Acting Out the Details of a Pediatric Check-up: The Impact of Interview Condition and Behavioral Style on Children’s Memory Reports

Andrea Follmer Greenhoot, Peter A. Ornstein, Betty N. Gordon, and Lynne Baker-Ward

This investigation was designed to determine whether an enactment interview condition involving a doll and props, in contrast to a verbal interview, would enhance 3- and 5-year-olds’ (N = 62) recall of a pediatric examination. An additional aim was to explore the influence of behavioral styles and language skills on children’s performance, and the extent to which these relations varied by age and interview condition. Both 1- and 6-weeks following their check-ups, the children in the enactment condition, particularly the 3-year-olds, provided more spontaneous, elaborate reports than did those assessed with a verbal protocol. Nonetheless, enactment also resulted in increased errors by the 3-year-olds at the first interview, and by children in both age groups after the 6-week delay. The age and interview condition effects, however, were moderated by the children’s behavioral characteristics. Among the younger children, a measure of manageability predicted performance in the enactment setting, whereas an indicator of persistence was associated with recall in the verbal condition. The results have implications for an understanding of children’s memory of events and of their ability to provide testimony in legal settings.

INTRODUCTION

Research on children’s long-term retention of personally experienced events has increased dramatically in the last decade. A considerable amount of evidence indicates that younger children provide less information about previous events, particularly in response to open-ended probes, than older children (e.g., Baker-Ward, Gordon, Ornstein, Larus, & Clubb, 1993). Converging evidence, however, suggests that the completeness and accuracy of children’s accounts of their experiences depend on a number of factors, including the conditions under which memory is assessed and the children’s cognitive and personality characteristics (e.g., Folds, Footo, Guttentag, & Ornstein, 1990; Gordon et al., 1993). The present investigation was designed to address these issues by (1) determining the extent to which age differences in children’s event recall could be minimized by manipulating the assessment context, and (2) exploring individual differences in relatively stable cognitive and personality attributes that may account for variation in children’s performance within age groups and interview conditions. This research, moreover, was designed to contribute to an understanding both of children’s long-term retention and forgetting of events and of their ability to provide testimony in legal settings.

Enhancing Children’s Recall

A number of researchers have suggested that young children’s memory abilities are actually quite similar to those of older children (e.g., Bender, Wallsten, & Ornstein, 1996; Mandler, 1990) and that age differences in recall performance may reflect developmental changes in the ability to report what is remembered. From this perspective, very young children’s reports of their experiences may not reflect what they actually remember, perhaps because they cannot meet the cognitive and behavioral demands of the typical interview situation. That is, young children’s recall in a conventional interview context may be limited by developmental factors such as restricted language skills, poor grasp of the principles of narrative structure, relative inability to self-conduct a memory search, and lack of interest in certain types of memory interviews. One way to evaluate this position is to examine the influence of characteristics of the interview setting on children’s recall. If age differences are due primarily to retrieval and reporting differences, they might be reduced by interview conditions that are designed to be supportive of these skills in very young children. Indeed, a growing literature indicates that children’s performance varies as a function of a wide range of contextual factors (e.g., Best & Ornstein, 1986; Ornstein, Baker-Ward, & Naus, 1988; Pipe & Wilson, 1994).

One aspect of the interview that can be associated with improved recall among preschoolers is the presence of mnemonic cues during the memory assessment. For example, Pipe and Wilson (1994) found that the completeness of 6- and 10-year-olds’ reports of a laboratory-based play experience (a visit with a magician) was increased if components of that event were
One reason why the 3-year-olds in Gordon et al.‘s (1993) study may not have benefited from the availability of dolls is that younger children may not have the necessary cognitive skills for the effective use of a toy as a representation of the self. Indeed, one implication of DeLoache’s (1990) work on the development of children’s understanding of scale models as representations of larger spaces is that young children may have difficulty in simultaneously viewing a model as an object in its own right and as a symbol of something else. Moreover, consistent with this interpretation, DeLoache and Marzolf (1995) reported that 2½- to 3-year-old children were more successful in indicating where they had been touched during a recently completed Simon Says game through verbal statements and gestures to their own bodies than by means of gestures to a doll.

It seems possible that the difficulties associated with the use of dolls in interviews of young children may be overcome by providing a great deal of contextual support for recall. Thus, performance may improve if 3-year-olds are interviewed in an enactment context in which they are given access to dolls, as well as props that are similar to objects used during the to-be-remembered event, and interviewed in a setting similar to the room in which the original experience occurred. The combination of dolls and props may encourage demonstration to a greater extent than a doll alone, thus creating a context that does not emphasize verbal recall. Furthermore, although a doll may provide some mnemonic support in that it may prompt memory for features of an event involving various body parts, props should present additional cues. In this regard, Saywitz, Goodman, Nicholas, and Moan (1991) have demonstrated the effectiveness of an enactment procedure involving dolls and medical instruments in 5- and 7-year-old children’s recall of a visit to the doctor. Children of both ages provided twice as much information in the enactment session as they did in a free recall session that was based on a verbal interview protocol. Similarly, Murachver, Pipe, Gordon, Owens, and Fivush (1996) found that 5- and 6-year-olds who either observed or were told about a novel play activity (“visiting the pirate”) reported more of the component actions of this event through enactment than verbal recall. However, enactment was also associated with increased errors in reporting, relative to verbal recall. These studies notwithstanding, enactment procedures involving props and dolls have not been examined extensively, especially with children as young as 3 years of age.

With these issues in mind, the present study was designed in part to address the question of whether enactment could enhance the recall performance of very young children. Specifically, 3- and 5-year-olds’ recall of a routine visit to the pediatrician was assessed with one of two interview protocols: a Verbal condition similar to that employed in earlier studies of children’s memory (e.g., Baker-Ward et al., 1993), and an Enactment condition involving the use of a doll and props. It was expected that the Enactment procedure would minimize age differences in recall performance.

Individual Differences in Children’s Recall Performance

Although age differences have been a primary concern of investigators of memory development, few researchers have speculated about the sources of within-group variation in children’s recall. Even when children are interviewed under equivalent conditions, there remains a great deal of unexplained variability in their memory reports, particularly among the younger age groups. It seems likely that individual differences in a number of cognitive and personality characteristics may account for some of this within-group variability in performance. For this investigation, two general individual difference factors were chosen as potential sources of within-group variability: language skill and behavioral style, sometimes referred to as temperament (Thomas, Chess, & Birch, 1968). In the context of a memory assessment, children’s language skills may affect their understanding of an interviewer’s questions as well as their

1Murachver et al. (1996) also included a condition in which the event was experienced directly in the sense that the participating children performed the actions involved in visiting the pirate. For these children, acting out the details of the pirate visit essentially involved reenactment of already performed event sequences, rather than enactment of other persons’ actions. Salmon, Bidrose, and Pipe (1995) also used a reenactment procedure to assess young children’s memory for previously performed actions.
abilities to describe past events. In support of this hypothesis, Gordon et al. (1993) reported that 5-year-olds’ language abilities were related to the amount of detail they provided in their accounts of their physical examinations. Moreover, there is a substantial body of literature illustrating that components of behavioral style are predictive of social adaptation (e.g., Bates, Maslin, & Frankel, 1985; Cameron, 1977; Thomas et al., 1968), and demonstrations of the relevance of behavioral characteristics to cognitive performance are now being reported. For example, Pasilin (1986) found that maternal ratings of their preschoolers’ persistence and attention to tasks were positively related to the children’s achievement test performance. In addition, the findings of Gordon et al. (1993) suggest that temperament may account for some within-group variation in recall performance, with components of children’s temperament accounting for 12% to 24% of the variance in same-aged children’s recall performance. To illustrate, 3-year-olds who tended to approach new situations provided more information in response to open-ended questions than 3-year-olds who tended to withdraw.

