details. Shift scores were not significantly related to any of the accuracy measures. These findings suggest that older children with higher Yield and Total Suggestibility scores have a greater tendency to provide accurate information about a nonexperienced event when they are questioned.

### *Hypothesis 6: Inaccuracy About the Suggested Event*

Our hypothesis that VSSC Yield, Shift, and Total Suggestibility scores would be related to children's inaccuracy about the nonexperienced event received some support, but again only among older children. For older children, total suggestibility was related to all of the inaccuracy indices and composite indices for the suggested event. The findings for yield and shift were mixed, with higher Yield scores being related to a larger number of probe-question-inaccurate utterances, inaccurate responses to the five yes–no leading questions, and the two composite inaccuracy indices. High-older shifters tended to have a higher number of inaccurate utterances in response to free-recall and probe questions, as well as the index combining these.

### *Hypothesis 7: Incremental Predictive Ability of Yield and Shift*

In view of the repeatedly established finding that chronological age is the single strongest predictor of suggestibility (Ceci & Bruck, 1993), we conducted four hierarchical regression analyses to examine whether the VSSC measures of yield and shift predict suggestibility significantly above and beyond simply knowing the child's age. For our dependent variables, we used the four composite inaccuracy indices of true-event-total-inaccurate utterances, true-event-total-inaccurate interviewer-supplied details, suggested-event-total-inaccurate utterances, and suggested-event-total-inaccurate interviewer-supplied details. As Table 4 shows, among older children there was a significant increase in $R^2$ when yield and shift were added to age in stepwise regressions for the total number of inaccurate utterances about the true event, total number of inaccurate utterances about the suggested event, and mentions of inaccurate interviewer-supplied details about the suggested event. Both yield and shift were significant at $p < .05$ in the three significant regressions, and effect sizes met Cohen's (1988) criteria for large effect sizes ($R^2 = .26$) for these three regressions. Among older children, the model predicting the number of mentions of inaccurate interviewer-suggested details about the true event was not significant. Among younger children, none of the models was significant.

### *Hypothesis 8: Predicting Assent to the Final Interview*

Finally, we were interested in examining whether the VSSC measures could predict which children were likely to make false reports about the suggested event in free recall when they were interviewed using the Cognitive Interview techniques in the final interview, which followed three highly leading interviews with a different interviewer. As Figure 1 shows, 92% (23 out of 25) of the older children assented to the true event in response to the first interview's free-recall questions. This rate of assent dropped to 88% in the second interview, then 76% in the third interview before increasing to 84% in the final interview. Only 28% of the older children assented to the suggested event in the first interview's free-recall questions. The rate of assent in free recall dropped to 16% in the second interview but then rose to 44% in the third interview and 60% (15 out of the 25) in the fourth interview, despite the fact that the interview was conducted by a different interviewer who stressed to the children that they should tell only what they remembered and not make things up. We entered age, yield, and shift into a maximum likelihood logistic regression predicting whether the child assented to the suggested event in the

Table 4
*Summary of Hierarchical Regression Analyses for Variables Predicting Composite Inaccuracy Indices*

