EMPIRICAL STUDY

Developmental changes in uncertainty monitoring during an event recall task

Nicole von der Linden · Claudia M. Roebers

Abstract The present study was conducted to examine developmental progression in children's metacognitive monitoring competencies in the context of a complex memory task. 7- and 9-year-olds rated their confidence after answering questions in two different question formats (unbiased and misleading) and two different question types (answerable and unanswerable). Feeling-of-knowing judgments were gathered for questions that had previously been answered with "don't know." The results showed that children from both age groups appropriately differentiated between correct and incorrect answers to unbiased questions in their confidence judgments, between answerable and unanswerable questions, and appropriately showed lower confidence levels in their confidence judgments than in their feeling-of-knowing judgments. 9-year-olds proved to be further able to discriminate metacognitively between correct and incorrect answers to misleading answerable questions in their confidence judgments while 7-year-olds were not. The comparison of feeling-of-knowing judgments before correct and incorrect recognition indicated that metacognitive differentiation at the lower end of the uncertainty–certainty continuum posed problems for these age groups. The observation of an adult confederate modeling appropriate metamemory monitoring did not improve children's metacognitive performance.

Assessment of subjective confidence plays an important role in many intentional and non-intentional cognitive tasks. For example, in structured learning situations, students' assessment of how well they already know the material helps them decide whether they

N. von der Linden (✉)
Lehrstuhl für Psychologie IV, Institut für Psychologie, Universität Würzburg,
Röntgenring 10, 97070 Würzburg, Germany
e-mail: linden@psychologie.uni-wuerzburg.de

C. M. Roebers
University of Berne, Berne, Switzerland

need to continue studying or not (C. H. Renner & M. J. Renner, 2001). In everyday memory tasks where accuracy of memory is crucial (departure times, shopping lists, academic tests, or eyewitness situations), appropriate monitoring of potential answers is the most important factor when deciding whether to withhold or to provide information (Koriat & Goldsmith, 1996; 1998).

Nelson and Narens (1990, 1994) provided a differentiated theoretical framework of procedural metacognitive processes. They distinguish between monitoring processes (ease-of-learning judgments, judgments-of-learning, feeling-of-knowing judgments, and confidence judgments) and control processes (allocation of study time, termination of search, "don't know" answers). While monitoring indicators are commonly assessed in the context of intentional learning, confidence judgments and feeling-of-knowing judgments are also relevant in non-intentional memory tasks and in many everyday memory tasks.

In recent years, developmental research has provided answers for some decisive questions on the development of procedural metacognitive processes (Schneider, 1998). Findings show that children as young as 8 years of age possess good metacognitive monitoring abilities (e.g. judgments-of-learning, feeling-of-knowing judgments) in the context of intentional self-directed learning activities. Yet, there is still need of further research concerning developmental progression in metacognitive processes in the context of complex, non-intentional, everyday memory tasks. This is because appropriate metacognitive monitoring and control processes are causally related to memory performances, independent of age and metacognitive indicator (DeMarie & Ferron, 2003; Pierce & Lange, 2000; Schneider, Schlagmüller, & Visé, 1998). As the effects of metacognition are comparable to those of capacity, strategies, and previous knowledge (for a review see Schneider & Bjorklund, 1998), the investigation of metacognitive monitoring processes in the context of everyday memory tasks is of additional interest. Therefore, the present study explores confidence judgments (CJs) in a complex episodic memory task across two age groups, paying specific attention to metacognitive monitoring of false memories. Monitoring faulty retrieval is especially important in eyewitness and academic contexts. Previous studies have documented relatively well-developed monitoring abilities for correct retrieval and revealed developmental progression for primary school age children in uncertainty monitoring abilities, that is, in the appropriate monitoring of false memories.

A central theoretical concept for CJs is *calibration*, which refers to the correlation between estimated and real performance. Generally, individuals tend to overestimate the accuracy of their performance, that is, they are miscalibrated or not well calibrated, independent of age (Ibabe & Sporer, 2004; Maki & Swett, 1987; Migueles & Garcia-Bajos, 1999; C. H. Renner & M. J. Renner, 2001; Schraw, Potenza, & Nebelsick-Gullet, 1993). The phenomenon is, however, more pronounced in children than in adults (Howie & Roebers, 2006; Pressley, Levin, Ghatala, & Ahmad, 1987; Roebers, 2002). Children's overestimation of their own performance is in correspondence with Flavell's (1979) early and general assumptions concerning children's performance optimism and can be considered as a protective factor against demotivation and frustration (Bjorklund & Bering, 2002). Yet, in many everyday situations and in every kind of test situation in academic contexts and in professional life realistic assessment of one's own performance plays a key role.

Another important aspect in the study of CJs is the ability to differentiate appropriately between correct and incorrect answers (*resolution*). This is measured by comparing mean CJs for correct and incorrect answers and by correlations between confidence and accuracy of recall (Howie & Roebers, 2006). Regarding the first index, adults typically give higher CJs after correct than after incorrect answers, pointing to appropriate metacognitive

differentiation (Granhag, Strömwall, & Allwood, 2000; Nolan & Markham, 1998; Robinson, Johnson, & Robertson, 2000). Some studies have documented developmental progression in metacognitive differentiation between the age of 7 and 10 (Pressley et al., 1987). In the early stages of this development, the ability to differentiate effectively between correct and incorrect answers seems to be strongly influenced by situational and task-relevant characteristics. For example, cognitively less demanding recognition tasks lead to earlier and better metacognitive differentiation in children than cued recall tasks (Roebers, 2002). Similar results have been found for the effects of question format. After unbiased questions, children as young as 8 years are able to differentiate between correct and incorrect answers in their CJs. After misleading questions, however, even 10-year-olds have difficulties doing so (Roebers, 2002). Furthermore, a mixture of unbiased and misleading questions leads to better metacognitive differentiation in primary school children than the uninterrupted use of misleading questions within a single interview (Roebers & Howie, 2003).