In the present study, estimates of language skills and behavioral style were related to older and younger children’s performance under both the Verbal and Enactment interview conditions because it seemed likely that the impact of these individual difference factors would vary with children’s age and the conditions under which their recall is assessed. More specifically, language skills may be associated with children’s performance in the Verbal condition, but not in the Enactment setting, in which linguistic demands are reduced. Similarly, behavioral styles may be differentially linked to children’s responses to the two interview conditions. For example, a highly distractible child may perform more poorly in the Enactment condition, in which the props can be viewed as attractive playthings, than a less distractible child, whereas distractibility may be unrelated to performance in the Verbal condition. Furthermore, both language skills and behavioral styles may be particularly important in explaining the performance of younger, rather than older, children, for whom the recall task seems to be most challenging.

To summarize, the present investigation had two general goals: (1) To determine whether a contextually supportive Enactment procedure could enhance the recall performance of young children and reduce age differences in remembering, and (2) to examine the relation between children’s language skills and stable behavioral characteristics and their performance in the Verbal and Enactment interview conditions.

METHOD

Participants

The children who took part in this study were sampled from among the 3- and 5-year-olds who were scheduled for routine physical examinations at two private pediatric practices in a suburban area. A total of 32 3-year-olds (M = 37.3 months, range = 34 to 42 months) and 30 5-year-olds (M = 62.2 months, range = 57 to 71 months) participated in this research. Most of the children in the sample were White (87.1%), and about half of the children were female (51.6%). The participants were primarily from middle-class families, as indicated by their parents’ average level of education. Approximately 95% of the mothers and 89% of the fathers had some college education. Recruitment involved telephoning parents prior to their children’s scheduled examinations and inviting them to take part in the research. Of the families contacted, 66% agreed to participate, and of these, 77% completed the study. The most frequently cited reasons for declining to participate and dropping out were lack of time and the inconvenience of visiting the university laboratory.

Experimental Design and Materials

Children were questioned about the details of their recently completed physical examinations according to either a verbal or an enactment interview protocol. Random assignment of 3- and 5-year-olds to either the Verbal or Enactment condition resulted in the establishment of four groups of participants, ranging in size from 10 to 12 children. The Enactment interview was designed to maximize the level of contextual support provided to the children. In this condition, the interview took place in a room that resembled a pediatrician’s examining suite. The room contained a sink, a physician’s examining table, a doll, and props with which the children could demonstrate the procedures included in their physical exams. Table 1 displays a list of the props in the enactment room. The props included medical instruments, some of which may have been used during the individual children’s examinations (e.g., reflex hammer), as well as distracter items that are never used during routine checkups (e.g., ace bandage). Children in the Verbal condition were interviewed in a room that was similar but did not contain an examining table or instruments and props resembling those used by medical personnel. These children were interviewed with a standard verbal protocol that has been used in previous studies of children’s memory for personally experienced events (e.g., Baker-Ward et al., 1993).
Enactment interview was identical in structure to the Verbal interview protocol, except that children were instructed to both show and tell what had happened during their checkups.

Four dolls were used in the study, and each participant in the Enactment condition was provided a doll that was consistent with his or her race and gender. The dolls were dressed in street clothing and were not sexually anatomically detailed.

The Physical Examinations

A total of 14 pediatricians and 10 nurses participated in this study. Prior to the study, discussion with physicians resulted in the definition of the physical examination in terms of a collection of subevents or features. These features, listed in Table 2, consisted of procedures that are typically administered during check-ups. Although the office visits varied somewhat across age and among children at any given age, all children received a subset of the procedures indicated in Table 2. In addition, a photograph was taken of each child by a nurse during the physical examination, in an effort to differentiate the to-be-remembered examination from other check-ups.

To facilitate the coding of the interviews, immediately after each check-up, the participating physicians and nurses completed checklists to indicate precisely the component features of each child’s physical examination. As an accuracy check, parents also were asked to complete these checklists during the check-ups. Parents’ reports were typically in agreement with those of nurses and physicians; indeed, agreement was approximately 95%. For the purposes of scoring the children’s reports, a feature was considered present if either the doctor, nurse, or parent included it in the checklist.

Procedure

An interviewer met with participating parents and children at the pediatric office just before the scheduled time of the child’s examination. The parents then provided written consent, verbal assent was obtained from the child, and the medical staff administered the check-ups. Following the child’s examination, several individual difference assessments were made (see section below on individual differences), and the first interview was scheduled for a time between 3 and 7 days after the check-up. A follow-up interview was conducted approximately 6 weeks (between 39 and 49 days) following the physical examination. Each child was seen on all occasions by one of six female interviewers. Recall assessments were conducted in one of the two university laboratory rooms described previously. All sessions were videotaped for subsequent analysis.

Interviews

Both the Verbal and Enactment interview protocols involved several levels of questions of increasing specificity. Each session began with the establishment of rapport, and an explanation about the purpose of the child’s visit with the interviewer. The child was directed to try to remember the details of the check-up during which the nurse had taken his or her picture. In the Verbal condition, the experimenter then began the interview. In the Enactment condition, the interviewer indicated that the room contained many of the things the doctor and the nurse had used in the check-up and that the child could use those objects and the doll to demonstrate the procedures included in his or her examination. The photo feature was then
used as a practice item: The interviewer demonstrated the photograph procedure for the child, and then asked the child to demonstrate, as well. Following this practice, the interview began.

The interviews for both conditions were hierarchically organized, beginning with general, open-ended questions, followed by more specific questions about those features not already recalled. Questioning in the Enactment condition differed from that in the Verbal condition only in that children were asked to both “show and tell” rather than just “tell” the interviewer what happened. To illustrate the mode of questioning, consider the following four levels of questions addressed to the children: (1) general demonstration probes (e.g., “Show (tell) me what happened during your check-up.”), (2) more specific verbal (yes/no) questions (e.g., “Did the doctor check your eyes?”), (3) specific elaboration probes (e.g., “Show (tell) me what the doctor did when he checked your eyes.”), and (4) pairs of leading questions (e.g., “Did the doctor shine a light in your eyes?” and “Did the doctor put some drops of water in your eyes?”). Note that in the final level of questioning, one of the two questions gave correct information about how a particular procedure may have been administered, whereas the other included incorrect detail about the procedure. The order of the two leading questions within the pair varied across features. Moreover, these questions were only asked when children admitted that they did not know the answer to or failed to respond to the previous question about that same procedure.