| Variable | $\Delta R^2$ | Final $R^2$ | $F$ statistic |
|---|---|---|---|
| True event total inaccurate utterances | | | |
| Older children ($n$ = 25) | | | |
| Step 1 | | | |
| Age | .06 | .06 | $F(1, 23) = 0.25$ |
| Step 2 | | | |
| Age, yield, shift | .56 | .62 | $\Delta F(2, 21) = 15.26$** |
| Younger children ($n$ = 25) | | | |
| Step 1 | | | |
| Age | .15 | .15 | $F(1, 23) = 4.21$ |
| Step 2 | | | |
| Age, yield, shift | .01 | .17 | $\Delta F(2, 21) = 0.13$ |
| True event total inaccurate interviewer-supplied details | | | |
| Older children ($n$ = 25) | | | |
| Step 1 | | | |
| Age | .00 | .00 | $F(1, 23) = 0.89$ |
| Step 2 | | | |
| Age, yield, shift | .15 | .15 | $\Delta F(2, 21) = 1.77$ |
| Younger children ($n$ = 25) | | | |
| Step 1 | | | |
| Age | .01 | .01 | $F(1, 23) = 0.67$ |
| Step 2 | | | |
| Age, yield, shift | .00 | .01 | $\Delta F(2, 21) = 0.00$ |
| Suggested event total inaccurate utterances | | | |
| Older children ($n$ = 25) | | | |
| Step 1 | | | |
| Age | .04 | .04 | $F(1, 23) = 0.36$ |
| Step 2 | | | |
| Age, yield, shift | .42 | .45 | $\Delta F(2, 21) = 8.01$** |
| Younger children ($n$ = 25) | | | |
| Step 1 | | | |
| Age | .03 | .03 | $F(1, 23) = 0.64$ |
| Step 2 | | | |
| Age, yield, shift | .02 | .05 | $\Delta F(2, 21) = 0.22$ |
| Suggested event total inaccurate interviewer-supplied details | | | |
| Older children ($n$ = 25) | | | |
| Step 1 | | | |
| Age | .01 | .01 | $F(1, 23) = 0.09$ |
| Step 2 | | | |
| Age, yield, shift | .38 | .30 | $\Delta F(2, 21) = 6.56$** |
| Younger children ($n$ = 25) | | | |
| Step 1 | | | |
| Age | .02 | .02 | $F(1, 23) = 0.43$ |
| Step 2 | | | |
| Age, yield, shift | .09 | .11 | $\Delta F(2, 21) = 1.04$ |

** $p < .05$.

free-recall part of the final interview. The model was significant for older children (−2 log likelihood = 20.21), $\chi^2(3, N = 25) = 13.44$, $p < .01$, with yield being the significant predictor variable (95% Wald confidence interval for adjusted odds ratio for yield = 1.15–2.94) and months and shift being nonsignificant. A standard measure of goodness of fit for logistic regressions is to examine whether (in this instance) out of all possible pairs of children in which one child denies participating in the suggested event and the other child assents to participating in the suggested event, the child who assented to the event has a greater predicted likelihood of assent as predicted by the logistic regression equation. In 89.3% of the 150 possible pairs of discordant older children, the child who assented to the event had a higher probability of assenting on the basis of his or her Yield, Shift, and age scores (Somers' $D$ = .78), indicating a good fit for the model.

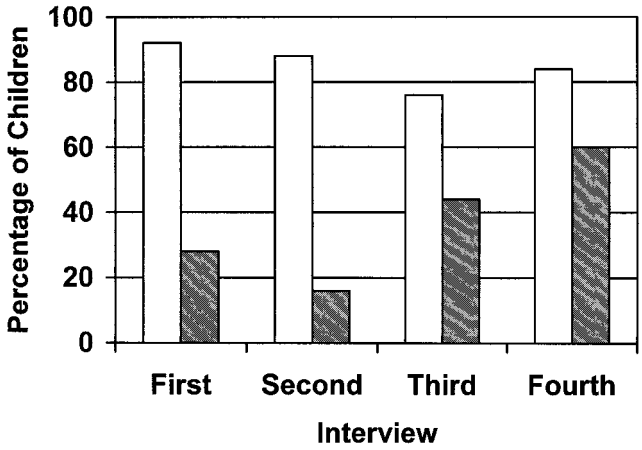



*Figure 1.* Percentage of older children ($n$ = 25) who assented to the true event (open bars) and the suggested event (solid bars) in the free-recall portion of each of four interviews.

As can be seen in Figure 2, among the younger children the assent rate for both the true event and the suggested event fluctuated across interviews, with the percentage of children assenting to the true event varying between 60% and 88% and the percentage of children assenting to the suggested event dropping from 64% to 24% between the first and second interview and then rising to 76% (19 out of 25) by the fourth interview. We entered age, yield, and shift into a maximum likelihood logistic regression predicting whether the child assented to the suggested event in the free-recall part of the final interview. For the younger group, this model was