Regarding the second index of resolution, correlations between confidence and accuracy for text and event recall show medium correlations for both primary school children and adults (Maki, Shields, Wheeler, & Zacchilli, 2005; Perfect & Hollins, 1999; Roebers, von der Linden, Howie, & Schneider, 2006). Concerning resolution for currently non-recallable items (feeling-of-knowing judgments), a number of studies suggest that accuracy of feeling-of-knowing judgments (FOKs) increases constantly from middle childhood to adolescence (Wellman, 1977; Zabrucky & Ratner, 1986). Yet, some newer findings question this assumption: they failed to detect developmental progression during the primary school years (Butterfield, Nelson, & Peck, 1988; Lockl & Schneider, 2002).

From the existing empirical evidence so far, it also appears that developmental changes result in increasing uncertainty after incorrect answers. While the level of confidence after correct answers hardly differs between primary school children and adults, confidence after incorrect answers decreases with age (Roebers, 2002; Roebers & Howie, 2003). In spite of the great practical relevance of uncertainty monitoring and its theoretical aspects (Flavell, 1979; Schneider, 1998), very few studies have addressed the question of its developmental progression. For this reason, the focus of this study is to investigate how well children of different ages can monitor and report uncertainty.

One possibility to extend the investigation of uncertainty beyond the commonly used CJs after answerable questions is the use of unanswerable questions. Participants cannot encode information to answer these questions because it was not provided and therefore the appropriate response to this kind of question is "I don't know." Two recent studies that included metacognitive judgments after wrong answers to unanswerable questions detected hints of an emerging ability to report low confidence, that is, uncertainty, from the age of 8 years (Roebers, von der Linden, & Howie, 2006). Especially under facilitating context conditions (unbiased question format, delayed CJs), CJs after incorrect answers to unanswerable questions proved to be less optimistic than after incorrect answers to answerable questions. In other words, children from the age of 8 years could metacognitively differentiate between the two question types. This implies that they were able to distinguish between an effortful retrieval process when answering a difficult question (leading to an incorrect answer) and filling memory gaps with general knowledge when confronted with unanswerable questions (again leading to an inadequate answer). In sum, these findings of significant developmental progression in metacognitive differentiation under favorable conditions suggest that a general metacognitive deficit in children can be ruled out.

Another aspect of uncertainty monitoring which the present study aimed to explore is developmental progression in memory monitoring after failed retrieval (i.e., "don't know" answers). Conceptually, "don't know" answers are classified as metacognitive control processes resulting from metacognitive monitoring processes (i.e., high uncertainty; Nelson & Narens, 1994). After unsuccessful retrieval efforts, metacognitive monitoring can be assessed with FOKs. When giving FOKs, participants judge how confident they are that they will identify the correct answer to a question in a recognition test although they currently cannot recall it. Adults show lower confidence in answers reported after unsuccessful retrieval efforts (C. H. Renner & M. J. Renner, 2001). Children also tend to give lower FOKs in case of omission and commission errors compared to correct answers in pair-associate learning tasks (Lockl & Schneider, 2002).

The present study is, to our knowledge, the first attempt to investigate whether children report greater uncertainty after unsuccessful retrieval processes compared to spontaneous answers in the context of event memory with FOKs. Additionally, FOKs before correct and incorrect recognition will be compared. We expect the level of confidence in FOKs to be lower than in CJs. Assuming appropriate metacognitive differentiation, FOKs before correct recognition should be higher than before incorrect recognition. We do not presume that 8-year-olds possess those metacognitive differentiation competencies since former studies have shown noticeable limits to their metacognitive monitoring abilities.

Another well-known metacognitive phenomenon in the literature is that younger children do not spontaneously use metacognitive abilities in new situations. For example, children succeed in effective metacognitive monitoring (judgments-of-learning) significantly earlier than they master execution into efficient learning behavior (allocation of study time, Lockl & Schneider, 2003). Even after systematic and successful learning of metacognitive strategies, children often fail to use them in similar learning situations (Schreblowski & Hasselhorn, 2001). Apparently, children have a deficit in using potentially available metacognitive abilities. But children who observed an adult confederate giving CJs were shown to be significantly influenced in both their recall performance and in their CJs (Schwarz & Roebers, 2006). Therefore, we wanted to explore in this study whether metacognitive competencies in children could be released through observing an adult model (triggering effect, Bandura, 1969): If children are in principle able to report more differentiated CJs than they did in previous studies, observation of an adult model using the whole range of the scale (thereby also modeling uncertainty) might activate potential competencies. It was of further interest if a model who additionally explained and elaborated her CJs (with retrieval speed and vividness of memory, see Robinson et al., 2000) would lead to more appropriate CJs in children than a model who simply used the whole scale range. We expected older children to benefit more from observing the adult model because their metacognitive competencies should be developed further and more easily released (Bandura, 1969).

In sum, this study aimed to investigate developmental changes in children's metacognitive abilities in the context of a complex everyday memory task. It was of special interest to explore metacognitive differentiation between correct and incorrect memories and the ability to report uncertainty. In order to do so, CJs after unanswerable questions and FOKs were included in an event recall task. Another innovative aspect of this study was the question whether observing an adult model could activate potentially available abilities. To ensure comparability with previous studies, 7- and 9-year-olds were questioned two weeks after observing an event and asked to give CJs and FOKs.