Throughout the interview, children in the Enactment condition were encouraged to use the doll and the instruments to demonstrate the procedures of their physical examinations. Children in both conditions were asked to provide as much information as possible in response to the general probes (e.g., “Show (tell) me what else happened.”) before they were questioned more specifically. For each procedure recalled, they were asked to elaborate, or describe in detail, that feature. The order of specific questioning was counterbalanced within each interview condition. One-half of the children were asked specific questions about the nurse features, followed by specific questions about the doctor features, whereas the other half were questioned in the reverse order.

In addition, to gauge the accuracy of the children’s responses, two types of yes/no questions were included about events that did not take place in the examination: (1) “Absent Feature” questions about procedures that could be included in a physical examination but which were not a part of a particular child’s examination (e.g., “Did you get a shot?” when, in fact, the child had not received a shot), and (2) “Extra-Event” questions about features that would be unlikely to occur in a physical examination, but which were also medically related (e.g., “Did the nurse take your temperature?”). These Extra-Event questions are listed in Table 3.

### Individual Difference Measures

#### Language Development

After each child’s physical examination, the Test of Early Language Development (TELD: Hresko, Reid, & Hammill, 1981) was administered. The TELD items probe children’s linguistic form and content in both the expressive and receptive modes, yielding a standardized score, or Language Quotient. This assessment is appropriate for measuring the language skills of children between the ages of 3 and 7 years. The standardized sample included primarily Caucasian children from middle-class families. Hresko et al. (1981) reported Cronbach’s coefficient alphas, calculated separately for five age groups ranging from 3 to 7 years, between .87 to .92, indicating that internal consistency reliability was high. Good content validity has been demonstrated by correlations between TELD scores and other language assessments ranging from .46 to .80 (e.g., Preschool Language Scale: Zimmerman, Steiner, & Pond, 1979).

#### Behavioral Style

During the children’s first interview, parents completed a standardized questionnaire concerning the behavioral style, or temperament, of their child. The Parent Form of the Temperament Assessment Battery For Children (TABC: Martin, 1988) contains 48 questions to which the parent is asked to respond, using a 7-point Likert scale. The measure yields scores on six personality-behavioral dimensions: Activity, Adaptability, Approach-Withdrawal, Emotional Intensity, Ease of Management, and Persistence. Martin (1988) reported reliability information based on a sample comparable to the participants in the present investigation. Internal consistencies of the six scales for the

<table>
<thead>
<tr>
<th>Table 3 Extra-Event Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did the nurse take your temperature?</td>
</tr>
<tr>
<td>Did the nurse ask you to drink some medicine?</td>
</tr>
<tr>
<td>Did the nurse wrap a bandage around your leg?</td>
</tr>
<tr>
<td>Did the nurse ask you to lift something to see how strong you are?</td>
</tr>
<tr>
<td>Did the doctor check your head for ticks?</td>
</tr>
<tr>
<td>Did the doctor ask you to touch your nose with your fingers?</td>
</tr>
<tr>
<td>Did the doctor put a washcloth on your head?</td>
</tr>
<tr>
<td>Did the doctor clean your belly button with some soap and water?</td>
</tr>
</tbody>
</table>
Parent Form were good, as illustrated by Cronbach’s coefficient alphas between .65 and .87. Interrater reliability of the Parent Form, calculated by comparing fathers’ and mothers’ ratings of their children on the six temperament dimensions, was adequate; correlations between fathers’ and mothers’ ratings ranged from .34 to .64.

Because preliminary analyses of the current data indicated that the six scales of the TABC were not independent of one another, a smaller set of nonredundant behavioral style measures was selected for analysis in this investigation. This selection process was based, in part, on an exploratory factor analysis with a varimax rotation. The results yielded a factor that seemed to reflect a Manageability (or Easiness-Difficulty) dimension of personality, because Ease-of-Management and Adaptability had high positive loadings, \( r(60) = .81 \) and \( .85 \), respectively, whereas Emotional Intensity had a high negative loading, \( r(60) = -.77 \). An Easiness-Difficulty dimension is well established in the temperament literature and has been shown to be related to adaptation across a range of settings (e.g., Bates et al., 1985; Cameron, 1977; Thomas et al., 1968). Because such a factor is suggested by both the current data and the temperament literature, each child’s Adaptability and Ease-of-Management scores and the inverse of his or her Emotional Intensity score were combined to form a composite Manageability measure. In addition, the Persistence scale was chosen for the analyses because of the six behavioral characteristics measured by the TABC, persistence would seem to be most directly related to performance on cognitive tasks. Indeed, a similar behavioral dimension has been found to be predictive of preschool children’s achievement test scores (Pasinlin, 1986). Moreover, Persistence did not load on the Manageability factor. Thus, the two aspects of behavioral style used in further analyses were Manageability and Persistence.

Coding of the Recall Protocols

Remembering

Each interview was scored with regard to performance on three types of features or procedures. Present Features consisted of procedures that had been carried out during the individual children’s examinations, and thus varied from child to child. The mean number of procedures administered for 3-year-olds was 16, whereas the average number of Present Features for 5-year-olds was 17. Absent Features were defined as medical procedures that have high probabilities of occurring during routine physical examinations but that did not occur during an individual child’s check-up. The number of Absent Features thus varied from child to child, as a function of the content of each check-up, with the average number being 3 and 2 for the 3- and 5-year-olds, respectively. The Extra-Event Features were the eight medically related procedures listed in Table 3. These procedures were considered to be medically plausible and could have been experienced by the children in other settings, but were highly unlikely to occur during a well-child examination.

Children’s responses to questions about Present Features were coded for features recalled and the type of probe (open-ended versus specific) needed to elicit the information. Total recall scores were calculated for each child by summing the proportion of features recalled in response to specific and open-ended questions. A miss was coded when a child incorrectly denied the occurrence of a Present Feature. There were three types of codes a child could receive for Absent Features: intrusions, false alarms, and correct rejections. Intrusions were reports of such features at the open-ended level of questioning. For example, an intrusion was scored if, when asked “What happened during your check-up?” a child reported getting a vision test when, in fact, he or she had not received such an assessment. Because the interviewers were blind to which procedures had been carried out during the individual children’s examinations, specific questions were asked about Absent Features not already reported in response to open-ended questions. False alarms occurred when a child incorrectly said “yes” to a specific question about an Absent Feature, whereas a correct rejection was coded if the child denied the occurrence of an Absent Feature. A similar system was used to code children’s responses with regard to the Extra-Event Features. Thus, possible codes on Extra-Event Features were the same as those on Absent Features (i.e., intrusions, correct rejections, and false alarms).

