Journal of Experimental Psychology: Applied
2011, Vol. 17, No. 4, 406–417

1076-898X/11/$12.00 DOI: 10.1037/a0025129

Response Latency as a Predictor of the Accuracy of Children's Reports

Rakefet Ackerman
Technion–Israel Institute of Technology

Asher Koriat
University of Haifa

Researchers have explored various diagnostic cues to the accuracy of information provided by child eyewitnesses. Previous studies indicated that children's confidence in their reports predicts the relative accuracy of these reports, and that the confidence-accuracy relationship generally improves as children grow older. In this study, we examined the added contribution of response latency to the prediction of children's accuracy over and above that of confidence ratings. In Experiments 1 and 2, 2nd and 5th graders studied picture–event pairs and were tested using forced-choice, 2-alternative, or 5-alternative questions. In Experiment 3, children watched a slideshow depicting a story and were tested by 5-alternative questions about story details. The children indicated their confidence in each response, and response latency was measured. The results of all experiments suggested that children in both age groups relied on response latency as a cue for confidence, and this reliance contributed to the success with which they monitored the accuracy of their reports. When the test format was easy (Experiment 1), 2nd graders were as accurate as 5th graders in monitoring the accuracy of their answers, and the latency of their responses was no less predictive of accuracy. When the task was more difficult, age differences emerged. Nevertheless, in all experiments and for both age groups, response latency was found to have added value for predicting accuracy over and above that of confidence. Theoretical and practical implications of these findings for predicting the accuracy of children's reports are discussed.

Keywords: children as witnesses, response latency, confidence-accuracy relationship, metacognitive development

There has been a dramatic increase in recent years in the involvement of children in criminal events, both as witnesses and as victims of crimes (see Karmen, 2009; La Rooy, Lamb, & Memon, 2011; Shao & Ceci, 2011; Wiley, 2009). Consequently, legal professionals often have to rely on the testimony of young children, and the question of the credibility of children's testimony has been attracting a great deal of attention by researchers and policymakers (Bottoms, Najdowski, & Goodman, 2009).

There is a common belief that young children's ability to reliably remember and report about event details is inferior to that of older children and adults. This belief has been observed among actual or potential jurors, legal interviewers, attorneys, police detectives, and judges (Field et al., 2010; Kassin, Tubb, Hosch, & Memon, 2001; Melinder, Goodman, Eilertsen, & Magnussen, 2004; Myers, Redlich, Goodman, Prizmich, & Imwinkelried, 1999; Newcombe & Bransgrove, 2007). Overall, this belief has been supported by empirical findings (see Goodman & Melinder, 2007, for a review). Compared with young children, older children and adults have a more elaborate knowledge base by which to interpret events, employ advanced strategies for memorizing and retrieving information, and are more efficient in the strategic regulation of reporting information from memory (Bruck & Ceci, 2004; Bukatko & Daehler, 1998; Koriat, Goldsmith, Schneider, & Nakash-Dura, 2001; Roderer & Roebers, 2009; Schneider & Bjorklund, 2003).

Because of the limited ability of children to communicate spontaneously a detailed description of events, people charged with questioning children often find it necessary to employ direct questioning procedures that prompt children's memory for specific details. Such focused questioning, including leading questions and option-posing procedures, are generally used after exhausting open-ended questions (Hershkowitz, Fisher, Lamb, & Horowitz, 2007; Holliday & Albon, 2004). While extensive focused questioning generally tends to yield a larger number of correct details, it has been found to produce a larger number of erroneous details as well (Ceci, Kulkofsky, Klemfuss, Sweeney, & Bruck, 2007; Goodman & Quas, 2008). Consequently, special effort has been invested in identifying diagnostic markers that can distinguish between true and false reports.

The literature distinguishes between markers that are intended for differentiating between truth tellers and liars and markers that are meant to help determine whether truthful witnesses remember certain details correctly or incorrectly. Among the markers that have been examined in relation to lie detection are characteristics of the report content (e.g., logical structure, the extent of structured production, the presence of spontaneous corrections, and the inclusion of unusual details) and various behavioral cues (e.g., nodding and foot, leg, and hand movements) (Kulkofsky, 2008;

This article was published Online First August 15, 2011.

Rakefet Ackerman, Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology, Haifa, Israel; Asher Koriat, Department of Psychology, University of Haifa, Haifa, Israel.

We are grateful to the Chief Scientist of the Israeli Ministry of Education and Culture for approval of the project. This study was conducted at the Institute of Information Processing and Decision Making, University of Haifa, Haifa, Israel. It was supported by European Commission Grant FP6-NEST: EYEWITMEM; 43460.

We thank Tamar Zamet for her help in conducting the experiments.

Correspondence concerning this article should be addressed to Rakefet Ackerman, Faculty of Industrial Engineering and Management, Technion, Technion City, Haifa 32000, Israel. E-mail: ackerman@ie.technion.ac.il

Vrij, 2005; Vrij, Akehurst, Soukara, & Bull, 2004; Sporer & Schwandt, 2007). However, the validity of some of the markers commonly assumed to be diagnostic of accuracy (e.g., gaze aversion) has been questioned (e.g., DePaulo et al., 2003; see Sporer & Schwandt, 2007 for a review).

Importantly, techniques that have been effective in identifying lying have been shown to be less effective in distinguishing between innocent types of children's false reports and true reports (Kulkofsky, 2008). Two markers of report accuracy have attracted particular attention in the study of accuracy markers among truthful eyewitnesses: confidence and response latency (e.g., Brewer & Weber, 2008; Lindsay, Nilsen, & Read, 2000; Sauerland & Sporer, 2009). Confidence is a metacognitive judgment reflecting the respondent's subjective assessment of the likelihood that the answer provided or selected is correct. It is commonly assumed that confidence judgments can rest either on declarative knowledge and beliefs about memory, or on mnemonic cues derived in real time from task performance (Koriat & Levy-Sadot, 1999). An age-related increase has been observed in the use of both types of judgment. For example, in judging the likelihood that an object will be forgotten, 4-year-olds but not 3-year-olds show an awareness that information decays over time (Lyon & Flavell, 1993; see Bjorklund, Muir-Broaddus, & Schneider, 1990, for a review). As far as the use of mnemonic cues is concerned, Koriat, Ackerman, Lockl, and Schneider (2009a) found that unlike 1st and 2nd graders (6–8 years old), older children relied on the memorizing-effort heuristic in predicting their own recall: Their predictions tended to decrease with increased self-paced study time. Reliance on this heuristic was found to enhance the accuracy of children's recall predictions. Thus, research shows that the accuracy of metacognitive judgments tends to be relatively low for younger children, exhibiting a significant improvement around the age of eight or nine (Koriat & Ackerman, 2010a; Koriat & Shitzer-Reichert, 2002; Roebers & Howie, 2003; von der Linden & Roebers, 2006).

Despite these developmental trends, several studies using confidence ratings have indicated that even young children can monitor the accuracy of their reports to some extent, evidencing a positive confidence-accuracy relationship across different questions (Howie & Roebers, 2007; Roebers, 2002; Roebers & Howie, 2003; von der Linden & Roebers, 2006). This result suggests that the within-person correlation between confidence and accuracy may be useful in forensic contexts even for young children.

The second predictor of accuracy, response latency, has the advantage of being an objective measure that can be used regardless of the quality of children's metacognitive assessment of their own performance. Among adults, response latency has been found to be diagnostic of accuracy in within-person analyses, such that reports associated with shorter response latencies are more likely to be correct than those associated with longer latencies (e.g., Robinson, Johnson, & Herndon, 1997). Response latency was also found to be diagnostic to some extent in between-person analyses among adults, but with important caveats (Brewer, Caon, Todd, & Weber, 2006; Sauerland & Sporer, 2009).

A recent study by Koriat and Ackerman (2010a) examined the relationships between confidence, response latency, and the accuracy of answers provided by children in a semantic memory task. In their study, 2nd, 3rd, and 5th graders were presented with two-alternative general-knowledge questions (e.g., Where is the Eiffel Tower located? Paris; London). The children chose the answer to each question and indicated their confidence. Both confidence and choice latency predicted accuracy reliably for all three age groups, but the predictive value of each of the two indexes increased with age.