Materials and methods

Overview A 2 (age group: 7-year-olds, 9-year-olds)×3 (experimental condition: CJs with rationale, CJs without rationale, control group)×2 (question format: unbiased, misleading)× 2 (question type: answerable, unanswerable) factorial design was utilized in this study, with age and experimental condition being between-subject factors, and question format and question type being within-subject factors. Dependent measures were accuracy of event recall, CJs and FOKs.

Sample A total of 120 participants (60 female and 60 male) from two age groups took part in the study. The 7-year-olds consisted of 60 children (30 girls) with a mean age of 7 years and 3 months (SD=7 months). There were 60 9-year-olds (30 girls) with a mean age of 9 years and 1 month (SD=7 months). In both age groups, 20 participants were randomly assigned to each experimental condition (CJs with rationale, CJs without rationale, control group) with the only constraint that male and female participants were equally distributed across conditions. Children were recruited from one large primary school in a small town in southern Germany, and came from predominantly lower to upper-middle-class backgrounds. Written consent was obtained from the parents prior to the study.

Procedure In small groups of up to ten individuals of one age group, participants watched a brief video (7 min) about children on a treasure hunt. Before the video started, participants were told to pay attention to the video, as the experimenter would later be interested in their opinion of it. (They were not informed about the real purpose of the study to prevent memory strategies that would affect results). At the end of the video, the group was briefly asked for their opinion of the video and thanked for their helpful comments.

An average of 14 days (SD=1 day) after stimulus presentation, children were individually questioned about the contents of the film in a quiet room in their school. Depending on the experimental condition, children were confronted with different situations when entering the room: One third of the children (control group) met only the interviewer, and after a short warm-up the interview started. Two thirds of the children (experimental conditions: CJs with rationale, CJs without rationale) saw the interviewer who questioned a female adult (confederate) who was unknown to the children. The adult served as a model for giving appropriate CJs. Each child was asked to take a seat and closely watch the model that reported CJs for an unknown video. The interviewer stressed that it was very good that the child had arrived early because he or she could now observe the adult participant who did a very good job at giving CJs. The child's attention was directed to the model and to the fact that she scattered her CJs along the entire scale. The model was praised for giving her CJs while the child was observing. In the experimental condition "CJs with rationale", the model in addition to the CJs gave reasons for choosing a particular CJ, for example, retrieval speed and vividness of memory ("I am uncertain because I had to ponder quite long to come up with an answer", "I am very sure because I can recall this detail vividly"). In contrast, the model in the experimental condition "CJs without rationale" did not elaborate her CJs. She only modeled using the whole scale range, especially the end representing uncertainty. In both experimental conditions, the model distributed her CJs equally across all categories of the scale. After judging confidence of six answers in the presence of the child, the adult confederate was thanked for participation and then left the room. Next, the interviewer explained to the child that she wanted to question him or her

about the video the child had seen 2 weeks prior and afterwards wanted to know how sure the child was that each answer was correct. The child was also instructed to remember the confederate's behavior when giving CJs and to give as many correct answers as possible. Then the memory interview was conducted (see below). After asking all questions, the interviewer explained to the child that now she wanted to find out which questions were hard and which were easy. The child should indicate for each answer how sure he or she was that the given answer was correct. For this purpose, participants were presented three cards (Fig. 1), which showed a pondering child and one of the following lines: "really sure", "somewhat sure" and "not sure." The scale was used successfully in previous studies with children of primary school age (Roebers & Howie, 2003; Roebers et al., 2006).

The interviewer checked that the child could read the words on each card, and then explained what each meant (for example, "somewhat sure" was defined as "when you are neither sure nor unsure"). Each child was asked a set of training questions about facts that were unrelated to the event to be recalled and that required different confidence ratings, e.g., "How old are you?" (Expected confidence rating: "really sure"), "How old am I?" (Expected rating: "not sure"), and "What are you going to do this afternoon?" (Expected rating: "somewhat sure"). After each answer, the child was asked "and how sure are you about that?" and asked to point to the appropriate card. If a child did not use all three confidence categories during training, or did not indicate the appropriate level of confidence for a question, he or she was corrected, given a rationale for the appropriate rating and asked another training question. After three appropriate training CJs, CJs for questions of the memory interview were collected. As in previous studies, children generally learned the use of the confidence scale with ease. For questions that had been answered with "don't know" during the memory interview no CJs but FOKs were gathered. After collection of all CJs the interviewer therefore explained to the child: "You could not recall the answer to a few questions. I will ask you these questions again and I want you to tell me how sure you are that you can recognize the correct answer among three possible answers." For each question that previously had been answered with "don't know" the child gave a FOK by indicating on the confidence scale (Fig. 1) how sure he or she could recognize the correct answer. Subsequently the child was asked to choose the correct answer from three possible alternatives. When the questioning was over, the child was praised for his/her good performance, given a small gift, thanked for his/her help and escorted back to the classroom.