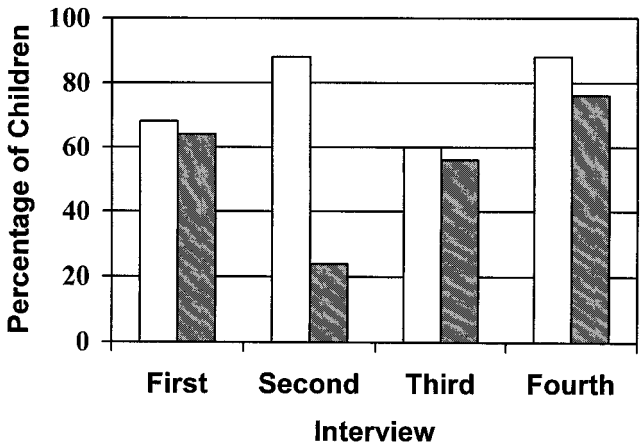



*Figure 2.* Percentage of younger children ($n$ = 25) who assented to the true event (open bars) and the suggested event (solid bars) in the free-recall portion of each of four interviews.

not significant ($-2$ log likelihood = 24.81) $\chi^2(3, N = 25) = 2.74$, *ns.* In summary, these findings suggest that the Yield subscale of the VSSC helps predict which older children are more likely to assent to a suggested event after repeated suggestive questioning.

## Discussion

This study examined how the individual difference measures of suggestibility from the VSSC were related to individual differences in children's performance in a repeated interview field study conducted by different interviewers in which the children were interviewed four different times about both a personally experienced true event and an event that was merely suggested by the interviewer. The three main objectives of this study were (a) to evaluate the construct validity of the VSSC by examining whether children who had higher scores on its measures of yield, shift, and total suggestibility had elevated levels of suggestibility in a field study in which they were interviewed about both true events and suggested events and by examining whether children who scored high on VSSC Memory Recall had better memory for the true event; (b) to determine whether yield and shift predicted levels of suggestibility about both true events and suggested events above and beyond knowing the child's age; and (c) to explore whether yield and shift predicted whether children still assented to a suggested event over and above age when they were interviewed using forensic techniques that have been shown to increase the amount of accurate testimony in slightly older children.

Overall, the results of this study provide substantial support for the construct validity of the VSSC among children over the age of 4 years 6 months. The Total Suggestibility measure was positively related to 10 of the 11 inaccuracy and suggestibility measures that we studied (and 3 of the 4 composite inaccuracy indices), with effect sizes in the medium–large range. However, this support for the VSSC must be qualified by the finding that there were no significant relationships between total suggestibility and the criteria variables for younger children, and our sample size lacked sufficient power to detect medium-sized effects. Below, we discuss the support for each of our hypotheses in turn. It should be pointed out that the suggestive interviewing techniques that we used in the field study produced a high rate of false assents, with the majority of children assenting to both the true event and the suggested event by the third interview. Moreover, the children frequently went beyond saying "yes," "no," or "I don't know" in response to questions and often provided elaborate details.

Both older and younger children in our study were more accurate about the true event than the suggested event and were more inaccurate about the suggested event than the true event. This was true in all cases except that there was no significant difference between younger children's tendency to say "yes" to leading questions that called for an affirmative answer in either the true or the false event.

In our examination of age differences in response to VSSC measures affecting the predictive ability of our scale, we found that younger children had lower Memory Recall scores than older children and tended to change their answers systematically from "no" or "I don't know" to "yes" and not vice versa. Among younger children, there was only one significant correlation in the predicted direction between the VSSC suggestibility measures and the criteria variables. Among older children, there were 32 significant correlations, meaning there may be major age differences in the predictive ability of the scales. However, our sample size was too small to rule out the possibility of medium effect size relationships (in the .30–.49 range) between the VSSC measures and the criteria variables among younger children.

Our prediction that the VSSC Memory Recall scale will be related to accuracy about the true event was partially supported. The Memory Recall scale, a test of how much children are able to verbally recall about a video of a birthday party seen 1 to 3 days earlier, was found to be positively related to the number of accurate utterances children made about a true event in response to free-recall and probe questions during the repeated interview field study, but only among older children. This provided evidence that the VSSC's measure of basic mnemonic performance is related to older children's utterances about a true event in a very different context.