Whereas the previous study focused on general knowledge, the present study extended the investigation to an episodic memory task in which children learn new information and are later tested on their memory for that information. This procedure is more similar to the procedure used in the developmental studies mentioned above in which the memory for an event depends on encoding, retention, and retrieval. The experimental paradigm we used in the first two experiments follows the standard list-learning paradigm used in typical learning and memory experiments, but the tasks were deliberately chosen to incorporate some of the features that are common in eyewitness memory situations. In the third experiment we used a more natural story that was presented in the form of a narrated slide show. Response latency was measured as children answered memory questions, and confidence in the correctness of each answer was collected.

Examination of the joint contribution of latency and confidence to the prediction of accuracy raises a question that stems from the possible dependency of confidence on response latency. In their study, Koriat and Ackerman (2010a) found that the children's confidence judgments were inversely related to response latency. This relationship is consistent with previous results with adults (Kelley & Lindsay, 1993; Koriat, Ma'ayan, & Nussinson, 2006), which were taken to suggest reliance on the *choice-latency heuristic* in making confidence judgments. Indeed, Koriat and Ackerman found that when the contribution of response latency was partialed out from the confidence-accuracy correlation, the predictive validity of confidence was reduced. However, they also observed (Experiment 1) that this reduction was smaller for the younger children than for the older children. This finding suggests that in making confidence ratings, the younger children did not exploit the mnemonic cue of response latency to the same extent as did the older children. Presumably, reliance on response latency as a basis for confidence increases with age.

In the present study, we assess the predictive validity of response latency and confidence when each is considered separately. In addition, however, we also examine whether response latency makes an added contribution to the prediction of accuracy beyond that of confidence judgments. Because the younger children in Koriat and Ackerman's (2010a) study did not exploit response latency in full in making their confidence ratings, it may be expected that the added contribution of response latency to the prediction of accuracy should be particularly pronounced for this age group. Thus, we asked whether the variance in response latency that is not shared with confidence judgments makes a unique contribution to the prediction of accuracy, and whether there is a developmental change in this respect.

Experiment 1

Experiment 1 simulated a situation in which a child eyewitness is required to judge which of two previously seen people was involved in a certain event. Participants—2nd and 5th graders—were presented with pictures of children, each accompanied by a brief description of an event that allegedly happened to that child. The participants were asked to memorize the association between each child's picture and the associated event. Unlike the general-knowledge task used in Koriat and Ackerman (2010a), which yielded marked age differences that may reflect differences in semantic knowledge, the episodic sequence of encoding, retention, and retrieval used in this study potentially allows for similar performance among the younger and older children.

In the memory test phase, each event description was presented along with two pictures, and the children were asked to choose the picture associated with the described event. Response latency was measured, and the children indicated their confidence in their choice. Whereas previous studies of children's confidence in forensic contexts used categorical confidence judgment scales (e.g., Howie & Roebers, 2007; Roebers & Howie, 2003), we used a continuous scale that was adapted for children (Koriat & Ackerman, 2010a; Koriat et al., 2009a; Koriat, Ackerman, Lockl, & Schneider, 2009b).

The experimental procedure attempted to simulate an eyewitness situation in which several people were present at the scene of a crime and the goal is to clarify who was doing what. Because the distracter pictures had been associated in the study phase with other events, the children could not rely solely on familiarity to make the correct choice, as is the case in some line-up procedures (e.g., Brewer & Day, 2005; Kneller, Memon, & Stevenage, 2001). Thus, the results of the study may have some bearing on the possible effects of misleading familiarity, such as those that occur when an innocent bystander or a person previously seen in a mug shot is presented during the child's questioning (Dysart, Lindsay, Hammond, & Dupuis, 2001; Ross, Ceci, Dunning, & Toglia, 1994).

We adopt Brunswik's (1956) lens model framework to describe the interrelationships between response latency, confidence, and accuracy (see Koriat & Ackerman, 2010a; Koriat et al., 2006). The within-person gamma correlation between confidence and accuracy is an index of *achievement*, the extent to which the children were successful in monitoring the accuracy of their reports. The correlation between response latency and accuracy is an index of *cue validity*, reflecting the extent to which latency predicts accuracy. The correlation between latency and confidence, in turn, reflects *cue utilization,* the extent to which the children relied on response latency as a cue for confidence. We used the gamma correlation because of the dichotomous nature of the accuracy scores (see Nelson, 1984). For the sake of consistency, we also used gamma for the latency-confidence correlation.

Method

Participants. Participants were 20 2nd graders (M_{age} = 7.8, *SD* = 0.37) and 20 5th graders (M_{age} = 10.7, *SD* = 0.27) recruited from elementary schools in Israel. There were 15 boys and 25 girls. The children were recruited from schools in the city of Haifa and came from predominantly middle-class or upper-middle class neighborhoods.

Materials and apparatus. The items used were 48 picture-event pairs. The pictures, in color, showed boys and girls of different ages (about 6 to 16 years) and contained some visual background. Half the pictures depicted girls and half boys. The pictures were about 1.2 in. × 1.2 in. (3 cm × 3 cm) and were presented on a high-resolution (1024 × 768) LCD screen controlled by an IBM-PC compatible laptop computer. The descriptions, in Hebrew, described everyday events in the past tense, for instance "danced at the party" or "prepared a pizza," and were two to five words long. The pictures were paired randomly with the descriptions (with appropriate wording according to gender). All participants saw the same random pairings. Two additional picture-event pairs were used for practice.

For the construction of the memory test, the 48 pictures (and their associated descriptions) were divided into two sets of 24 pictures each. Half the children were tested on one set, and half on the other, so that each child was tested on 24 pairs of pictures—12 boy–boy pairs and 12 girl–girl pairs. In each test pair, one picture was the target and the other was the distracter, with the assignments reversed for half the children (that is, half the children encountered questions regarding pictures from one set, and half saw questions about pictures in the other.) The cue for each test item was the verbal description, and the child was asked which of the two pictures shown corresponded to that description (e.g., "Who danced at the party?"). The question was phrased in the masculine or the feminine in keeping with the gender of the children shown in the pairs (e.g., the Hebrew for *danced* is RAKAD in the masculine and RAKDA in the feminine).

Procedure. The consent of the parents and the school was obtained before beginning the study. The children volunteered to participate in the experiment, and they were explicitly told that they were allowed to withdraw at any stage. None of the children asked to withdraw during the session. Children were tested individually in a quiet school room during school hours. They were told that they would see a series of pictures of children on a computer screen, each accompanied by a short description of an event that happened to that child. The children were asked to pay close attention to the pictures and the accompanying descriptions and were told that during the memory test phase they would be presented with event descriptions and would be asked to identify which of two pictured children was associated with that event. The experimenter then illustrated and explained the full procedure using two practice items.

The children initiated each trial by using the mouse to click a box labeled *Display* (in Hebrew). The picture then appeared on the right side of the screen, and the event description appeared on the left. (Hebrew reads from right to left, so the picture occupied the position of the subject in a sentence.) To minimize the effects of reading skill, the experimenter read the description aloud, and the child was encouraged to pay attention to the picture. Each item was presented for 4 seconds, after which the *Display* box appeared again. The items were presented in random order.

When the study phase was completed, the children were given a 1-min filler task (making a free-line drawing). The test phase then followed. In the test phase, the child initiated the presentation of each question by clicking a box labeled *Display question*. The experimenter read each question aloud, and the child clicked a box labeled *Answers* to bring up the target and distracter pictures. The children were instructed to click the *Answers* box as soon as the experimenter finished reading and not to try to determine the answer before seeing the pictures. The pictures appeared on the screen side by side, and the children indicated their answer by clicking a rectangular bar beneath the chosen picture. The children were allowed to change their choice before clicking a *Continue* box. Choice latency was defined as the time interval between the *Answers* and *Continue* responses.

Confidence was measured as follows: After the *Continue* box was clicked, a red frame appeared around the chosen picture, along with the following question: *How sure are you that the answer is correct?* The measurement of confidence capitalized on the cold–hot game familiar to children by using a thermometer procedure (see Koriat & Ackerman, 2010a; Koriat et al., 2009a, a2009b). Children indicated their rating by sliding a pointer on a colored scale (in the shape of a horizontal thermometer) using the mouse. The pointer was initially positioned in the middle of the scale (colored white), and children could slide it toward one end of the scale (deep blue for "very cold") or toward the other end (deep red for "very hot"). The position on the scale was transformed into a confidence percentage score, 0 for very cold and 100 for very hot. After marking their confidence rating, the children clicked the *Next question* box to initiate the next trial.