Elaboration

Children’s recall was also characterized in terms of the amount of elaborative detail provided verbally or through demonstration about each feature. A 2 was scored if the elaboration about a feature was complete (e.g., “The doctor hit my knee with a hammer.”), a 1 indicated that some detail was present but that the description was incomplete (e.g., “The doctor hit my knee.”), and a 0 indicated that the child was unable to describe or demonstrate the procedure. For the purposes of calculating average elaboration scores, incorrect elaborations were scored as 0s, but were tagged
Greenhoot et al. 369

for further analyses. Children in both interview conditions could receive credit for both verbal and non-verbal information.

Reliability

Reliability in the coding of recall and elaboration was quite good. The Present, Absent, and Extra-Event Feature scoring was completed by six research assistants, and a “master” coder scored 20% of each assistant’s work. Percent agreement between each coder and the master coder was at least 90%, averaging 93%. Similarly, the elaboration data were scored by five research assistants, and a master coder scored 20% of each assistant’s assignment. Percent agreement between each research assistant and the master coder was, at minimum, 90% and averaged 94%.

Memory/Play Discriminations

Because children in the Enactment condition were at times “off task,” several criteria were adopted in order to distinguish between memory reports and play. Demonstrations that were in direct response to an interviewer’s probe were counted in the children’s memory reports. For example, if the interviewer instructed a child to “show me what the doctor did,” and the child responded by immediately picking up the otoscope and checking the dolls’ ears, that procedure was counted in the memory report. Moreover, if in response to that same prompt, the child had used the otoscope to check the doll’s hair for ticks, the demonstration was also counted in the child’s memory report. Children could also verbally indicate that a particular demonstration represented a procedure present in their examinations. For example if the a child said “she did this . . .” while enacting a procedure, the demonstration was included in the memory report. If neither of the latter two conditions applied, the interviewer asked the child whether he or she was demonstrating a procedure that had occurred during his or her exam. Only if the child responded affirmatively was the demonstration in question counted in the memory report. At the initial interview, 10% of the 3-year-olds’ demonstrations and 6% of the 5-year-olds’ demonstrations did not meet these criteria and were therefore coded as play. The proportion of “off-task” demonstrations during the second interview averaged 10% for the younger children and 4% for the older children.

RESULTS

Preliminary analyses indicated no differences in performance as a function of interviewer, order of questioning, pediatric practice, gender, or ethnicity. Consequently, for further analyses, the data were collapsed across these variables. The general analysis strategy was carried out in two stages. In the first step, the effects of age and interview condition on children’s accounts at the two interviews were examined. In the second stage, the relations between the two domains of individual difference variables and children’s memory reports were explored. The results of each of these steps will be discussed in turn.

Age and Interview Condition

Repeated measures analyses of variance (ANOVAs) were used to relate age and interview condition to selected measures of recall performance at the two assessments and to examine patterns of change over the 6-week delay period. For each performance variable, age and interview group were treated as between-participants factors, and time of interview was considered a within-participants factor. Significant main effects of time or interactions involving time were followed up with separate 2 × 2 (age × interview condition) univariate ANOVAs on the 1-week and 6-week data.

Present Features

Recall scores for Interviews 1 and 2 are shown in Figure 1A and B, respectively. As can be seen, 5-year-olds showed better overall recall performance than 3-year-olds, $F(1, 58) = 4.14, p < .05$. The dark portions of the bars in Figure 1 represent the proportion of features recalled in response to open-ended probes. As the graphs illustrate, the Enactment condition seemed to enhance open-ended recall, particularly for the 3-year-olds, as reflected in a significant age × condition interaction, $F(1, 58) = 4.94, p < .05$. It is also apparent in Figure 1 that the patterns of correct responses to specific questions (yes/no) are essentially the inverse of those made to open-ended questions, and that incorrect responses (i.e., misses) to these questions did not vary with age or interview condition. Moreover, recall performance did not change significantly between the 1-week and 6-week interviews. At both interviews, the majority of the 3-year-olds’ recall in the Verbal condition was elicited by specific questions, whereas 3-year-olds in the Enactment condition and 5-year-olds recalled a greater proportion in response to open-ended questions than specific questions.

To understand further the children’s memory reports, an additional analysis explored the extent to which recall was accompanied by elaborative detail.
The 0–2 coding scheme described earlier was used to construct an elaboration score for each child. Table 4 shows the average elaboration scores by age and interview condition for the 1-week and 6-week interviews. As can be seen, 5-year-olds seemed to provide more elaborative accounts of their check-ups than 3-year-olds. However, children at both age levels provided more detail if interviewed in the Enactment condition than if questioned with the Verbal protocol. Consistent with visual inspection of the data, the elaboration scores differed as a function of both age and interview group, $F_{s}(1, 57) = 41.49, p < .0001$, and these patterns of performance did not change over the 6-week delay period.

In sum, the analyses of Present Feature recall suggested that children were able to provide richer, more spontaneous accounts of their physical examinations when given the opportunity to act out their check-ups than when simply asked to talk about them. Enactment was particularly beneficial to the open-ended reports of the younger children, and was supportive of the provision of elaborative detail by both 3- and 5-year-olds.

Nonadministered Features

Because the set of medically related Absent Features was quite small and the children in each of the four groups responded comparably to questions about these procedures and the eight Extra-Event features, the data with regard to these two types of features were combined. This resulting set of Nonadministered Features, or procedures that had not been carried out, averaged 11 for the 3-year-olds and 10 for the 5-year-olds. A summary of children’s responses to questions about these features is shown in Figure 2.

As can be seen in the dark portions of the bars, the Enactment condition was associated with a higher proportion of intrusions of Nonadministered Features for the 3-year-olds at the 1-week interview, and for both age groups at the second assessment, in comparison with the other children. Consistent with these observations, across-time analyses of the children’s intrusions revealed a main effect of interview group, $F_{1, 58} = 8.71, p < .01$, and a marginal age × group × time interaction, $F_{1, 58} = 3.67, p = .06$. The interaction reflected the finding that at the first interview there was an age × group interaction, $F_{1, 58} = 9.13, p < .01$, whereas at the second assessment, a main effect of interview group was obtained, $F_{1, 58} = 7.66, p < .01$. Inspection of Figure 2 also indicates that the 3-year-olds tended to have lower proportions of correct rejections and correspondingly higher percentages of false alarms than the 5-year-olds, $F_{s}(1, 58) = 9.35, p < .01$. Although there were no significant changes in the proportions of correct rejections over time, the overall rate of false alarms increased signif-