The VSSC's Yield, Shift, and Total Suggestibility measures predicted some aspects of children's suggestibility and lack of accuracy about the true event. Among older children, yield, shift, and total suggestibility were all related to children's inaccurate utterances in response to the interviewer's probe questions and the children's inaccurate affirmations of misleading yes–no questions, as well as the summary index of all of the children's inaccurate free-recall and probe utterances. Among older children, both shift and total suggestibility were additionally related to the number of inaccurate utterances made in response to the interviewer's free-recall questions. Among younger children, the only significant relationship was between yield and "yes" responses to the interviewer's misleading yes–no questions.

We also examined the degree to which children were able to successfully provide accurate details about a suggested event by either repeating details mentioned by the interviewer, guessing, or hearing details from others. Among older children, Yield and Total Suggestibility scores were related to all of the measures of children's accuracy about the suggested event. Older children who scored higher on Yield and Total Suggestibility were more likely to make accurate utterances about the suggested event in response to free-recall and probe questions and report more accurate details that had been previously mentioned by the interviewer. Shift was not related to older children's accuracy about the suggested event, and none of the VSSC measures predicted accuracy about the suggested event in younger children.

Total Suggestibility scores were also significantly related to all of our measures of older children's inaccuracy about the suggested event, including the number of inaccurate utterances made about the suggested event in response to both free-recall and probe questions, number of inaccurate interviewer-supplied details mentioned by the child, and the number of inaccurate affirmative responses to misleading questions that called for a "yes" answer. Total suggestibility was also related to the composite indices of total inaccurate utterances and the total number of inaccurate interviewer-supplied details. Yield and shift were only related to some of the measures of older children's inaccuracy. Older high yielders had higher numbers of inaccurate responses to probe questions, more affirmative answers to leading yes–no questions, higher numbers of total inaccurate utterances, and higher numbers of mentions of inaccurate interviewer-supplied details. Older high

shifters were associated with higher numbers of inaccurate utterances in response to free-recall and probe questions, as well as the composite of these. Among younger children, there were no significant relationships between the VSSC measures and inaccuracy about the suggested event.

Overall, the Total Suggestibility scale of the VSSC received broad support for tapping into the construct of older children's suggestibility and lack of accuracy in a series of suggestive forensic interviews about both true and suggested events, correlating significantly and positively with all of the relevant outcome criteria except for mentions of inaccurate interviewer-supplied details about the true event. The support for yield and shift considered individually was mixed, as yield was correlated with 7 out of 11 measures (not counting the composite indices) and shift was correlated with only 5 out of 11 measures for older children.

We conducted four stepwise regression analyses to examine whether children's Yield and Shift scores could predict the degree to which children were inaccurate about the true and suggested field events. These analyses showed that for the true event, yield and shift significantly contributed to the prediction of the total number of inaccurate utterances made by older children. Higher tendencies both to yield and shift were related to children's tendency to report inaccurate information. However, for older and younger children both yield and shift failed to contribute to the prediction of the number of inaccurate details about the true event that had also been mentioned by the interviewer. The scale does not directly test this measure, though, because children never received biasing information about what happened in the video, except in the form of leading questions. For the suggested event, the full models of both hierarchical regressions for older children were significant. In predicting both the number of total inaccurate utterances children made and the total number of inaccurate interviewer-supplied details mentioned by the child, the addition of yield and shift increased the predictive power of the model substantially, and both yield and shift were statistically significant.

Finally, yield predicted whether an older child was likely to assent to the suggested event when interviewed in a modified forensic manner after three leading interviews. In Phase 1 of the study, the fourth and final interview was conducted by a different interviewer and incorporated some context reinstatement techniques derived from the Cognitive Interview for children (Köhnken, 1993). The techniques appeared to work well at eliciting detailed narratives from children about both the true event and the suggested event in free recall, but also resulted in the final interview having the highest number of children who assented to the suggested event. We found that only among the older children, the Yield subscale was a significant predictor of whether a child assented to the suggested event in the free-recall portion of the final interview.