The first question was about one of the two practice items, with the second picture used as the distracter. Test items were randomly ordered for each child. The participants received a small gift at the conclusion of the experiment.

Results and Discussion

The results of two 2nd graders were eliminated from the analyses because there was little variation in their confidence ratings (both indicated 100 as their confidence for all answers except one). The results for two 5th graders were eliminated as well, because they achieved 100% accuracy, which prevents the calculation of gamma correlations (see Nelson, 1984). Thus, the results reported below are based on 18 children in each age group.

The means for choice latency, confidence, and accuracy for each age group are presented in Table 1. Whereas choice latencies were longer for 2nd graders than for 5th graders, confidence and accuracy were similar for the two grades. The within-person standard deviation of response latency averaged 3.8 seconds for both age groups, indicating similar variability in response speed. The use of the confidence scale was also similar for the two age groups, as suggested by their similar mean confidence ratings, their similar levels of memory accuracy, and their similar standard deviation of confidence ratings (22.8), $t < 1$.

To foreshadow, except for the difference in choice latency, noted above, no other differences between the 2nd and 5th graders were found. However, we included grade as a factor in all the analyses to enable comparisons with the subsequent experiments.

In the following analyses, we first consider the separate contributions of confidence and response latency to the accuracy of the answers. We then examine the joint prediction of accuracy by latency and confidence, focusing on the unique, added contribution of response latency.

Confidence as a predictor of accuracy. Figure 1A presents the mean confidence-accuracy gamma correlation for each age group. The correlation averaged .74 (SD = .22) for 2nd graders and .77 (SD = .26) for 5th graders. Both correlations were significantly different from 0, $p < .0001$, and did not differ significantly from each other, $t(34) = 0.31$. These results testify to the ability of children, even 2nd graders, to distinguish correct from incorrect answers in episodic memory tests. This result differs from that reported by Koriat and Ackerman (2010a) for semantic memory. In that study, the confidence-accuracy correlation was measurably lower for 2nd graders (.48) than for 5th graders (.74).

To examine the predictive validity of confidence, we used a repeated-measures logistic regression (Proc GENMOD in SAS 9.1; Orelien, 2001) with confidence as a within-participant predictor and grade as a between-participants predictor. Confidence was a reliable predictor of accuracy, $\chi^2(1) = 103.61$, $p < .0001$, with no difference between the grades, $\chi^2(1) = 0.18$, *ns*, and no interactive effect, $\chi^2(1) = 0.44$, *ns*. The odds ratio (OR) of confidence was 1.04, indicating that every point of increase in confidence yielded a 4% increase in accuracy.

Response latency as a predictor of accuracy. Turning next to the predictive validity of latency, the gamma correlation between response latency and accuracy averaged −.55 (SD = .28) for 2nd graders and −.59 (SD = .29) for 5th graders, both significantly different from 0 ($p < .0001$), with no significant difference between them, $t(34) = 0.44$ (see Figure 1B). This result too differs from the significant developmental trend reported by Koriat and Ackerman (2010a) for a semantic memory task (−.24 for 2nd graders and −.65 for 5th graders).

As we did for confidence, we used a repeated-measures logistic regression to evaluate the cue validity of latency. Accuracy decreased significantly with increased response latency, $\chi^2(1) = 17.13$, $p < .0001$, with no difference between the grades, $\chi^2(1) =$

Table 1
Means and Standard Deviations (in Parentheses) of Choice Latency, Confidence, and Accuracy for Each Age Group in Each of the Three Experiments

Age group	Choice latency (s)	Confidence	Accuracy (%)
Experiment 1			
2nd graders	6.5 (1.7)	81.4 (9.9)	83.1[a] (8.4)
5th graders	5.4 (1.4)*	80.6 (9.8)	82.2[a] (11.9)
Experiment 2			
2nd graders	8.9 (2.2)	73.3 (16.0)	62.3[a] (18.3)
5th graders	10.4 (5.7)	75.9 (11.7)	73.5[a] (15.4)*
Experiment 3			
2nd graders	8.5 (2.1)	82.1 (10.0)	71.3[a] (9.5)
5th graders	5.4 (1.2)***	84.0 (8.7)	80.0[a] (6.1)***

[a] Accuracy is better than the chance level for the experiment, 50% for Experiment 1 and 20% for Experiments 2 and 3, $p < .0001$.
* $p < .05$. *** $p < .001$ for the difference between the grades.

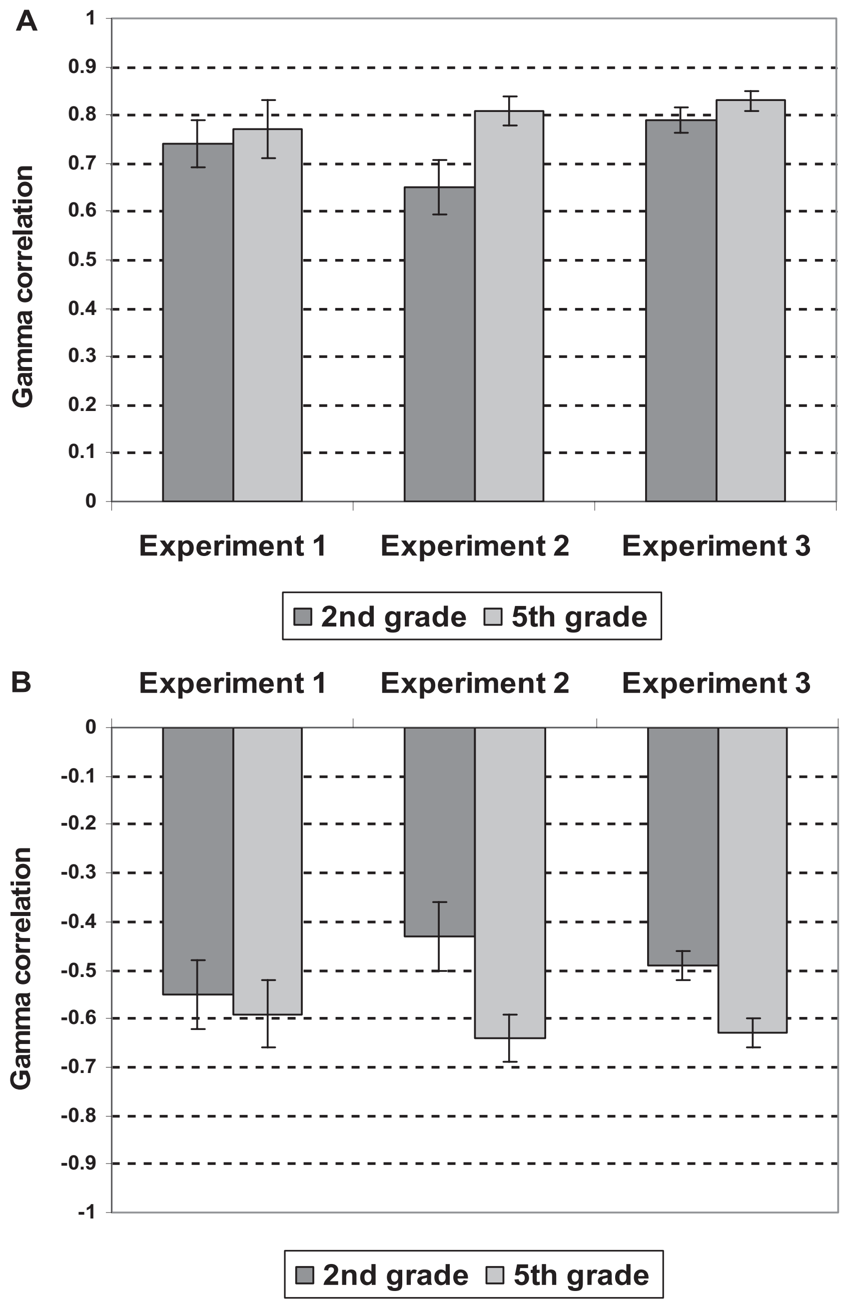

Figure 1. Means of within-participant gamma correlations between confidence and answer accuracy (panel A) and between choice latency and answer accuracy (panel B) for 2nd graders and for 5th graders across the three experiments. Error bars represent standard error of the means.

1.38, $p < .24$, and no interactive effect, $\chi^2(1) = 0.78$, *ns*. The odds ratio of latency was 0.88, indicating that every second of increase in latency reduced the likelihood an answer would be correct by 12%.