Materials The memory test consisted of 28 target questions, 16 of which were answerable and 12 were unanswerable. Of the 16 answerable questions, 8 were asked in a misleading
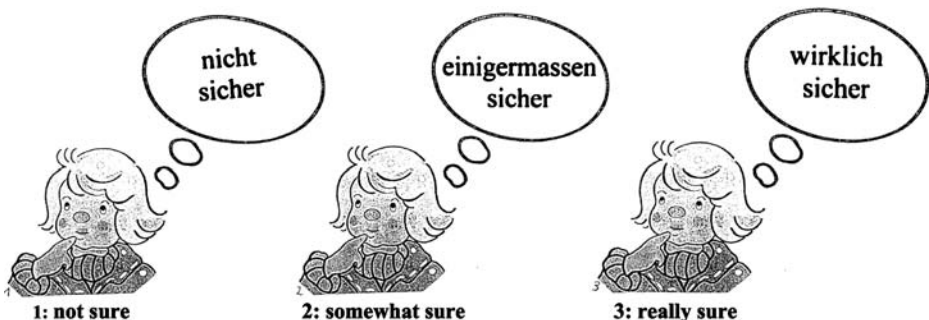

Fig. 1 Confidence instrument used in the study (note: higher values indicate higher confidence)

question format, and the remaining 8 in an unbiased question format. Unbiased questions asked for a specific detail without providing the correct answer ("How was the castle destroyed"?). In contrast, misleading questions always suggested an incorrect answer ("The children used candles to see in the dark, didn't they?" when in fact they used flashlights). Of the 12 unanswerable questions, 6 were asked in a misleading question format, and the remaining 6 in an unbiased question format. Half of the participants in each age group were given one version (set) of the questions, and the remaining participants were given the other version. In order to control for question content, the questions asked in misleading question format in set 1 were asked in a unbiased format in set 2, and questions asked in unbiased question format in set 1 were asked in misleading question format in set 2. Additionally, both sets included the same six correct leading questions which served as filler questions, in order to counteract any impression that "no" was the correct answer to all leading questions, to maintain the interviewer's credibility and to have some easy questions included in the interview. Answers to these questions were not included in the analyses. These question sets have been employed in previous studies (Roebers & Fernandez, 2002; Roebers & Howie, 2003; Roebers et al., 2006; Schwarz, Roebers, & Schneider, 2004).

Results[1]

We will first present memory performance in terms of correct answers to answerable questions and "don't know"-answers to unanswerable questions. Then we will analyze CJs and FOKs as measures of metacognitive monitoring to examine whether children can differentiate varying between degrees of certainty. All measures are reported as a function of age and experimental condition to explore developmental progression and the influence of observing an adult model. Preliminary analysis assessing the effect of gender did not reveal any systematic differences between male and female participants. Thus, data were collapsed across this variable. As the post-hoc procedure and in order to follow-up on main effects, Student-Newman-Keuls tests were used in order to ensure neither too progressive nor too conservative testing. Level of significance was set to $p<0.05$. As answerable and unanswerable questions differ conceptually from each other (Roebers & Fernandez, 2002; Waterman, Blades, & Spencer, 2001; 2004), and in order to facilitate interpretation of results, the two question types will be reported separately.

Recall: answerable questions Table 1 presents the mean percentages of correct answers to answerable questions as a function of experimental condition, question format, and age. An ANOVA was conducted on the percentages of *correct* answers, with age and experimental condition as between-subject factors, and question format as within-subject factor. It revealed a significant main effect of question format, $F (1, 114)=27.93$, $\eta^2=0.20$, and a significant main effect of age, $F (1, 114)=21.62$, $\eta^2=0.16$. The main effect of experimental condition and all interactions failed to reach significance. As in previous studies, participants gave less correct answers to unbiased [30.4%] than to misleading questions [42.4%]. 9-year-olds [42.4%] gave more correct answers than 7-year-olds [30.4%].

Unanswerable questions Table 2 presents the mean percentages of "don't know" answers to unanswerable questions as a function of experimental condition, question format, and

---

[1] Participants included in the different analysis vary according to the number of children who reported the corresponding judgments.

age. An ANOVA on the percentages of "don't know" answers with age, experimental condition, and question format (within-subject factor) revealed a significant main effect of question format, $F (1, 114)=546.10$, $\eta^2=0.83$, a significant main effect of age, $F (1, 114)= 5.75$, $\eta^2=.05$, and a significant interaction between age and experimental condition, $F (1, 114)=3.49$, $\eta^2=0.06$. The main effect of experimental condition and all other interactions failed to reach significance. Across conditions and in all age groups, unbiased questions [66.7%] were answered more often with "don't know" than misleading questions [13.2%]. Following up on the interaction between age and condition, post-hoc tests revealed that in the experimental condition "CJs without rationale" 7-year-olds [43.8%] answered more often with "don't know" than 9-year olds [28.0%]. In contrast in the condition "CJs with rationale" [45.0 vs. 39.6%, $t (38)=0.98$, n.s.] and in the control group [40.8 vs. 42.5%, $t (38)=-0.36$, n.s.] age groups did not differ significantly from each other.

Metacognitive monitoring Metacognitive judgments were also analyzed separately for each question type because in the case of the answerable questions, confidence judgments were given both after correct and incorrect answers. In the case of unanswerable questions, however, confidence judgments were only given for incorrect answers. Readers are reminded that for this question type, "I don't know" is the appropriate answer.

CJs after answerable questions Figure 2 presents the mean CJs after correct and incorrect answers to answerable questions and after incorrect answers to unanswerable questions as a function of experimental condition, question format, and age. For answerable questions an ANOVA was conducted with age and experimental condition as between-subject factors, and question format as within-subject factor to explore if children were able to distinguish correct and incorrect answers in their CJs. It revealed main effects of correctness of answer, $F (1, 93)=27.27$, $\eta^2=0.23$, question format, $F (1, 93)=53.30$, $\eta^2=0.36$, an interaction between question format and age, $F (1, 93)=7.11$, $\eta^2=0.07$, an interaction between correctness of answer and age, $F (1, 93)=5.86$, $\eta^2=0.06$, an interaction between correctness of answer and question format, $F (1, 93)=15.23$, $\eta^2=0.14$, and a three-way interaction between question format, correctness of answer, and age, $F (1, 93)=5.36$, $\eta^2=0.06$. To further explore the three-way interaction that qualified the main effects and two-way interactions unbiased and misleading questions were analyzed separately. For the *unbiased* questions, there was a significant main effect of correctness of answer, $F (1, 102)=68.16$, $\eta^2=0.40$. As in previous studies, children were able to give higher CJs after correct [2.79] than after incorrect answers [2.37]. However, no age differences were found for unbiased