### *Limitations of the Study*

The small sample size of the two age groups limits the conclusions we can draw about the null findings in the younger age group because of our low statistical power to detect medium effect sizes. The finding that the strongest relationships with the criteria variables were among children over 4 years 6 months of age means that further research with larger samples may be necessary to establish whether the scale has practical utility with children under this age.

Among children under 4 years 6 months, the validity of the VSSC Memory Recall scale may be compromised by floor effects, and the Shift measure may be adversely affected because children this age tend to systematically change their initial "no" or "I don't know" responses to "yes" responses after receiving negative feedback, resulting in a strong negative correlation between Yield and Shift scores among most of the younger children. Young children's difficulty with theory of mind and executive-functioning tasks may mean that they would have difficulty with the mental processes necessary to evaluate incorrect initial responses in light of negative feedback from an interviewer. Although the scale may not be as valid for younger children for these reasons, studies have shown that children who lack the ability to perform theory-of-mind tasks are more suggestible than those who can (Templeton & Wilcox, 2000; Welch-Ross, 1999). However, once children (and adults) have attained theory of mind, many of them are still suggestible. It may be at this point that suggestibility becomes more accurately measurable by the VSSC. Among older children, yield and shift were uncorrelated, perhaps indicating that these older children had better developed mnemonic strategies for evaluating the accuracy of their responses. Future research could address whether there is a qualitative change in the nature of children's suggestibility after a certain age or attainment of certain cognitive skills.

Thus, the findings presented here suggest that an important research question to address in the future is whether there is a particular age or level of attainment of a certain cognitive functioning at which time suggestibility becomes a more traitlike characteristic. If so, it may be worth examining how the relationship between VSSC measures and other measures of suggestibility changes at different ages.

Pezdek and her colleagues have argued that it may be relatively easy to suggest a change in an event if the child has experienced the event but substantially harder to plant memories of a false event into a child's memory (Pezdek, Finger, & Hodge, 1997). These authors asserted that when implausible or unlikely events are suggested, children are much less likely to be misled and believe that the event occurred. However, in this study, some children came up with very elaborate accounts for unlikely events having to do with finding a monkey (who by some accounts was alive) or helping someone who injured her ankle after spilling Play-Doh, both implausible events that were unlikely to be part of children's actual or even observed experiences. Their stories often went beyond script-based knowledge. For example, during the fourth interview one child (who had not participated in the monkey event) described her encounter with the monkey lady as follows:

> She said, "Let's find the monkey," and everybody said "yes," except John said "no." And then John vacillated. He said "yes." And then John went home. And then everybody went back to play. And then I went inside again to help her find the monkey again inside here. And then it was inside here. In there [points to right].

One may wonder whether the VSSC is simply capturing a trait related to talkative "yeah-saying." Although we cannot rule this out, it could be the case that a tendency to be talkative and willing to go along with an interviewer's suggestions is an important

characteristic of believable, suggestible children who provide false statements in interrogative situations such as the child above.

### *Summary*

The results of this study show that the VSSC is a valuable addition to the research on children's suggestibility, especially suggestibility in forensic contexts. By establishing a variety of validation criteria in measures of accuracy, inaccuracy, and suggestibility in an independent repeated-suggestive-interviews experiment about true and suggested events, we were able to establish its utility as a measure of individual differences of children's inaccuracy and suggestibility. More important, there were medium and strong effect size relationships between the VSSC's Total Suggestibility scale and multiple validation criteria among children over 4 years 6 months of age. Although neither yield nor shift correlated with as many measures of suggestibility in the field study as did total suggestibility, our results do suggest that they are both tapping into important factors that contribute to children's suggestibility in forensic situations. This provides support for the view that there are meaningful individual differences in older children's suggestibility in forensic interview situations and that this type of suggestibility has at least two major, statistically independent dimensions that correspond to the Yield and Shift subscale measures, each of which contributes independently to the prediction of suggestibility (Gudjonsson, 1992b; Register & Kihlstrom, 1988).

The finding that there are individual differences in suggestibility that correlate strongly across situations after 4 years 6 months of age indicates that the nature of suggestibility may change and become more traitlike around this age, perhaps because of the development of theory of mind (Welch-Ross, 1999) or better executive functioning (Perner & Lang, 1999). An investigation of the relationships between VSSC scores and factors like these may shed some light on how the nature of suggestibility changes at this time.