The latency-confidence relationship. As for cue utilization, the results suggest that children relied on response latency as a cue for confidence. The latency-confidence gamma correlation averaged $-.53$ ($SD = .18$) for 2nd graders and $-.50$ ($SD = .29$) for 5th graders, with no difference between them, $t(34) = 0.44$, while both significantly differed from 0 ($p < .0001$).

We examined the extent to which reliance on choice latency contributed to the predictive accuracy of confidence ratings. This examination required the use of within-participant Pearson correlations to compare the strength of the confidence-accuracy relationship before and after partialing out response latency (see Zakay & Tuvia, 1998). The raw Pearson correlation averaged .51 (SD = .24) across participants for 2nd graders and .50 ($SD = .22$) for 5th graders, both significant at $p < .0001$. When latency was partialed out, the respective correlations were lower but still significant, averaging .36 for each of the two age groups ($SD = .27$ for 2nd graders and .31 for 5th graders), $p < .0001$. A two-way Analysis of Variance (ANOVA) on these correlations, Correlation Type (raw vs. partialed out) × Grade, yielded a main effect for correlation type, $F(1, 34) = 13.88$, $MSE = 0.027$, $p = .001$, $\eta_p^2 = .29$, and no significant effects for grade or the interaction (both Fs < 1). These results suggest that monitoring accuracy derived partly, but not completely, from the children's reliance on choice latency as a cue for confidence. This finding is consistent with the results of Koriat and Ackerman (2010a), but in the task used in the present experiment the two age groups did not differ in this respect.

The added predictive validity of choice latency over that of confidence. Given that (a) children seem to rely on response latency as a cue for confidence and (b) this reliance contributes to the predictive accuracy of confidence, the question emerges whether response latency can still be used by external observers to improve the assessment of children's accuracy. We examined this question by comparing the log likelihood of the logistic regression model based on confidence alone (reported above) to that of a model that combines confidence and latency. The contribution of latency to the predictive validity of the regression model was significant, $\chi^2(2) = 8.89$, $p = .01$. When both choice latency and confidence were combined, the regression model yielded significant effects for both predictors $-\chi^2(1) = 8.26$, $p < .01$, OR = 0.93 for latency, and $\chi^2(1) = 87.83$, $p < .0001$, OR = 1.04 for confidence—and no interactive effect with grade, $\chi^2(1) = 0.89$, *ns*, and $\chi^2(1) = 0.30$, *ns*, respectively. Thus, taking into account latency in addition to confidence can improve assessments of accuracy in children's reports for both age groups.

The results of Experiment 1 generalize the findings for semantic memory (Koriat & Ackerman, 2010a) to a task in which children memorized person-event associations. First, the latency-confidence correlation was significant, suggesting that children in both age groups relied on response latency as a mnemonic cue for confidence. Second, however, choice latency made an independent contribution to children's memory accuracy.

Unlike Koriat and Ackerman (2010a), in the present study, 2nd graders demonstrated a pattern of results similar to that of the older children. It is of interest to examine whether the difference between the two studies is specific to the task used in Experiment 1 or reflects a more general difference between tasks that tap general knowledge and those that tap memory for recently studied information. Experiments 2 and 3 examined the possibility that the absence of age differences in Experiment 1 derived from the fact that the memory task used was relatively easy.

Experiment 2

Experiment 2 examined the generality of the findings using a more challenging memory test. Because our focus is on the processes that occur during testing rather than during learning, we used the same study conditions as in Experiment 1, which yielded a similar level of performance for both age groups, but modified the test format to include five alternatives rather than only two. This format change was expected to increase test difficulty by reducing the likelihood that a child would choose the correct answer by chance or by reliance on an elimination strategy that takes advantage of partial knowledge (see Budescu & Bar-Hillel, 1993).

Method

Participants. The participants were 20 2nd graders ($M_{age} = 8.1$, $SD = 0.57$) and 20 5th graders ($M_{age} = 11.0$, $SD = 0.26$), 12 boys and 28 girls, drawn from the same population as in Experiment 1.

Materials. The items were the picture-event pairs used for Experiment 1. Two items were added, resulting in a total of 50 experimental items and two practice items.

Procedure. The procedure for the study phase and the filler task were identical to those in Experiment 1. The procedure for the test phase was similar except that five same-gender pictures appeared side by side, in a five-alternative forced-choice format. The test phase comprised 25 items. Each set of 10 consecutive items included all 50 pictures (five pictures from the study phase for each question). Each set of 10 items was randomized for each participant. Thus, each studied picture appeared two or three times during the test phase, once as the correct answer and once or twice as a distracter. Half of the studied pictures were the correct answers for the questions.

Results and Discussion

There were no significant differences in latency, accuracy, or confidence between items in which the distracters appeared once in the memory test and those in which they appeared twice, and no interactive effects with age group, all ps > .20. Therefore, the results were pooled across all items.

In this experiment, no differences were found between 2nd graders and 5th graders in response latency or confidence. However, we observed an age difference in accuracy (see Table 1). We conducted two-way ANOVAs, Experiment (1 vs. 2) × Age Group, to compare the results of the two experiments. With regard to latency, the results indicated that it took longer to answer the five-alternative questions than the two-alternative questions, $F(1, 72) = 2.94$, $MSE = 11.18$, $p < .0001$, $\eta_p^2 = .24$, but the difference tended to be larger for the 5th graders than for the 2nd graders, $F(1, 72) = 2.94$, $MSE = 11.18$, $p = .09$, $\eta_p^2 = .04$ for the interaction. Accuracy was lower in Experiment 2 than in Experiment 1, $F(1, 72) = 20.40$, $MSE = 202.09$, $p < .0001$, $\eta_p^2 = .22$, and the interaction approached significance, $F(1, 72) = 3.48$, $MSE = 202.09$, $p < .07$, $\eta_p^2 = .05$, indicating that the younger group's performance suffered more from the use of five rather than two alternatives. However, although confidence was lower in Experiment 2 than in Experiment 1, $F(1, 72) = 5.22$, $MSE = 149.33$, $p < .05$, $\eta_p^2 = .07$, no interaction with age was found.

Confidence as a predictor of accuracy. Children in both age groups were accurate in monitoring the accuracy of their answers: The confidence-accuracy gamma correlation was significant for both age groups, $p < .0001$. However, the correlation was lower for the younger children ($M = .65$, $SD = .25$) than for the older children ($M = .81$, $SD = .15$), $t(38) = 2.59$, $p = .01$, Cohen's $d = 0.80$ (see Figure 1A).

The repeated-measures logistic regression yielded a significant effect for confidence, $\chi^2(1) = 40.46$, $p < .0001$, OR = 1.03 (3% increase in accuracy probability with a 1-point increase in confidence), with no effect for grade, $\chi^2(1) = 0.83$, *ns*. The interaction approached significance, $\chi^2(1) = 2.85$, $p = .09$, suggesting a trend in which confidence was a weaker predictor of accuracy for the younger than for the older children.

Response latency as a predictor of accuracy. The latency–accuracy gamma correlation revealed a significant age difference (see Figure 1B): The correlation was significantly weaker for the 2nd graders ($M = -.43$, $SD = .30$) than for the 5th graders ($M = -.64$, $SD = .22$), $t(38) = 2.59$, $p = .01$, $d = 0.82$, but the correlation was significantly different from 0 for both groups (both $p < .0001$). This result differs from what was observed in Experiment 1 but is in line with the finding of an age increase in the predictive validity of latency for general-knowledge questions (Koriat & Ackerman, 2010a). Note, however, that the latency-accuracy correlation for 2nd graders was stronger in the present experiment with five-alternative items than it was in that study with two-alternative answers (−.24).

The repeated-measures logistic regression indicated that latency was a reliable predictor of accuracy, $\chi^2(1) = 33.34$, $p < .0001$, OR = 0.91 (9% decrease in accuracy probability for each second increase in latency). A marginal difference was found between the grades, $\chi^2(1) = 3.45$, $p = .06$, OR = 2.13, with no interactive effect, $\chi^2(1) = 0.01$, *ns*.

The latency-confidence relationship. The inverse relationship between latency and confidence was replicated. However, the within-person latency-confidence gamma correlation was now weaker for 2nd graders ($M = -.44$, $SD = .20$) than for 5th graders ($M = -.62$, $SD = .16$), $t(38) = 3.12$, $p < .01$, $d = 1.02$, although both were significantly different from 0, $p < .0001$. This result also accords with the results reported by Koriat and Ackerman (2010a) for the general-knowledge task, suggesting an age increase in reliance on latency as a basis for confidence.