Table 1 Mean percentage of the correct answers to answerable questions as a function of experimental condition, question format, and age (standard error of mean in parentheses; $N=8$ unbiased and $N=8$ misleading questions)

	Control group	Confidence judgments without rationale	Confidence judgments with rationale	Overall
Unbiased questions				
7-year-olds	26.9 (3.9)	23.8 (3.1)	26.3 (3.4)	25.6 (2.0)
9-year-olds	35.0 (3.1)	35.6 (2.8)	35.0 (3.0)	35.2 (1.7)
Misleading questions				
7-year-olds	37.5 (4.9)	36.3 (5.0)	31.9 (4.9)	35.2 (2.8)
9-year-olds	48.8 (5.9)	43.1 (4.0)	56.9 (5.1)	49.9 (3.0)

Table 2 Mean percentage of the "don't know" answers to unanswerable questions as a function of experimental condition, question format, and age (standard error of mean in parentheses; $N=6$ unbiased questions and $N=6$ misleading questions)

	Control group	Confidence judgments without rationale	Confidence judgments with rationale	Overall
Unbiased questions				
7-year-olds	71.7 (4.4)	71.7 (3.7)	70.0 (6.0)	71.1 (2.7)
9-year-olds	67.5 (4.1)	51.7 (4.3)	67.5 (5.3)	62.2 (2.8)
Misleading questions				
7-year-olds	10.0 (3.9)	15.8 (4.6)	20.0 (4.9)	15.3 (2.6)
9-year-olds	17.5 (5.1)	4.2 (1.7)	11.1 (2.4)	11.1 (2.0)

questions. For the *misleading* questions, conversely, a significant interaction between age and correctness of answer was found, F (1, 105)=12.66, η^2=0.11. 7-year-olds [2.15 vs. 2.30, t (53)=1.45, n.s.] did not differ in their CJs between correct and incorrect answers, while 9-year-olds [2.50 vs. 2.19, t (56)=−4.19] did.

CJs after incorrect answers to unanswerable questions To explore whether CJs after answerable and unanswerable questions differed from each other, another ANOVA was conducted with age and experimental condition as between-subject factors, and question type and question format as within-subject factors. There was a main effect of question type, F (1, 94)=26.66, η^2=0.22, but no other significant main effects or interactions. CJs after answerable questions [2.33] were higher than after unanswerable questions [2.13] but were not influenced by age or experimental condition.

FOKs With regard to FOK-judgments it was of interest if children could distinguish in the level of their metacognitive judgments between correct and incorrect recognition and between CJs and FOKs. Because the numbers of FOKs were limited to the numbers of "I don't know" responses during recall, Chi^2- analyses were conducted to account for the overall low frequencies. Each FOK-judgment was compared to the answer given in a subsequent recognition test with three alternatives. Figure 3 (upper half) presents the distribution of FOKs over the three points of the confidence scale (not sure, somewhat sure, really sure) as a function of question type, age, and correctness of recognition. For unanswerable questions, there were only incorrect recognitions because no information had been encoded for this question type. For answerable questions, a Chi^2-test was conducted to compare FOKs before correct and incorrect recognition. Both 7-year-olds [χ^2(2)=3.07, n.s.] and 9-year-olds [χ^2(2) =1.75, n.s.] distributed their FOKs before correct and incorrect recognition with equal probabilities across the three points of the confidence scale. This indicates no reliable differentiation in metacognitive judgments between correct or incorrect recognition.