It is important to note that although scores on the Yield, Shift, and Total Suggestibility subscales were related to lack of accuracy about both true and suggested events, this relationship was far from perfect. Even a child who makes inaccurate statements about a true event may report much of the gist correctly. However, the results of this study show that the VSSC scale measures—especially the composite Total Suggestibility measure—are tapping into meaningful aspects of suggestibility for children over the age of 4 years 6 months. Thus, the measure may provide important information to forensic interviewers and psychologists about a child's tendency to provide inaccurate information about true and suggested events. The VSSC may also be a potentially valuable training tool because observing children's responses to its questions can provide investigators with a sense of how sensitive some children are to suggestive influences.

## References


Baker-Ward, L., Gordon, B. N., Ornstein, P. A., Larus, D. M., & Clubb, P. A. (1993). Young children's long-term retention of a pediatric examination. *Child Development, 64,* 1519–1533.

Bruck, M., Ceci, S. J., & Hembrooke, H. A. (in press). *On the nature and ontogeny of true and false narratives.* Developmental Review.

Bruck, M., Ceci, S. J., & Melnyk, L. (1997). External and internal sources of variation in the creation of false reports in children. *Learning and Individual Differences, 9,* 289–316.

Candel, I., Merckelbach, H., & Muris, P. (2000). Measuring interrogative suggestibility in children: Reliability and validity of the Bonn Test of Statement Suggestibility. *Psychology, Crime and Law, 6,* 61–70.

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113,* 403–439.

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony.* Washington, DC: American Psychological Association.

Ceci, S. J., & Friedman, R. D. (2000). The suggestibility of children: Scientific research and legal implications. *Cornell Law Review, 86,* 34–108.

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

Ceci, S. J., Loftus, E. F., Leichtman, M., & Bruck, M. (1994). The possible role of source misattributions in the creation of false beliefs among preschoolers. *International Journal of Clinical and Experimental Hypnosis, 42,* 304–320.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cunningham, J. L. (1988). Contributions to the history of psychology: XLVL. The pioneer work of Alfred Binet on children as eyewitnesses. *Psychological Reports, 62,* 271–277.

Danielsdottir, G., Sigurgeirsdottir, S., Einarsdottir, H. R., & Haraldsson, E. (1993). Interrogative suggestibility in children and its relationship with memory and vocabulary. *Personality and Individual Differences, 14,* 499–502.

Dunlop, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177.

Endres, J., Poggenpohl, C., & Erben, C. (1999). Repetitions, warnings, and video: Cognitive and motivational components in preschool children's suggestibility. *Legal and Criminological Psychology, 4,* 129–146.

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

Garven, S., Wood, J. M., Malpass, R. S., & Shaw, J. S. (1998). More than suggestion: The effect of interviewing techniques from the McMartin Preschool case. *Journal of Applied Psychology, 83,* 347–359.

Goodman, G. S., Bottoms, B. L., Schwartz-Kenney, B. M., & Rudy, L. (1991). Children's testimony about a stressful event: Improving children's reports. *Journal of Narrative and Life History, 1,* 69–99.

Gudjonsson, G. H. (1984). A new scale of interrogative suggestibility. *Personality and Individual Differences, 5,* 303–314.

Gudjonsson, G. H. (1987). A parallel form of the Gudjonsson Suggestibility Scale. *British Journal of Clinical Psychology, 26,* 215–221.

Gudjonsson, G. H. (1992a). Interrogative suggestibility: Factor analysis of the Gudjonsson Suggestibility Scale (GSS 2). *Personality and Individual Differences, 13,* 479–481.

Gudjonsson, G. H. (1992b). *The psychology of interrogations, confessions and testimony.* Chichester, England: Wiley.

Gudjonsson, G. H., & Clark, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour, 1,* 83–104.

Köhnken, G. (1993). *The cognitive interview: A step-by-step introduction.* Unpublished manuscript.

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology, 52,* 297–318.