---

### Table 4  Average Elaboration Scores (and Standard Deviations) by Age and Interview Condition

<table>
<thead>
<tr>
<th>Age</th>
<th>Week 1</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Verbal</td>
<td>Enactment</td>
</tr>
<tr>
<td>3-Year-olds</td>
<td>0.36 (0.39)</td>
<td>1.26 (0.52)</td>
</tr>
<tr>
<td>5-Year-olds</td>
<td>1.15 (0.31)</td>
<td>1.66 (0.25)</td>
</tr>
</tbody>
</table>

---

**Figure 1**  Present Feature performance at interviews 1 and 2, as a function of age and interview group. Proportions of open-ended recall, specific recall, and misses may not add to 100% due to the occurrence of “don’t know” responses.
Unrelated Features

Children sometimes “recalled” procedures that were not included in the predefined set of Nonadministered Features and that were not reported by either their parents or the medical personnel. For example, in response to an open-ended probe, one child demonstrated how the doctor presumably wrapped her head in a bandage, whereas another child described how the nurse spread lotion all over her stomach! For each child, the number of these Unrelated Feature intrusions was tabulated. Note that as these items were not a part of the predefined set of procedures, the scores represent frequencies rather than proportions. As shown in Table 5, across the two interviews, 3- and 5-year-olds in the Enactment condition made this kind of intrusion with a much greater frequency than children in the Verbal condition, although the difference was significantly more pronounced for the 3-year-olds, $F(1, 58) = 8.72, p < .01$. Moreover, there was a significant interaction between time and interview condition, $F(1, 58) = 7.21, p < .01$, such that the frequency of Unrelated Feature intrusions increased from the 1-week interview to the 6-week assessment in the Enactment condition, but not in the Verbal condition.

These findings indicate that although enactment seemed to enhance reporting of Present Features, it was also associated with poor performance in regard to the Nonadministered Features. The Enactment condition resulted in an increased likelihood of open-ended or spontaneous reports of incorrect information for both older and younger children, particularly after a long delay period, although this effect was more pronounced for the 3-year-olds. Taken together, these results seem to suggest that enactment, in contrast to a verbal interview, may result in declines in the accuracy of young children’s open-ended reports.

In order to examine further the effects of enactment on the accuracy of recall, three additional sets of analyses related the children’s reports of Present Features to their performance on questions about Nonadministered Features.

Accuracy of Open-Ended Reports

The analysis of the accuracy of children’s open-ended reports involved the determination of the proportion of spontaneously reported information that was incorrect. For each child, the number of examination features that were produced in error (i.e., intrusions) was taken as a proportion of the total number of procedures reported in response to open-ended questions. The resulting proportions of incorrect open-ended responses are shown as a function of age, inter-

Table 5  Mean Frequency (and Standard Deviations) of Unrelated Feature Intrusions

<table>
<thead>
<tr>
<th></th>
<th>Week 1</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Year-olds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>.06 (.25)</td>
<td>.25 (1.00)</td>
</tr>
<tr>
<td>Enactment</td>
<td>2.56 (2.19)</td>
<td>4.31 (4.09)</td>
</tr>
<tr>
<td>5-Year-olds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>.07 (.26)</td>
<td>.00 (.00)</td>
</tr>
<tr>
<td>Enactment</td>
<td>.27 (.80)</td>
<td>1.13 (2.67)</td>
</tr>
</tbody>
</table>

Figure 2  Nonadministered Feature performance at interviews 1 and 2, as a function of age and interview group. Proportions of open-ended recall, specific recall, and misses may not add to 100% due to the occurrence of “don’t know” responses.
view group, and time in Table 6. As these means illustrate, the 3-year-olds in the Verbal condition and the 5-year-olds in both interview groups were rarely incorrect in their reports. A notable exception to this pattern of accuracy, however, can be seen in the performance of the 3-year-olds in the Enactment condition. These children were more likely than the other participants to be in error when they reported a feature in response to an open-ended probe, as reflected in a significant age × group interaction, $F(1, 47) = 7.63, p < .01$. In addition, the analysis revealed an interaction between time and group, $F(1, 47) = 6.43, p < .02$, reflecting the finding that the proportion of incorrect reports of features at the open-ended level increased over time for both the 3- and the 5-year-olds in the Enactment condition, but not for those in the Verbal condition.

### Accuracy of Responses to Questions about Open-Ended Demonstrations

As described earlier, the children in the Enactment condition were sometimes asked whether procedures they spontaneously demonstrated had actually been carried out during their examinations. On average, the 3-year-olds were questioned about eight demonstrations at each interview, whereas the 5-year-olds were asked about three demonstrations. The accuracy of children’s responses to these specific demonstration probes was evaluated by calculating the proportions of correct and incorrect answers to questions about demonstrations of Present and Nonadministered procedures. These proportions are shown in Table 7, as a function of age and delay. As can be seen in the top portion of Table 7, both the older and the younger children were quite accurate in their responses to specific demonstration probes about Present Features. However, as shown in the bottom portion of the table, when questioned about demonstrations of procedures that had not actually been performed during their examinations, both the 3- and the 5-year-olds often incorrectly claimed that the procedures had occurred.

### Accuracy of Responses to Specific (Yes/No) Questions

Paralleling the analyses of the children’s open-ended recall, the accuracy of their responses to specific (i.e., yes/no) probes was examined by calculating the proportions of correct and incorrect answers given to questions about both Present and Nonadministered Features. Note that these proportions differ from those presented earlier (i.e., the specific recall and misses illustrated in Figure 1, and the false alarms and correct rejections shown in Figure 2) in that their calculation was based only on the subsets of Present and Nonadministered Features about which specific questions were asked, rather than the total numbers of Present and Nonadministered Features. On average, the children were asked 11 specific questions about presented features and 10 yes/no questions about procedures that had not been administered. The rescaling based on these subsets of features permitted the use of $t$ tests to compare the means against chance levels of responding (.50). The proportions of correct and incorrect responses to questions about Present and Nonadministered Features (i.e., the rescaled specific recall, miss, false alarm, and correct rejection scores) are shown in Table 8. As can be seen in the top panel of Table 8, there were no age or interview condition differences in the accuracy of responding to questions about Present Features at either assessment point. Moreover, a series of $t$ tests indicated that neither the 3- nor the 5-year-olds responded to these questions at better than chance levels at the 1- and 6-week assessments. However, as is apparent in the bottom panel, in comparison to the 3-year-olds, the older children did provide more correct and fewer incorrect

### Table 6 Mean Proportions (and Standard Deviations) of Features Reported in Response to Open-ended Questions that were Recalled in Error

<table>
<thead>
<tr>
<th></th>
<th>Week 1</th>
<th>Week 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Year-olds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>.10 (.16)</td>
<td>.11 (.21)</td>
</tr>
<tr>
<td>Enactment</td>
<td>.33 (18)</td>
<td>.43 (22)</td>
</tr>
<tr>
<td>5-Year-olds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>.09 (15)</td>
<td>.04 (.09)</td>
</tr>
<tr>
<td>Enactment</td>
<td>.07 (11)</td>
<td>.12 (17)</td>
</tr>
</tbody>
</table>