Figure 3 (lower half) presents the distribution of CJs across the three points of the confidence scale (not sure, somewhat sure, really sure) as a function of question type, age, and correctness of answer. To explore whether CJs and FOKs were distributed differently across the three points of the confidence scale in a way that indicated lower levels of confidence for FOKs than for CJs (and thus metacognitive differentiation between both judgments), Chi^2-tests were conducted separately for each question type and correctness of answer. For correct answers to answerable questions, distribution of CJs differed significantly from the FOK-judgments´ distribution [7-year-olds: χ^2(2)=57.57; 9-year-olds: χ^2(2)=75.81]. Participants in both age groups used the categories 'somewhat sure'
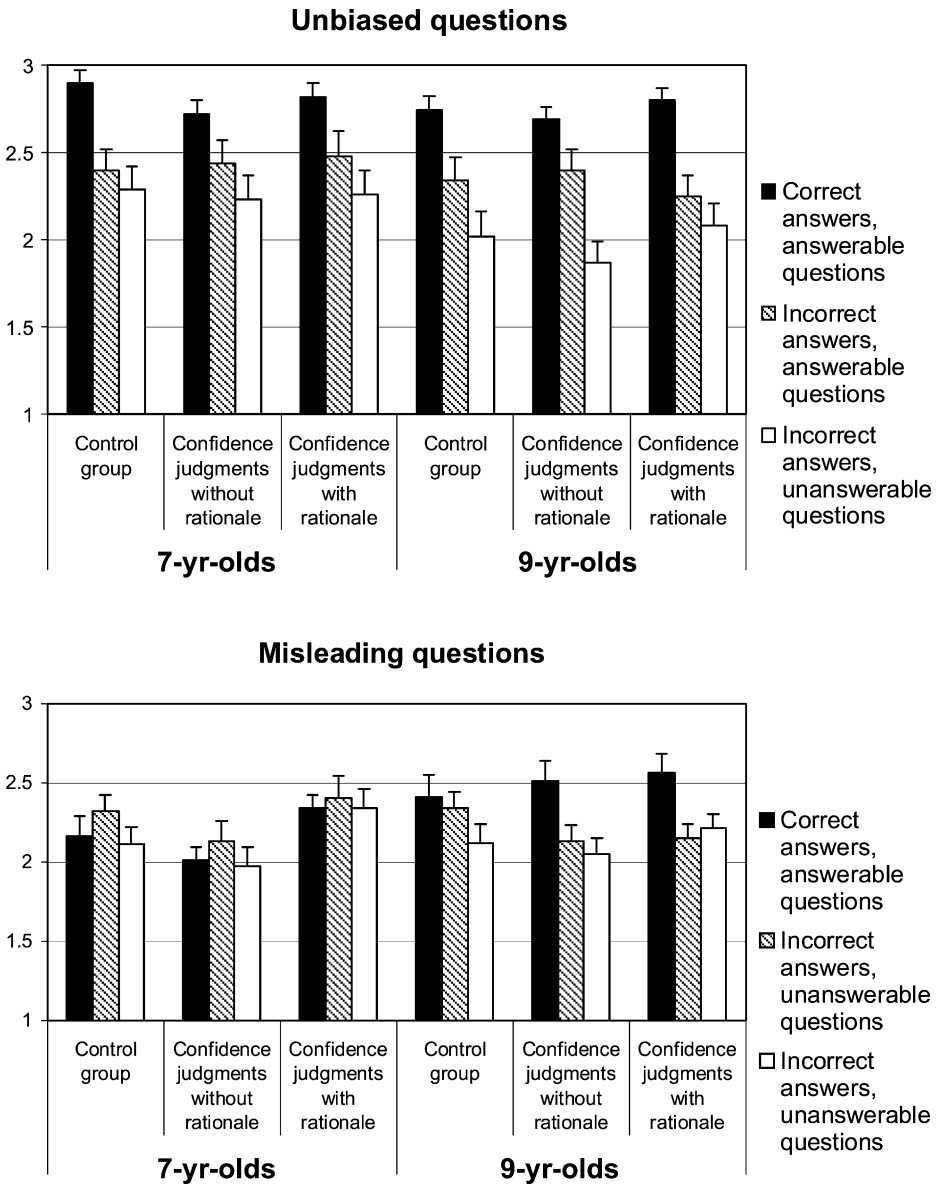
Unbiased questions

Misleading questions

Fig. 2 Mean CJs after correct and incorrect answers to answerable questions and after incorrect answers to unanswerable questions as a function of experimental condition, question format, and age group

and 'not sure' more often in their FOKs than in their CJs. After incorrect answers distribution of CJs differed significantly from distribution of FOKs in both age groups [7-year-olds: $\chi^2(2)=39.42$; 9-year-olds: $\chi^2(2)=12.44$]. Again FOKs compared to CJs were more strongly distributed towards lower confidence.
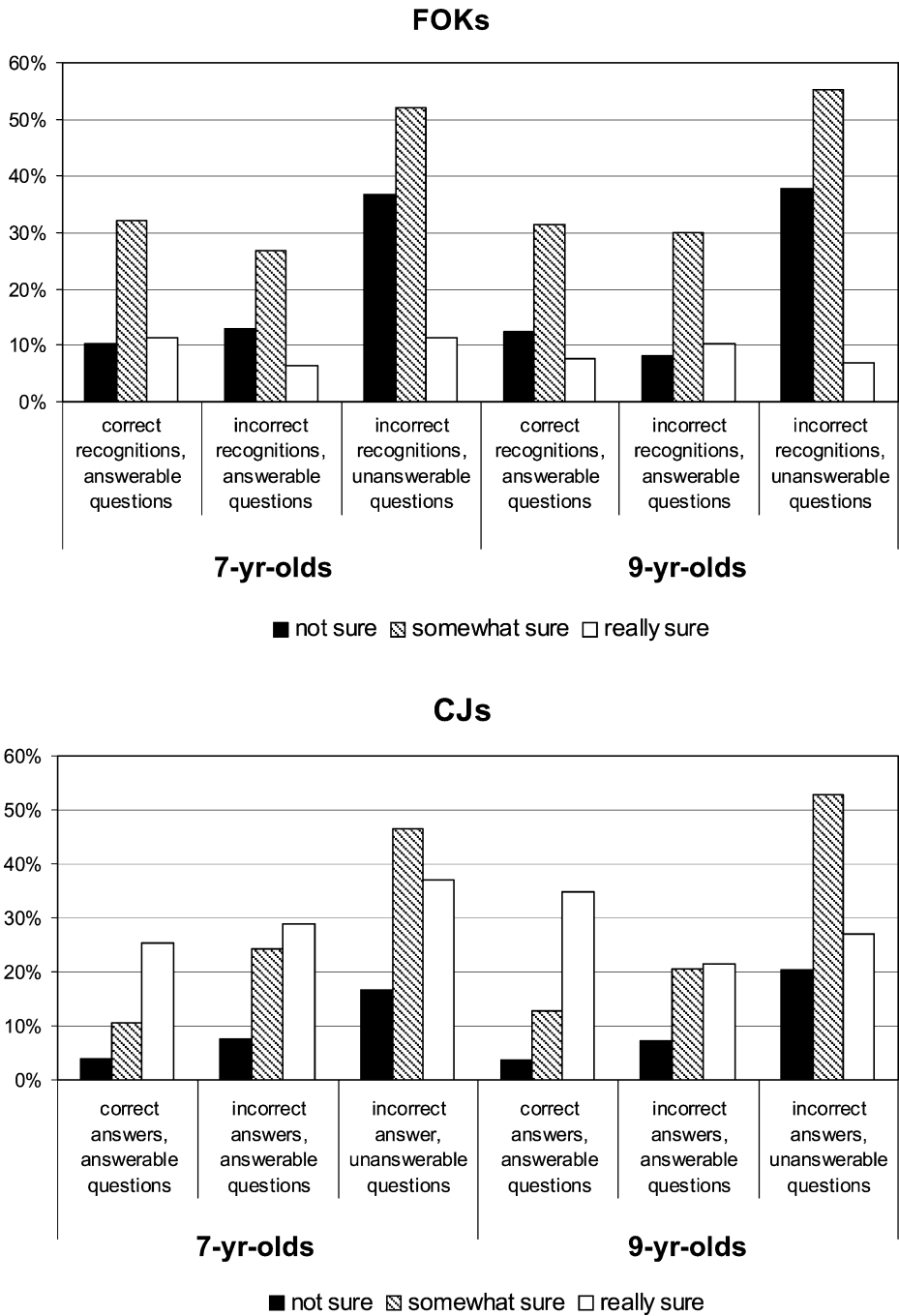
Fig. 3 Distributions of FOKs and CJs over the three points of the confidence scale (*not sure*, *somewhat sure*, *really sure*) as a function of question format, age, and correctness of answer