Loftus, E. F. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

McCauley, M. R., & Fisher, R. P. (1995). Facilitating children's eyewit-

ness recall with the revised cognitive interview. *Journal of Applied Psychology, 80,* 510–516.

Memon, A., Holley, A., Wark, L., & Bull, R. (1996). Reducing suggestibility in child witness interviews. *Applied Cognitive Psychology, 10,* 503–518.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 9,* 337–344.

Pezdek, K., Finger, K., & Hodge, D. (1997). Planting false childhood memories: The role of event plausibility. *Psychological Science, 8,* 437–441.

Quas, J. A., Qin, J., Schaaf, J. M., & Goodman, G. S. (1997). Individual differences in children's and adults' suggestibility and false event memory. *Learning and Individual Differences, 9,* 359–390.

Register, P. A., & Kihlstrom, J. F. (1988). Hypnosis and interrogative suggestibility. *Personality and Individual Differences, 9,* 549–558.

Richardson, G., Gudjonsson, G. H., & Kelly, T. P. (1995). Interrogative suggestibility in an adolescent population. *Journal of Adolescence, 18,* 211–216.

Scullin, M. H., & Ceci, S. J. (2001). A suggestibility scale for children. *Personality and Individual Differences, 30,* 843–856.

Templeton, L. M., & Wilcox, S. A. (2000). A tale of two representations: The misinformation effect and children's developing theory of mind. *Child Development, 71,* 402–416.

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

Welch-Ross, M. K. (1999). Preschoolers' understanding of mind: Implications for suggestibility. *Cognitive Development, 14,* 101–131.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655–684.

Zaragoza, M. S., & Mitchell, K. J. (1996). Repeated exposure to suggestion and the creation of false memories. *Psychological Science, 7,* 294–300.

Received October 3, 2001
Revision received June 24, 2002
Accepted June 25, 2002 ■

## New Editors Appointed, 2004–2009

The Publications and Communications Board of the American Psychological Association announces the appointment of five new editors for 6-year terms beginning in 2004.

Except where noted, as of January 1, 2003, manuscripts should be directed to the following individuals:

- For *Psychology and Aging* (http://www.apa.org/journals/pag.html), submit manuscripts to **Rose T. Zacks, PhD,** Department of Psychology, Michigan State University, East Lansing, MI 48824-1117.
- For *Psychological Assessment* (http://www.apa.org/journals/pas.html), submit manuscripts to **Milton E. Strauss, PhD,** Department of Psychology, Case Western Reserve University, Cleveland, OH 44106-7123.
- For *Journal of Family Psychology* (http://www.apa.org/journals/fam.html), submit manuscripts to **Anne Kazak, PhD,** ABPP, Oncology Psychosocial Research, The Children's Hospital of Philadelphia, Room 1486 (Market Street), 34th and Civic Center Boulevard, Philadelphia, PA 19104. For overnight couriers: Room 1486, 3535 Market Street, Philadelphia, PA 19104.
- For *Journal of Experimental Psychology: Animal Behavior Practices* (http://www.apa.org/journals/xan.html), submit manuscripts as of January 15 to **Nicholas Mackintosh,** Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, United Kingdom.
- For *Journal of Personality and Social Psychology: Personality Processes and Individual Differences* section (http://www.apa.org/journals/psp.html), submit manuscripts to **Charles S. Carver, PhD,** Department of Psychology, University of Miami, P.O. Box 248185, Coral Gables, FL 33124-2070.

**Electronic submission:** As of January 1, 2003, authors will be expected to submit manuscripts electronically through the journal's Manuscript Submission Portal (see the Web site listed above with each journal title). Authors who are unable to do so should correspond with the editor's office about alternatives.

Manuscript submission patterns make the precise date of completion of the 2003 volumes uncertain. Current editors Leah L. Light, PhD, Stephen N. Haynes, PhD, Ross D. Parke, PhD, Mark E. Bouton, PhD, and Ed Diener, PhD, respectively, will receive and consider manuscripts through December 31, 2002. Should 2003 volumes be completed before that date, manuscripts will be redirected to the new editors for consideration in 2004 volumes.