We examined the extent to which monitoring accuracy was mediated by choice latency by comparing Pearson correlations before and after partialing out the effects of latency, as was done earlier. The raw Pearson correlation averaged .45 ($SD = .17$) and .60 ($SD = .19$) for 2nd and 5th graders, respectively. As in Experiment 1, these correlations were lower after partialing out the effects of latency, .36 ($SD = .20$) and .41 ($SD = .28$), respectively, but both were significant at $p < .0001$. A two-way ANOVA on these correlations yielded a main effect for correlation type (raw vs. partialed out), $F(1, 38) = 31.47$, $MSE = 0.012$, $p < .0001$, $\eta_p^2 = .45$. The effect for grade approached significance, $F(1, 38) = 2.45$, $MSE = 0.081$, $p < .13$, $\eta_p^2 = .06$, and the interaction was significant, $F(1, 38) = 4.83$, $MSE = 0.012$, $p < .05$, $\eta_p^2 = .11$. Although the difference between the two correlations was significant for both age groups, it was smaller for 2nd graders, $t(19) = 2.58$, $p < .05$, $d = 0.50$, than for 5th graders, $t(19) = 5.21$, $p < .0001$, $d = 0.81$. These results are similar to the findings of Koriat and Ackerman (2010a) with regard to confidence and of Koriat et al. (2009a, 2009b) with regard to judgments of learning (JOLs), suggesting that improved monitoring accuracy with age is partly attributable to an age increase in the utilization of time (response time or study time) as a cue for metacognitive judgments.

The added predictive validity of choice latency over that of confidence. A regression analysis similar to that used in Experiment 1 indicated again a significant contribution of latency to accuracy over and above that of confidence, $\chi^2(2) = 25.37$, $p < .0001$. The combination of latency and confidence as predictors of accuracy yielded significant effects for both latency, $\chi^2(1) = 12.67$, $p < .001$, OR = 0.95, and confidence, $\chi^2(1) = 24.33$, $p < .0001$, OR = 1.03. However, somewhat unexpectedly, there were no interactive effects with grade, $\chi^2(1) = 0.45$, *ns* for latency, and $\chi^2(1) = 1.36$, *ns* for confidence.

Assuming that the learning process was similar in Experiments 1 and 2, the results of Experiment 2 highlight the importance of test characteristics in bringing to the fore age differences in the interrelationships between confidence, latency, and accuracy. Nevertheless, the validity of the two predictors was high for both age groups, even in Experiment 2. The results suggested that as in Experiment 1, choice latency can add to the prediction of accuracy over and above that provided by confidence. However, in contrast to our expectation, there was no difference in this respect between the two age groups. This was so despite the fact that there were age differences in the utilization of choice latency as a basis for confidence ratings.

Experiment 3

The purpose of Experiment 3 was to extend the findings to a situation that is closer still to a natural eyewitness scenario. In this experiment, a slideshow was used to present the sequence of events as a continuous story. Roebers, Gelhaar, and Schneider (2004) compared memory performance for 5- to 10-year-old children who watched a slideshow, a video, or a live show. They found that although the more natural watching conditions improved memory performance, children's confidence judgments and their accuracy were not influenced by the modality of presentation.

Unlike in Experiments 1 and 2, participants did not know in advance on what details of the slideshow they would be tested. After watching the slideshow, the participants answered five-alternative multiple-choice questions regarding various details of the story. Confidence ratings were elicited immediately after each answer, as in Experiments 1 and 2. This task resembles real-life situations in which, for example, several objects (e.g., a hammer, a golf club, a chair, a gun) were potentially present at the scene, and the child is asked which of these objects was used for a certain act (e.g., breaking a window).

Method

Participants. A total of 80 children from elementary schools in Israel participated in the study, including 40 2nd graders ($M_{age} = 7.7$, $SD = 0.42$) and 40 5th graders ($M_{age} = 10.7$, $SD = 0.32$). They were drawn from the same population as in the previous experiments and included 45 boys and 35 girls.

Materials. The study used a computerized slideshow lasting about 5 minutes, as in Koriat et al. (2001). The slideshow consisted of 27 color photographs, presented on a high-resolution (1024 × 768) LCD screen controlled by an IBM-PC compatible computer. The pictures were presented at a rate of 8 seconds per picture. Each picture was accompanied by a recorded narrative spoken in Hebrew by a professional radio broadcaster. The show, entitled "On the Way to the Picnic," depicted a staged incident in the life of a family. The story began with an introduction to each family member (the parents and three daughters). It then showed how, as the family was setting out on a picnic, the family cat climbed up an electricity supply pole and had to be enticed by various means to come down. A sequence of events then followed, culminating in the successful rescue of the cat.

A 30-item memory test was adapted from Koriat et al. (2001). The questions used in the present study were in a five-alternative recognition format and depicted either central aspects of the story or peripheral details.[1] The answer options were short, with a mean of 1.83 (*SD* = 1.05) words per answer. Two general knowledge questions and two questions on the details of the story were added for practice.

Procedure. The children were told that they would see a story presented on the computer screen in the form of a narrated slideshow, and that when the story was over, they would be asked to answer several questions about it. The children were instructed to pay close attention to the details of the spoken narrative as well as to the pictures. The procedures for answering the questions and giving confidence ratings were explained and demonstrated with the help of two general knowledge questions.

After the slideshow, the questions were displayed one at a time. To initiate each trial, the children clicked a box labeled *Display question* to reveal the question. After the experimenter read the question aloud, the children clicked a box labeled *Answers* to display five alternative answers in one column, labeled with the first five letters of the Hebrew alphabet. The children were instructed to click the *Answers* box as soon as the experimenter finished reading and not to think about the answer before seeing the alternatives. The first two questions were used as practice. In cases where the child had difficulty reading the answers for the two practice questions, the experimenter read the answers in addition to reading the questions throughout the test. The children chose an answer by clicking the letter next to it. The rest of the procedure was identical to that of Experiment 1. The 30 experimental questions appeared in the same order for all participants, following the sequence of events in the slideshow. Children went through the test phase at their own pace.

Results and Discussion

As can be seen in Table 1, 2nd graders took longer than 5th graders to choose their answers. As in Experiment 2, confidence ratings were similar for the two age groups, though 2nd graders achieved a lower accuracy rate than 5th graders. Despite the challenge of not knowing which items would be solicited at test, overall, the children achieved similar levels of accuracy and reported similar confidence levels to those found in the previous experiments (see Table 1). This finding puts the comparison between the two tasks on a common ground of perceived and actual knowledge levels.

Confidence as a predictor of accuracy. The children were accurate in discriminating between correct and wrong answers (Figure 1A), and the developmental improvement was quite small. The confidence-accuracy gamma correlation averaged .79 (*SD* = .16) for 2nd graders and .83 (*SD* = .14) for 5th graders. It was significantly different from 0 for both groups, $p < .0001$, and the difference between them was not significant, $t(78) = 1.24$, $p = .22$, $d = 0.27$. The logistic regression model, as before, yielded a significant effect for confidence, $\chi^2(1) = 244.92$, $p < .0001$, OR = 1.05 (5% increase in accuracy with every increase of one point in confidence), with no effect of grade, $\chi^2(1) = 0.03$, *ns*, and no significant interaction, $\chi^2(1) = 1.97$, $p = .16$.

Response latency as a predictor of accuracy. As can be seen in Figure 1B, the latency-accuracy gamma correlation was weaker for 2nd graders ($M = -.49$, $SD = .21$) than for 5th graders ($M = -.63$, $SD = 21$), $t(78) = 2.94$, $p < .01$, $d = 0.68$. The correlation was significant for both groups, $p < .0001$. The logistic regression model yielded a significant prediction of accuracy by latency, $\chi^2(1) = 63.75$, $p < .0001$, OR = 0.84, a significant prediction by grade, $\chi^2(1) = 10.30$, $p = .001$, OR = 2.50, and also an interactive effect, $\chi^2(1) = 8.61$, $p < .01$. With regard to age differences, the regression models indicated a significant prediction for both age groups, but the prediction was weaker for 2nd graders, $\chi^2(1) = 20.88$, $p < .0001$, OR = 0.90 (10% decrease in accuracy per second), than for 5th graders, $\chi^2(1) = 42.90$, $p < .0001$, OR = 0.79 (21% decrease in accuracy probability for each extra second). Thus, latency was predictive of accuracy for both age groups, but it was less predictive for 2nd graders.