For unanswerable questions, CJs after incorrect answers were compared to FOKs before incorrect recognition using a Chi^2-Test. Comparable to answerable questions, distributions of CJs and FOKs differed significantly in both age groups [7-year-olds: $\chi^2(2)=63.02$; 9-year-olds: $\chi^2(2)=47.13$]. Distribution of FOKs compared to CJs showed a shift towards uncertainty, that is, the lower categories of the confidence scale were used more frequently.

CJs accuracy in relation to answerable questions In order to assess CJs accuracy as a function of age group, question format, and experimental condition, Goodman–Kruskal gamma correlations between CJs (1=not sure, 2=somewhat sure, 3=really sure) and recall performance were computed for each participant, and then averaged for each single cell in the experimental design. Gamma correlations are considered to be the most appropriate measure of metacognitive accuracy (Nelson, 1984) and are commonly used in the contemporary literature (Dunlosky, Rawson, & Middleton, 2005; Koriat, Ma'ayan, Sheffer, & Bjork, 2006; Lockl & Schneider, 2003; Maki et al., 2005). A positive correlation indicates that higher CJs were given for items that were recalled correctly than for those recalled incorrectly. Table 3 shows mean gamma correlations for CJs after answerable questions as a function of age, question format, and experimental condition. Firstly, for unbiased questions all mean gamma correlations were significantly different from zero (using one-tailed t tests), while for misleading questions this was only the case for 7-year-olds in the control condition and for 9-year-olds in the conditions "CJs with and without rationale." An ANOVAs was conducted with age and experimental condition as between-subject factors and question format as within-subject factor. It revealed a significant main effect of question format, F (1, 73)=17.11, η^2=0.19, a significant interaction between question format and experimental condition, F (1, 73)=4.19, η^2=0.10, and a significant interaction between question format and age, F (1, 73)=7.95, η^2=0.09. Gamma correlations for unbiased questions [0.59] were higher than for misleading questions [0.16]. For unbiased questions there was no difference between the two age groups (0.72 vs. .57, t (89)=1.09, n.s.), while for misleading questions 7-year-olds had lower gamma correlations than 9-year-olds (−0.11 vs. .45, t (96)=−3.97). In the control condition gamma correlations for unbiased questions were higher than for misleading questions (0.69 vs.−0.14, t (26)=4.70), while in the conditions CJs with (0.63 vs. .36, t (25)=1.65, n.s.) and without rationale (0.45 vs. 0.36, t (25)=0.44, n.s.) gamma correlations for both questions formats did not differ significantly from each other.

Table 3 Mean gamma correlations for confidence judgments after answerable questions as a function of the age, question format, and experimental condition (standard error of mean in parenthesis)

	Control group	Confidence judgments without rationale	Confidence judgments with rationale	Overall
Unbiased questions				
7-year-olds	0.82 (0.17)	0.62 (0.18)	0.70 (0.17)	0.71 (0.10)
9-year-olds	0.64 (0.16)	0.41 (0.16)	0.67 (0.16)	0.57 (0.09)
Misleading questions				
7-year-olds	−0.33 (0.17)	−0.04 (0.17)	0.06 (0.19)	−0.10 (0.10)
9-year-olds	0.26 (0.17)	0.50 (0.17)	0.58 (0.16)	0.45 (0.10)

Discussion

One aim of the present study was to replicate previous findings concerning developmental progression in metacognitive monitoring in the context of an event recall task over the primary school years. Thereby, the emergence of metacognitive competencies in experiencing and reporting low confidence or uncertainty for incorrect answers was investigated in detail. To achieve these goals, CJs as a function of age, question format, and correctness of answer were assessed as in previous studies. Our results fully replicate developmental changes documented in previous studies (Roebers, 2002; Roebers & Howie, 2003) with children of both age groups being able to differentiate correct and incorrect answers to unbiased questions in their CJs but only 9-year-olds were able to do so for misleading questions. This argues for the reliability of our data.