### Table 7 Accuracy of Responses to Specific Demonstration Questions in the Enactment Condition

<table>
<thead>
<tr>
<th></th>
<th>Week 1 Response</th>
<th>Week 6 Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct (“Yes”)</td>
<td>Incorrect (“No”)</td>
</tr>
<tr>
<td></td>
<td>Incorrect (“Yes”)</td>
<td>Correct (“No”)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Present Features</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Year-olds</td>
<td>.76 (.30)</td>
<td>.06 (.13)</td>
</tr>
<tr>
<td>5-Year-olds</td>
<td>.72 (.44)</td>
<td>.11 (.33)</td>
</tr>
<tr>
<td></td>
<td>.81 (.25)</td>
<td>.09 (.14)</td>
</tr>
<tr>
<td></td>
<td>.85 (.15)</td>
<td>.15 (.23)</td>
</tr>
<tr>
<td>3-Year-olds</td>
<td>.35 (.34)</td>
<td>.44 (.32)</td>
</tr>
<tr>
<td>5-Year-olds</td>
<td>.47 (.45)</td>
<td>.36 (.43)</td>
</tr>
<tr>
<td>3-Year-olds</td>
<td>.18 (.27)</td>
<td>.65 (.33)</td>
</tr>
<tr>
<td>5-Year-olds</td>
<td>.64 (.38)</td>
<td>.36 (.38)</td>
</tr>
</tbody>
</table>

**Note:** Standard deviations are in parentheses.
responses to questions about Nonadministered Features, \( F(1, 58) = 9.90, p < .01 \). A series of \( t \) tests indicated that the younger children responded randomly to yes/no questions about Nonadministered Features, whereas the older children performed well in response to these questions.

In order to explore further the accuracy of the children’s responses to specific, yes/no questions, an additional analysis involved calculating a measure of response bias suggested by signal detection theory. According to this perspective (see, e.g., Snodgrass & Corwin, 1988), “items” that children are asked about during the interview are assumed to vary along a continuum of familiarity or strength. Although remembering is determined in part by memory strength, each participant’s responding to yes/no questions about Present and Nonadministered Features is assumed to be based on a personal criterion of strength such that a procedure is identified as having been administered during the check-up if its familiarity exceeds this response criterion. Within the theory, the distance between the criterion and the intersection between the strength distributions associated with the Present and Nonadministered Features, or \( C \), is thought to reflect response bias: a child with a neutral bias should have \( C \) equal to 0, a child with a liberal bias (biased to respond “yes”) should have a positive \( C \) score, and a child with a conservative bias (biased to respond “no”) should have a negative \( C \) score. To examine overall response bias, \( C \) values were calculated for each child at each interview, and are summarized in Table 9. As can be seen, mean \( C \) scores for the 3- and 5-year-olds at both interviews were quite low in absolute value; moreover, an ANOVA confirmed that there were no age or interview condition differences in response bias.

### Individual Differences

Several analyses of covariance (ANCOVAs) with heterogeneous slopes were carried out in order to examine the relation between various aspects of the children’s memory reports and the individual difference measures. The goal of these analyses was to test individual differences in the TELD Language Quotient and selected behavioral style measures as moderators of the effects of age and interview condition on recall performance. The analyses were carried out separately for measures of memory at the 1-week and 6-week interviews. For each measure of recall, an ANCOVA was performed in which age and interview condition were categorical factors, and one of the individual difference measures was included as a continuous variable. In each ANCOVA, the slope of the relation between an individual difference variable and the recall measure was allowed to vary according to age and interview group. That is, four slopes were estimated (one for each of the four cells formed by crossing age and interview condition), and their equivalence was then tested in order to determine whether the relations between each individual difference measure and recall varied across the cells. The analyses involving language and behavioral style will be discussed in turn.

### Language Skills

The results of the ANCOVA involving the TELD language quotient did not reveal any significant rela-
tions between language scores and measures of recall performance, either within or across age groups and interview conditions.

Behavioral Style

For each measure of recall at each interview, two ANCOVAs with heterogeneous slopes were performed. In the first ANCOVA, Manageability was used as the continuous variable, whereas Persistence was included in the second analysis. The results of these ANCOVAs indicated that both Manageability and Persistence were related to important aspects of the children’s memory reports, but that the relation between these two aspects of behavioral style and recall performance varied substantially according to both age and interview condition. Overall, behavioral styles were important in predicting the performance of the 3-year-olds, but not of the 5-year-olds.

Manageability and remembering. Manageability was positively associated with the 3-year-olds’ intrusions of both Nonadministered and Unrelated Features at the 1-week interview, but only in the Enactment condition. For each of these two outcome measures, the slope and intercept parameter estimates yielded by the ANCOVA were used to plot separate regression lines for each of the four cells. As suggested by the plot in Figure 3A, a test of heterogeneity of slopes indicated that the relation between the intrusions of Nonadministered Features and Manageability differed marginally as a function of age and interview condition, $F(1, 52) = 3.76, p < .06$. Follow-up tests confirmed that Manageability was marginally related to Nonadministered Feature intrusions for the 3-year-olds in the Enactment condition, $F(1, 52) = 3.42, p < .07$, but not for the other three groups of children. As the plot in Figure 3A illustrates, increases in Manageability were associated with increases in intrusions among the younger children interviewed in the Enactment context. Similarly, Figure 3B illustrates the relation between Manageability and the children’s intrusions of Unrelated Features. As can be seen, the slopes varied as a function of age and interview group, $F(1, 52) = 4.16, p < .05$, indicating that the 3-year-olds in the Enactment condition differed from the other children. Thus, Manageability was positively related to the number Unrelated Feature intrusions made only by the 3-year-olds in the Enactment condition, $F(1, 52) = 9.85, p < .01$. However, in contrast to these analyses of the 1-week data, comparable ANCOVAs carried out on the 6-week data failed to reveal any significant relationships between Manageability and memory performance.

Persistence and remembering. Whereas the ratings of Manageability predicted the performance of the 3-year-olds in the Enactment condition, judgments of Persistence were related to the younger children’s performance in the Verbal condition. As with the analyses involving Manageability, the slope and intercept estimates obtained in the ANCOVAs were used to plot the regression lines for each of the four cells. The graph in Figure 4A shows the regression lines for the prediction of total recall at the initial interview. As can be seen, the slopes differed, $F(1, 52) = 8.02, p < .01$, with follow-up tests indicating that Persistence was negatively related to total recall for 3-year-olds in the Verbal condition, $F(1, 52) = 17.27, p < .0001$, and was unrelated to performance in the three other cells of the design. Note, however, that the direction of this relationship was unexpected: the more persistent the child, the lower the recall. Similarly, the
slope of the relation between Persistence and specific recall at the 1-week interview differed among the four groups of participants, F(1, 52) = 12.98, p < .001. As illustrated by the plot in Figure 4B, Persistence was negatively related to the proportion of specific (i.e., yes/no) recall for 3-year-olds in the Verbal condition, F(1, 52) = 20.58, p < .0001, but not for the other three groups of participants. (A comparable analysis was conducted on the rescaled specific recall scores used in the accuracy analyses, with similar results being obtained.) The similarity of this pattern to that for total recall should not be surprising, given that specific recall accounted for most of the 3-year-olds’ total recall scores in the Verbal interview group. Again, consistent with the Manageability findings, it should be noted that the linkages between Persistence and recall were not observed at the second interview.