The latency-confidence relationship. The mean gamma correlation between choice latency and confidence was weaker for 2nd graders ($M = -.44$, $SD = .21$) than for 5th graders ($M = -.57$, $SD = .18$), $t(78) = 2.90$, $p < .01$, $d = 0.67$. Both correlations differed significantly from 0, $p < .0001$. This pattern replicates the age differences that were found in Experiment 2.

We again examined the extent to which the confidence–accuracy correlation was mediated by reliance on response latency, using the same procedure as in the previous experiments. The raw Pearson correlation averaged .55 (*SD* = .17) for 2nd graders and .61 (*SD* = .16) for 5th graders. When latency was partialed out, the respective correlations were lower but still significant, averaging .46 (*SD* = .24) and .47 (*SD* = .24). A two-way ANOVA on these correlations yielded a main effect for correlation type, $F(1, 78) = 49.56$, $MSE = 0.01$, $p < .0001$, $\eta_p^2 = .39$, with no difference between the grades ($F < 1$) and no interaction, $F(1, 78) = 2.42$, $MSE = 0.01$, $p = .12$, $\eta_p^2 = .03$. Overall, the results suggest that in both age groups monitoring accuracy derived in part from reliance on choice latency as a cue for confidence.

The added predictive validity of choice latency over that of confidence. Comparing the regression models with and without latency according to log likelihood ratio again yielded a significant contribution of latency over that of confidence, $\chi^2(2) = 48.44$, $p < .0001$. The combined model of latency and confidence as predictors of accuracy yielded significant effects for both factors, $\chi^2(1) = 14.80$, $p = .0001$, OR = 0.92 for latency, and $\chi^2(1) = 195.63$, $p < .0001$, OR = 1.04 for confidence, with no interactive effects with grade, $\chi^2(1) = 1.30$, *ns*, and $\chi^2(1) = 0.68$, *ns*, for latency and confidence, respectively. Thus, latency allows better assessment of children's accuracy when used in addition to confidence ratings.

Experiment 3 generalized the conclusions of Experiment 2 to a task that is more similar to eyewitness testimony (Roebers et al., 2004). In the scenario used in Experiment 3, confidence was equally valid as a predictor of accuracy for both age groups. The results again support the finding that choice latency is a valid cue for report accuracy even for 2nd graders. The results are consistent with Experiments 1 and 2, indicating that latency contributes to the

[1] The questions were classified by two judges as involving central details tied to the main story (e.g., the cat climbed up an electricity pole) and more peripheral information (e.g., the color of the mother's shirt). There was high interjudge reliability, and disagreements were settled after discussion, resulting in 11 central items and 19 peripheral items. The results did not differ for the two types of items. Therefore only the results across all items are reported.

prediction of accuracy over and above confidence, and the added contribution does not differ for the two age groups.

General Discussion

In this study, we focused on situations in which children honestly report several pieces of information. For such situations, we examined two cues that can discriminate between correct and false reports: The latency of choosing an answer and the confidence in that answer. In previous research, confidence ratings were found to be predictive of accuracy of the report even for young children (Howie & Roebers, 2007; Roebers, 2002; Roebers & Howie, 2003; von der Linden & Roebers, 2006). The present study focused on response latency as a second predictor and examined its predictive validity both when considered separately and when considered in conjunction with confidence ratings.

The memory tasks we used in Experiments 1 and 2 required children to decide which of two pictured children had been associated with a particular event. Increasing the number of alternatives from two (Experiment 1) to five (Experiment 2) increased response latency, reduced confidence in the answers, and decreased overall report accuracy, suggesting that it increased the difficulty of the task. The slideshow used in Experiment 3 was richer in content and was meant to make the scenario more similar to that of an eyewitness situation in which the witness does not know beforehand which items he or she will be asked to report on during an interview. Whereas Experiment 1 yielded no age difference in the validity of either confidence or latency as predictors of accuracy, Experiment 2 revealed an age increase in the confidence–accuracy relationship, and both Experiments 2 and 3 yielded an age increase in the latency–accuracy relationship. Regardless of these differences, however, the results overall suggest that under all the studied conditions both confidence and response latency reliably discriminate between correct and false memory reports, and this is so even for 2nd graders.

In addition to evaluating the separate predictive validity of confidence and latency, we asked whether response latency makes an added contribution to the prediction of accuracy over and above that of confidence ratings. This question is important in view of findings suggesting that participants rely on response latency as a basis for their confidence (Kelley & Lindsay, 1993; Koriat et al., 2006). Because our previous results with general-knowledge questions suggested that children, particularly the youngest group, did not exploit choice latency in full in making confidence ratings (Koriat & Ackerman, 2010a), we expected the added contribution of response latency to be especially pronounced for the younger children. The analyses of the interrelationships between latency, confidence, and accuracy, however, did not support this possibility for any of the tasks used in this study. Rather, in all three experiments, response latency was found to make a reliable added contribution to the prediction of accuracy over and above that of confidence ratings, and the extent of this contribution did not differ for the two age groups. The implication of these findings is that external observer can benefit from considering children's response latency as a cue for report accuracy even when confidence ratings are available.

In what follows, we first discuss some of the theoretical issues raised by the results and then mention some of their potential applications. The present study contributes to a theoretical understanding of the development of metacognitive monitoring. It joins a recent line of research suggesting that children, like adults, base their metacognitive judgments on the feedback from task performance (Hoffmann-Biencourt, Lockl, Schneider, Ackerman, & Koriat, 2010; Koriat & Ackerman, 2010a; Koriat et al., 2006; Koriat et al., 2009a, 2009b). Mnemonic cues, such as processing fluency, accessibility, and experienced effort, which reflect the feedback from task performance, are assumed to give rise directly to metacognitive feelings (e.g., feeling of knowing, judgment of learning, and confidence) through a process that operates below full consciousness (Kelley & Lindsay, 1993; Koriat & Ackerman, 2010b; Koriat & Levy-Sadot, 1999; Thomas, Bulevich, & Chan, 2010). In previous studies, it was found that as children grow older they rely more heavily on these mnemonic cues as a basis for their metacognitive judgments. The present study did not yield consistent support for this age increase; only in Experiment 2 was there increased reliance with age on response latency as a basis of confidence. The discrepancy between the results of the present study and those reported by Koriat and Ackerman (2010a) may reflect a difference in the underlying memory representations. Clearly, in the general-knowledge task of Koriat and Ackerman (2010a), children draw on a long-term database, which changes with age, whereas in the picture-event task used in this study, they rely on newly learned, episodic associations. Perhaps this is why the younger group achieved much better monitoring accuracy for the episodic task than when faced with a semantic task. Further research comparing metacognitive monitoring for semantic and episodic tasks may shed light on the processes underlying the development of metacognitive skills.

An important theoretical issue raised by the results of the present study concerns the mechanism underlying the validity of response latency as a cue for accuracy. According to the self-consistency model of subjective confidence (Koriat, 2011a, 2011b), both confidence and response latency are based on the amount of deliberation and conflict a respondent experiences in attempting to choose the correct answer. Response latency would seem to reflect how difficult it is to reach a choice, and hence the likelihood that the choice is correct (Wright & Ayton, 1988). However, it would seem that the response latency of older children is more tuned to the intrinsic difficulty of the task than that of younger children. For example, 5th graders invested twice as much time in the five-alternative task of Experiment 2 than in the two-alternative task of Experiment 1, whereas the increase was smaller for the 2nd graders (see Table 1). This pattern parallels the pattern obtained for self-paced study time (e.g., Lockl & Schneider, 2002, 2004; Masur, McIntyre, & Flavell, 1973). For example, Dufresne and Kobasigawa (1989) asked children to study either easy or hard paired-associates until they were sure they could remember all pairs perfectly. They found that 5th and 7th graders spent more time studying the hard pairs than the easy pairs, whereas 1st and 3rd graders spent roughly the same amount of time on both types of pairs. It should be noted that both confidence and accuracy decrease with response latency only when response latency is data-driven, tuned to the processing difficulty of the item in a bottom-up fashion. In contrast, when differences in response latency reflect goal-driven, top-down variations, both confidence and accuracy *increase* with the amount of time invested (Koriat et al., 2006). Thus, if the child's motivation to be accurate differs for different questions, this should compromise the generally negative correlation between response time on the one hand, and confidence and accuracy on the other.