Figure 5 shows the estimated regression lines for the prediction of false alarms at the initial and follow-up interviews, respectively. Tests of the heterogeneity of slopes were significant for both the 1- and 6-week data, Fs(1, 52) ≥ 6.60, ps < .05. At both assessments, Persistence was negatively related to the likelihood of false alarms for 3-year-olds questioned with the Verbal interview, Fs(1, 52) ≥ 10.33, ps < .01.

These findings indicate that in the Verbal condition, less persistent 3-year-olds scored higher on total and specific recall, and were more likely to have false alarms.

Paralleling the exploration of the linkage between Persistence and specific recall, additional analyses focused on the association between Persistence and the proportion of incorrect “yes” responses to specific questions about Nonadministered features (i.e., the rescaled false alarms). As was the case with the original false alarm data, at both interviews Persistence was negatively related to the proportion of “yes” responses to questions about Nonadministered features.
alarms, than 3-year-olds who are better able to attend to tasks. This pattern is consistent with the view that the less persistent 3-year-olds may have been more likely to respond “yes” to specific questions, regardless of whether the procedure in question had been administered during their check-ups. Thus, to determine the extent to which the effects of Persistence, discussed above, may be accounted for by “yes” response bias, an ANCOVA with C as the dependent variable, age and interview group as categorical factors, and Persistence as a continuous predictor, was performed. As in the previous analyses involving Persistence, the slope of the relation between Persistence and C was allowed to vary according to age and interview condition, and the resulting parameter estimates were used to plot separate regression lines for the four cells. As suggested by the graph in Figure 6A, for the 1-week data, the slopes differed as a function of age and interview group, $F(1, 52) = 13.00, p < .001$, with follow-up tests indicating that Persistence was positively related to C for 3-year-olds interviewed with the Verbal protocol, $F(1, 52) = 22.58, p < .0001$, but not for the other three groups of children. At the 6-week interview, as shown in Figure 6B, the slopes differed according to interview group, $F(1, 52) = 8.70, p < .01$. However, follow-up tests indicated that Persistence significantly contributed to the prediction of C only for 3-year-olds in the Verbal condition, $F(1, 52) = 4.98, p < .05$, suggesting that the variation in slopes according to group was due primarily to this cell. Thus, in the Verbal interview context at both time points, the less persistent a 3-year-old was, the lower the C value, or the more liberal the child’s criterion was, and the more likely he or she was to respond affirmatively to a specific question, regardless of whether that particular feature had actually occurred.

**DISCUSSION**

The results of this study indicate that the context of memory assessment may have marked effects on children's recall performance, and that these effects may vary as a function of personal characteristics of the children. Previous findings regarding age differences in memory reporting (e.g., Baker-Ward et al., 1993) were replicated in that at both 1 and 6 weeks following their examinations, the younger children provided descriptions of their check-ups that were less spontaneous, complete, and elaborate than those generated by the older children. Moreover, the younger children’s responses to yes/no questions were consistently less accurate than those of the older children, and, in fact, the 3-year-olds’ answers to specific questions about both Present and Nonadministered Features did not differ from chance levels. Further, even the 5-year-olds performed poorly on specific questions about Present Features. It must be noted, however, that because yes/no questions about Present Features were only asked about procedures not mentioned in response to more general questions, features probed at the specific level of questioning may be those for which the children have weaker memory traces. In addition, there were relatively few changes in the children’s recall of Present Features in the interval between the first and second interviews. That performance changed little over time should not be surprising, given that the first assessment did not take place until 3 to 7 days after the original event, and studies of retention since the time of Ebbinghaus (1913) have shown that the steepest declines in performance occur during the period immediately after an event or original learning.
Consistent with expectations regarding the effects of the interview context, enactment did lead to reductions in age differences in performance, particularly at the open-ended level of questioning. Children who were given the opportunity to act out the details of their check-ups were able to provide more spontaneous, elaborate reports of procedures that had occurred during their examinations than those who were asked to talk about them, and this pattern was particularly true among the 3-year-olds. However, the Enactment condition also resulted in an increased likelihood of reports of incorrect information by younger children at the first interview, and by children in both age groups after the 6-week delay period. Thus, children in the Enactment condition were more likely than those who received the Verbal interview to spontaneously report medical procedures that had not been administered during their examinations, as well as actions that had not been included in the predefined set of features. Moreover, the analyses of the overall accuracy of children’s open-ended reports indicated that when the 3-year-olds in the Enactment condition reported a procedure in response to a general probe, they were incorrect over 30% of the time at the first interview and over 40% of the time at the second assessment! In contrast to the latter group, the 3-year-olds in the Verbal condition and the 5-year-olds were much less likely to be incorrect when they reported a feature at the open-ended level of questioning. Therefore, even though the Enactment context was designed to be supportive of young children’s recall, the overall pattern of performance indicated that the accuracy of their reports declined in this setting, particularly following long delays.

These findings notwithstanding, the exploratory analyses involving behavioral style suggest that the effects of age and interview condition may differ substantially as a function of personal characteristics of the children. For example, although there were no clear patterns in response bias as a function of age or group, an interesting pattern emerged when individual differences in Persistence were considered. As indicated by the analyses involving C, the younger children who were rated by their parents as having difficulty in completing tasks (i.e., those who were low in Persistence) were particularly at risk for adopting a “yes” response set during the Verbal interview. Likewise, the effect of enactment on young children’s reports of procedures that had not been administered seemed to be modified by another dimension of behavioral style. Thus, the 3-year-olds in the Enactment condition were more likely to make intrusions at the initial interview than were the other three groups of children, but this effect was especially pronounced among children who scored high on Manageability. Admittedly, the individual difference analyses need to be interpreted cautiously because some of the findings were unexpected, and others were inconsistent with those of prior investigations. For instance, children’s language skills were not related to their recall performance, whereas Gordon et al. (1993) observed that language ability predicted children’s elaboration of the details of their check-ups. The reasons for this discrepancy are not clear, particularly as it seems intuitive that aspects of language skill should mediate verbal recall. Of course, it is possible that better measures of language ability would be more consistently associated with children’s recall, but these inconsistencies do point to a general difficulty in identifying the sources of individual differences in remembering.

It is interesting that interview context, as well as behavioral style, seemed to be more heavily related to the performance of younger children than that of older children, particularly at the initial recall assessment. Given that the recall task is quite challenging for younger children, perhaps because their linguistic, narrative, and information processing skills are less well developed than those of older children, the present findings are consistent with accounts of the “zone of proximal development” (Vygotsky, 1978; see also Cox, Ornstein, & Valsiner, 1991). From this perspective, conditions that reduce task demands may serve to “scaffold” young children as they attempt challenging activities that they would be unable to complete alone. In contrast, older children should perform well on the same tasks under a broad range of conditions. Demonstrations of greater context specificity among younger, as opposed to older, children are commonplace in the literature on children’s recall and strategy use in laboratory settings (Cox et al., 1991; Folds et al., 1990). The current findings illustrate, as well, that young children are more sensitive than older children to contextual factors that both interfere with and support the remembering of salient “real-world” events. Extending the Vygotskian perspective still further, young children seem to be more sensitive than older children to the “goodness-of-fit” between personal characteristics and the specifics of the assessment condition, because the external supports associated with optimal performance seem to vary with the children’s temperament.

Why does the accuracy of some children’s reports suffer in the Enactment condition? One possibility is that many of the children were playing or acting upon their general knowledge of visits to the doctor, rather than attempting to provide accurate accounts of their own examinations. Moreover, the distracter items might have been so suggestive that the children could
and although casual observation of the enactment in-
planations for the children's behavioral productions,
routines. There are thus a number of competing ex-
od of Present Features may reflect play that is based to
pretation of the intrusion data, the children's reports
ments condition raise the possibility that the children's
may decline, reducing individual differences.
reprende to the interview setting. In addition, the "easy"
children may have been particularly vulnerable to the
suggestive nature of the props during the first inter-
view because they were more compliant and eager-
to-please than the "difficult" children. In support of
this interpretation, the literature on individual differ-
ces in interrogative suggestibility among adults in-
dicates that susceptibility to suggestion is positively
related to personality measures such as compliance
and fear of negative evaluation (Gudjonsson, 1988,
1991). After 6 weeks, however, it seems possible that
the children's weakened memory traces rendered
them unable to easily remember their check-ups, thus
leading to either an increased tendency to play or to a
greater susceptibility to suggestion. To a point, then,
task difficulty should lead to greater expression of in-
dividual differences in performance, but as the task
becomes increasingly hard, all children's performance
may decline, reducing individual differences.

Taken alone, the increased reporting of Present
Features that was associated with enactment might
indicate that either the cueing properties of the props
or the opportunity to act out the details of a check-up
facilitated the children’s memory. However, the in-
creased errors that were also observed in the Enact-
ment condition raise the possibility that the children's
reports may not reflect entirely their memories for a
specific check-up. That is, consistent with the inter-
pretation of the intrusion data, the children's reports
of Present Features may reflect play that is based to
some extent on their prior knowledge of physicians’
routines. There are thus a number of competing ex-
planations for the children’s behavioral productions,
and although casual observation of the enactment in-
terviews suggests that there may be examples of these
various behaviors in the reports of individual chil-
dren, there does not seem to be a reliable way to tease
apart memory, knowledge, and play responses on a
frame-by-frame basis. Although an attempt was made
to identify episodes of play during individual inter-
views, the occurrence of Unrelated Feature intrusions
suggests that the objective criteria used to discrimi-
nate between memory and off-task behavior did not
result in the elimination of all play behaviors. More-
ever, even when specifically asked about the accuracy
of some of their demonstrations, the children often
claimed that their play-like behaviors represented ac-
tual procedures from their check-ups. Accordingly,
a global analysis of performance is necessary to differ-
entiate between memory, knowledge, and play be-
behavior. For example, by taking into account the chil-
dren's performance with respect to both present and
nonadministered procedures, it would seem possible
to increase the discrimination between reports that re-
fect remembering and those that reflect knowledge
or are generated in play.

The analyses of the accuracy of the children’s open-
ended reports were conducted with these points in
mind, and serve to supplement to the impressions re-
sulting from the individual analyses of correct and in-
correct recall. Specifically, the finding that the 3-
year-olds in the Enactment condition were very often
incorrect in their open-ended recall further supports
the conclusion that the reports generated in the Enact-
ment condition, particularly those of the younger
children, were not indicative of memories of the
check-ups. Further, because the overall patterns of be-
avioral productions for the 3-year-olds in the Enact-
ment condition do not lead to a “diagnosis” of accu-
r ade remembering, it is still unclear whether age
differences in recall in general are more a result of de-
velopmental changes in memory or in the ability to
report what is remembered. Although it is clear that
in the Enactment condition some of the 3-year-olds
did perform comparably to the older children, addi-
tional work with alternative assessment techniques is
needed to elucidate the sources of age differences in
children's memory reports.

In conclusion, the finding that children were likely
to report spurious information under conditions that
were deliberately designed to enhance performance
is especially noteworthy. The interview protocol was
free of aggressively suggestive questioning. In fact,
the largest decrements in accuracy were seen at the
open-ended level of questioning, which by clinical
standards is considered to yield more trustworthy in-
formation than that obtained through specific ques-
tioning (Wehrspann, Steinhauser, & Klayner-Diamond,
Thus, information about complex events obtained from young children through behavioral enactment must be interpreted with caution. These findings are particularly relevant for clinicians who use behavioral reenactment techniques in forensic settings to evaluate allegations of sexual abuse in very young children. Indeed, the exploratory analyses of behavioral style suggest that the use of dolls and props may be particularly problematic in interviews of “easy” children. Similarly, a standard verbal interview may not be a productive way to interview young children who get bored easily. Accordingly, additional work is needed to examine the goodness-of-fit between aspects of the interview context and the personal characteristics of the children, so as to lead to the identification of optimal conditions under which to assess individual children.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the cooperation of the physicians, nurses, and office staff at Chapel Hill Pediatrics, P.A., and Durham Pediatrics, P.A., without whose interest and support this study would not have been possible. We also extend our gratitude to the children and parents who willingly participated in this research. Thanks are also due Laurie Austin, Patricia Clubb, Minda DeCubellis, Elizabeth Furtado, Paula Hollar, Aimee MacVean, Becky Rose, Lauren Shapiro, Gita Tlusty, Nancy Watt, and Jeff White for assistance in data collection, coding, and data entry; and Margaret Burchinal for her valuable comments regarding data analysis. In addition, we thank two anonymous reviewers for their constructive comments on earlier versions of this manuscript.

This research was supported in part by grants MH 49304 and HD 32214 from the United States Public Health Service. Preliminary reports of the data discussed here were presented at the Conference on Human Development, Pittsburgh, April 1994, and at the meeting of the Society for Applied Research in Memory and Cognition, Vancouver, BC, July 1995.

REFERENCES


Mandler, J. (1990). Recall and its verbal expression. In R. Fivush & J. A. Hudson (Eds.), Knowing and remembering...