We turn next to some potential applications of these results. As noted in the introduction, a great deal of effort has been invested in identifying potential markers for the accuracy of memory reports. Much of that effort has been carried out in connection with deception and lie detection. Both behavioral cues (e.g., signs of nervousness) and verbal cues (e.g., less detailed reports) have been explored (see Porter & ten Brinke, 2010; Sporer & Schwandt, 2007; Vrij, Granhag, & Porter, 2010). The advantage of confidence ratings and response latency is that both represent continuous measures that can be obtained for each question. Response latency can be easily measured in computerized environments for conducting eyewitness identification, particularly in connection with sequential lineups, as is done, for example, by the software for computer-based lineup administration suggested by MacLin, Zimmerman, and Malpass (2005) and those used for the study of reliability of eyewitness identification (e.g., Brewer et al., 2006). In the present study, we examined the validity of response latency only for option-posing memory questions. For questioning procedures that use such memory questions, the results suggest that response speed can have some value in monitoring the accuracy of young children's reports. However, previous results with adults suggest that response speed is also diagnostic of accuracy for open questions. In fact, Robinson et al. (1997) found response latency to be more diagnostic of accuracy for recall than for recognition testing. In view of the recommendation to prefer open-ended questions in questioning procedures (e.g., Hershkowitz et al., 2007), it is important to confirm that the results of the present study regarding the validity of response speed also hold true for open-ended questions.

From a practical point of view, several researchers have discussed factors that should be emphasized in training professional interviewers (see Fisher, 2010; Goodman & Melinder, 2007). In connection with these discussions, it would be important to determine whether interviewers can be trained to use their own spontaneous assessment of response speed in assessing the accuracy of children's reports. Previous results suggest that training might be beneficial. For example, in their study of judgments of learning, Koriat and Bjork (2006) found that learners were able to overcome some of the biases inherent in monitoring their competence after receiving a description of the factors that lead to these biases. That such training might be beneficial is also suggested by the results of a recent study (Koriat & Ackerman, 2010b) in which observers watched a video depicting a person engaged in self-paced study of paired associates who spent different amounts of time in the study of different pairs. The observers, who were required to predict the likelihood that the target person would recall the pair at test, failed to use study time as a cue for recall predictions. However, they were successful in using study time as a cue for recall predictions (under the heuristic that recall should *decrease* with increased study time) after having the opportunity to study paired-associates themselves under self-paced conditions. This line of research (see also Undorf & Erdfelder, in press) suggests that practice and training may also enhance interviewers' sensitivity to response speed as a cue for report accuracy.

The present study, however, has several limitations. First, the tasks used in this study were relatively benign, with no stressful events or direct involvement of the child in the scene, and there were no misleading questions and no anticipation of lying. It is unclear how these factors affect the validity of response speed as a cue for accuracy. Second, we used a short retention interval in our experiments. In many real-life situations, in contrast, eyewitnesses are questioned about an event after longer retention intervals. The validity of response time may change with increased retention interval (Brewer et al., 2006). We suspect that response speed is sensitive to the differential forgetting of different pieces of information, so that its validity might actually increase with retention interval, but this speculation needs investigation. Finally, response latency is clearly sensitive to task difficulty, but it is unclear how the effects of task difficulty might interact with age. We proposed that the age differences observed between Experiments 1 and 2 derive primarily from differences in the difficulty of the memory tests used. To substantiate this proposition, however, it is important to examine whether a similar pattern is found when difficulty is manipulated by other variables, such as speed of exposure to the events during the study phase, viewing conditions, number of stimuli, and so on.

In sum, unlike many studies that have focused on the fallibility of eyewitness memory, particularly among young children, the present study joins with other attempts (see Brewer & Weber, 2008) to define conditions that promote reliable eyewitness reporting by identifying useful markers for memory accuracy. The results indicate that both confidence and response latency are reliable markers of accuracy among young children, and that consideration of response latency can improve discrimination between correct and false memory reports even when the contribution of confidence ratings is taken into account.

References

Bjorklund, D. F., Muir-Broaddus, J. E., & Schneider, W. (1990). The role of knowledge in the development of strategies. In D. F. Bjorklund (Ed.), *Children's strategies: Contemporary views of cognitive development* (pp. 93–128). Hillsdale, NJ: Erlbaum.

Bottoms, B. L., Najdowski, C. J., & Goodman, G. S. (2009). *Children as victims, witnesses, and offenders: Psychological science and the law.* New York, NY: The Guilford Press.

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior, 30,* 31–50. doi:10.1007/s10979-006-9002-7

Brewer, N., & Day, K. (2005). The confidence-accuracy and decision latency-accuracy relationships in children's eyewitness identification. *Psychiatry, Psychology and Law, 12,* 119–128. doi:10.1375/pplt.2005.12.1.119

Brewer, N., & Weber, N. (2008). Eyewitness confidence and latency: Indices of memory processes not just markers of accuracy. *Applied Cognitive Psychology, 22,* 827–840. doi:10.1002/acp.1486

Bruck, M., & Ceci, S. (2004). Forensic developmental psychology. *Current Directions in Psychological Science, 13,* 229–232. doi:10.1111/j.0963-7214.2004.00314.x

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* Berkeley, CA: University of California Press.

Budescu, D., & Bar-Hillel, M. (1993). To guess or not to guess: A decision-theoretic view of formula scoring. *Journal of Educational Measurement, 30,* 277–291. doi:10.1111/j.1745-3984.1993.tb00427.x

Bukatko, D., & Daehler, M. W. (1998). *Child development: A thematic approach* (3rd ed.). Boston, MA: Houghton Mifflin Company.

Ceci, S. J., Kulkofsky, S., Klemfuss, J. Z., Sweeney, C. D., & Bruck, M. (2007). Unwarranted assumptions about children's testimonial accuracy.

Annual Review of Clinical Psychology, 3, 311–328. doi:10.1146/annurev.clinpsy.3.022806.091354

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129,* 74–118. doi:10.1037/0033-2909.129.1.74

Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology, 47,* 274–296. doi:10.1016/0022-0965(89)90033-7

Dysart, J. E., Lindsay, R. C. L., Hammond, R., & Dupuis, P. (2001). Mug shot exposure prior to lineup identification: Interference, transference, and commitment effects. *Journal of Applied Psychology, 86,* 1280–1284. doi:10.1037/0021-9010.86.6.1280

Field, T., Malphurs, J. E., Yando, R., Bendell, D., Carraway, K., & Cohen, R. (2010). Legal interviewers use children's affect and eye contact cues to assess credibility of their testimony. *Early Child Development and Care, 180,* 397–404. doi:10.1080/03004430801996355

Fisher, R. P. (2010). Interviewing cooperative witnesses. *Legal and Criminological Psychology, 15,* 25–38. doi:10.1348/135532509X441891

Goodman, G. S., & Melinder, A. (2007). Child witness research and forensic interviews of young children: A review. *Legal and Criminological Psychology, 12,* 1–19. doi:10.1348/135532506X156620

Goodman, G. S., & Quas, J. A. (2008). Repeated Interviews and Children's Memory. *Current Directions in Psychological Science, 17,* 386–390. doi:10.1111/j.1467-8721.2008.00611.x

Hershkowitz, I., Fisher, S., Lamb, M. E., & Horowitz, D. (2007). Improving credibility assessment in child sexual abuse allegations: The role of the NICHD investigative interview protocol. *Child Abuse & Neglect, 31,* 99–110. doi:10.1016/j.chiabu.2006.09.005

Hoffmann-Biencourt, A., Lockl, K., Schneider, W., Ackerman, R., & Koriat, A. (2010). Self-paced study time as a cue for recall predictions across school age. *British Journal of Developmental Psychology, 28,* 767–784. doi:10.1348/026151009X479042

Holliday, R. E., & Albon, A. J. (2004). Minimising misinformation effects in young children with cognitive interview mnemonics. *Applied Cognitive Psychology, 18,* 263–281. doi:10.1002/acp.973

Howie, P., & Roebers, C. M. (2007). Developmental progression in the confidence-accuracy relationship in event recall: Insights provided by a calibration perspective. *Applied Cognitive Psychology, 21,* 871–893. doi:10.1002/acp.1302

Karmen, A. (2009). *Crime victims: An introduction to victimology* (7th ed.). Belmont, CA: Wadsworth.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: A new survey of the experts. *American Psychologist, 56,* 405–416. doi:10.1037/0003-066X.56.5.405

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32,* 1–24. doi:10.1006/jmla.1993.1001

Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential line-ups: Decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology, 15,* 659–671. doi:10.1002/acp.739

Koriat, A. (2011a). Subjective confidence in perceptual judgments: A test of the self-consistency model. *Journal of Experimental Psychology: General, 140,* 117–139. doi:10.1037/a0022171

Koriat, A. (2011b). The self-consistency model of subjective confidence. *Psychological Review* (under revision).

Koriat, A., & Ackerman, R. (2010a). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science, 13,* 441–453. doi:10.1111/j.1467-7687.2009.00907.x

Koriat, A., & Ackerman, R. (2010b). Metacognition and mindreading: Judgments of learning for Self and other during self-paced study. *Consciousness and Cognition, 19,* 251–264. doi:10.1016/j.concog.2009.12.010

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009a). The easily learned, easily-remembered heuristic in children. *Cognitive Development, 24,* 169–182. doi:10.1016/j.cogdev.2009.01.001

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009b). The memorizing effort heuristic in judgments of learning: A developmental perspective. *Journal of Experimental Child Psychology, 102,* 265–279. doi:10.1016/j.jecp.2008.10.005

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1133–1145. doi:10.1037/0278-7393.32.5.1133

Koriat, A., Goldsmith, M., Schneider, W., & Nakash-Dura, M. (2001). The credibility of children's testimony: Can children control the accuracy of their memory reports? *Journal of Experimental Child Psychology, 79,* 405–437. doi:10.1006/jecp.2000.2612

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 483–502). New York, NY: Guilford Press.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135,* 36–69. doi:10.1037/0096-3445.135.1.36

Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy. In P. Chambres, M. Izaute, & P.-J. Marescaux (Eds.), *Metacognition: Process, function and use* (Vol. xvi, pp. 1–17). Dordrecht, The Netherlands: Kluwer Academic Publishers. doi:10.1007/978-1-4615-1099-4_1

Kulkofsky, S. (2008). Credible but Inaccurate: Can Criterion-based Content Analysis (CBCA) Distinguish True and False Memories? In M. J. Smith (Ed.), *Child sexual abuse: Issues and challenges* (pp. 21–42). New York, NY: Nova Science Publishers.

La Rooy, D., Lamb, M. E., & Memon, A. (2011). Forensic interviews with children in Scotland: A survey of interview practices among police. *Journal of Police and Criminal Psychology, 26,* 26–34. doi:10.1007/s11896-010-9072-9

Lindsay, D. S., Nilsen, E., & Read, J. D. (2000). Witnessing-condition heterogeneity and witnesses' versus investigators' confidence in the accuracy of witnesses' identification decisions. *Law and Human Behavior, 24,* 685–697. doi:10.1023/A:1005504320565

Lockl, K., & Schneider, W. (2002). Zur Entwicklung des selbstregulierten Lernens im Grundschulalter: Zusammenhänge zwischen Aufgabenschwierigkeit und Lernzeiteinteilung [The development of self-regulated learning in elementary school children: Associations between task difficulty and allocation of study time]. *Psychologie in Erziehung und Unterricht, 49,* 3–16.

Lockl, K., & Schneider, W. (2004). The effects of incentives and instructions on children's allocation of study time. *European Journal of Developmental Psychology, 1,* 153–169. doi:10.1080/17405620444000085

Lyon, T. D., & Flavell, J. H. (1993). Young children's understanding of forgetting over time. *Child Development, 64,* 789–800. doi:10.2307/1131218

MacLin, O. H., Zimmerman, L. A., & Malpass, R. S. (2005). PC_Eyewitness and the sequential superiority effect: Computer-based lineup administration. *Law and Human Behavior, 29,* 303–321. doi:10.1007/s10979-005-3319-5

Masur, E. F., McIntyre, C. W., & Flavell, J. H. (1973). Developmental changes in apportionment of study time among items in a multitrial free recall task. *Journal of Experimental Child Psychology, 15,* 237–246. doi:10.1016/0022-0965(73)90145-8

Melinder, A., Goodman, G. S., Eilertsen, D. E., & Magnussen, S. (2004).

Beliefs about child witnesses: A survey of professionals. *Psychology, Crime & Law, 10,* 347–365. doi:10.1080/10683160310001618717

Myers, J. E. B., Redlich, A. D., Goodman, G. S., Prizmich, L. P., & Imwinkelried, E. (1999). Jurors' perceptions of hearsay in child sexual abuse cases. *Psychology, Public Policy, and Law, 5,* 388–419. doi:10.1037/1076-8971.5.2.388

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109–133. doi:10.1037/0033-2909.95.1.109

Newcombe, P. A., & Bransgrove, J. (2007). Perceptions of witness credibility: Variations across age. *Journal of Applied Developmental Psychology, 28,* 318–331. doi:10.1016/j.appdev.2007.04.003

Orelien, J. G. (2001). Model fitting in PROC GENMOD. *Proceedings of the 26th Annual SAS Users Group International Conference.* Cary, NC: SAS Institute; 2001.

Porter, S., & ten Brinke, L. (2010). The truth about lies: What works in detecting high-stakes deception? *Legal and Criminological Psychology, 15,* 57–75. doi:10.1348/135532509X433151

Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology, 82,* 416–425. doi:10.1037/0021-9010.82.3.416

Roderer, T., & Roebers, C. M. (2009). Children's strategic regulation of memory accuracy. In M. R. Kelley (Ed.), *Applied memory* (pp. 253–274). Hauppauge, NY: Nova Science Publishers.

Roebers, C. M. (2002). Confidence judgments in children's and adults' event recall and suggestibility. *Developmental Psychology, 38,* 1052–1067. doi:10.1037/0012-1649.38.6.1052

Roebers, C. M., Gelhaar, T., & Schneider, W. (2004). "It's magic!" The effects of presentation modality on children's event memory, suggestibility, and metacognition. *Journal of Experimental Child Psychology, 87,* 320–335. doi:10.1016/j.jecp.2004.01.004

Roebers, C. M., & Howie, P. (2003). Confidence judgments in event recall: Developmental progression in the impact of question format. *Journal of Experimental Child Psychology, 85,* 352–371. doi:10.1016/S0022-0965(03)00076-6

Ross, D. F., Ceci, S. J., Dunning, D., & Toglia, M. P. (1994). Unconscious transference and mistaken identity: When a witness misidentifies a familiar but innocent person. *Journal of Applied Psychology, 79,* 918–930. doi:10.1037/0021-9010.79.6.918

Sauerland, M., & Sporer, S. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied, 15,* 46–62. doi:10.1037/a0014560

Schneider, W., & Bjorklund, D. F. (2003). Memory and knowledge development. In J. Valsiner (Ed.), *Handbook of developmental psychology* (pp. 370–402). London, UK: Sage.

Shao, Y., & Ceci, S. J. (2011). Adult credibility assessments of misinformed, deceptive and truthful children. *Applied Cognitive Psychology, 25,* 135–145. doi:10.1002/acp.1652

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law, 13,* 1–34. doi:10.1037/1076-8971.13.1.1

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

Undorf, M., & Erdfelder, E. (in press). Judgments of learning reflect encoding fluency: Conclusive evidence for the ease-of-processing hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Von der Linden, N., & Roebers, C. M. (2006). Developmental changes in uncertainty monitoring during an event recall task. *Metacognition and Learning, 1,* 213–228. doi:10.1007/s11409-006-9001-6

Vrij, A. (2005). Criteria-based content analysis. A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law, 11,* 3–41. doi:10.1037/1076-8971.11.1.3

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research, 30,* 8–41. doi:10.1111/j.1468-2958.2004.tb00723.x

Vrij, A., Granhag, P. A., & Porter, S. B. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest, 11,* 89–121. doi:10.1177/1529100610390861

Wiley, T. R. A. (2009). Legal and social service responses to child sexual abuse: A primer and discussion of relevant research. *Journal of Child Sexual Abuse, 18,* 267–289. doi:10.1080/10538710902901754

Wright, G., & Ayton, P. (1988). Decision time, subjective probability, and task difficulty. *Memory & Cognition, 16,* 176–185. doi:10.3758/BF03213487

Zakay, D., & Tuvia, R. (1998). Choice latency times as determinants of post-decisional confidence. *Acta Psychologica, 98,* 103–115. doi:10.1016/S0001-6918(97)00037-1

Received July 18, 2010
Revision received May 31, 2011
Accepted July 18, 2011 ■