Gamma correlations as an indicator of CJs accuracy showed the same pattern. For unbiased questions both age groups reached relatively high gamma correlations and age differences were negligible. In contrast, 9-year-olds performed better on the monitoring of answers to misleading questions than 7-year-olds. Level of CJs was similar to those found in previous studies (Maki et al., 2005; Roebers et al., 2006). However, the fact that the gammas were far from perfect in both age groups even under favorable conditions (unbiased questions) points to the difficulties of achieving accurate memory performance postdictions assessments. Yet, this does not point to specific deficits in children but this overestimation of individuals' competencies is generally found in adults as well (C. H. Renner & M. J. Renner, 2001; Schraw et al., 1993).

To explore metacognitive abilities concerning low confidence, CJs after incorrect answers to answerable questions and unanswerable questions were compared. Additionally, FOKs before correct and incorrect recognition were contrasted. FOKs had been collected after "don't know" answers in the memory interview with the "don't know" answers indicating failed retrieval from long-term memory. Furthermore CJs and FOKs were compared.

CJs after incorrect answers to unanswerable questions were significantly lower than CJs after incorrect answers to answerable questions. This was true for both question formats and both age groups. Seven-year-olds showed correct differentiation but especially 9-year-olds reliably discriminated in their CJs between feeling uncertain about retrieved information and guessing. This implies that they were able to distinguish between incorrect retrieval of event-related information when answering a difficult question and filling memory gaps with general knowledge when confronted with unanswerable questions. This new finding is in line with results from Schwanenflugel, Henderson, and Fabricius (1998) who showed that it is only from an age of about 9 years that children can distinguish meanings of mental verbs that express different degrees of confidence (e.g., knowing vs. guessing; remembering vs. reasoning).

We expected that children of both age groups could successfully differentiate between successful retrieval followed by a CJ and failed retrieval ("don't know" answer) followed by a FOK, resulting in generally lower FOKs than CJs. Consistent with this expectation, comparisons between FOKs and CJs revealed that 7- and 9-year-old children can metacognitively differentiate these degrees of confidence. Although the database is somewhat limited to relatively few FOKs, descriptive analyses and distribution of judgments across the scale clearly point to the fact that FOKs in both age groups, for both question formats, and the two question types were significantly lower than the corresponding CJs. For questions that were initially answered with "don't know," indicating unsuccessful retrieval from long-term memory, lower confidence was reported than after questions for which participants provided specific answers even if they were unsure or had

to guess. These findings suggest that reporting relatively low confidence after providing an incorrect answer is not necessarily children's cognitive limit to report uncertainty. It appears that, like adults (C. H. Renner & M. J. Renner, 2001), children of only 7 and 9 years of age seem to be able to use the experience of unsuccessful retrieval efforts to develop more appropriate metacognitive monitoring leading to a further decrease in confidence compared to questions to which answers were given. The results can be interpreted as showing that at least for the unbiased, less demanding question format, the execution of monitoring is successful in children from 7 years of age on.

Limits of metacognitive competencies concerning uncertainty monitoring become obvious when differentiating FOKs before correct and incorrect recognition. Participants in both age groups distributed their judgments equally before correct and incorrect recognition across the whole confidence scale. This result suggests that metacognitive monitoring of FOKs poses greater metacognitive demands than monitoring in the context of successful memory retrieval (children typically succeed in differentiating correct and incorrect answers in CJs) and develops later in childhood. Our findings in the present study are incompatible with the assumption that children have a general problem with uncertainty monitoring, yet, the results concerning FOKs suggest that the lower the confidence the more demanding the monitoring processes seem to be. This interpretation is supported by results from previous studies showing more appropriate uncertainty monitoring under favorable compared to less favorable conditions (unbiased question format, delayed CJs, Roebers et al., 2006).

Another goal of the present study was to investigate whether one instance of a short-term observation of an adult expressing CJs leads to activation of existing metacognitive abilities in children that result in more differentiated CJs. No effect of observing the adult model was found for event recall or CJs. Children do not immediately and openly benefit from observing a model through releasing available skills in the sense of a triggering effect. It is possible that the necessary metacognitive abilities do exist but are too complex to be activated by a short-term observation. It would be interesting to see if and how an intensive metacognitive training including practice and precise feedback could improve children's metacognitive abilities in the context of complex everyday memory tasks.

Another possible explanation is that children have a general metacognitive deficit and therefore could not be influenced by the model. Findings of former studies contradict this assumption by showing improved metacognitive abilities when simplifying tasks (Roebers, 2002; Roebers & Howie, 2003). Furthermore, Allwood, Jonsson, and Granhag (2005) augmented accuracy of children's metacognitive judgments through feedback, and Schwarz and Roebers (2006) demonstrated a change in CJs when an adult confederate was present during the interview. These findings support the existence of metacognitive abilities in primary school children and the assumption that they can in principle be influenced positively. Possibly, repeated experience and feedback either in everyday life or in intensive training sessions is necessary to achieve metacognitive improvements in this regard. In sum, the presented data suggests that non-interactive observations have no immediate or positive influence on the developing metacognitive abilities of primary school children.

The ability to assess confidence in given answers and to perceive when one is guessing is of great importance in everyday life, both for adults and children. In general, a positive perception of one's performance may be protective against loss of motivation and adverse effects on self-concept (Bjorklund & Bering, 2002). Yet, in several contexts assessment of uncertainty is more important than assessment of certainty. In academic contexts, assessment of how well one knows new learning material is important to effectively implement further study. For eyewitness testimony situations, these metacognitive abilities

are decisive determinants of credibility and quality. Answers that may turn out to be wrong should be withheld in order to augment accuracy of one's verdict. In sum, the present study reveals both unknown competencies and slow developmental progression in the lower half of the certainty-uncertainty continuum in children.

References

Allwood, C. M., Jonsson, A.-C., & Granhag, P. A. (2005). The effects of source and type of feedback on child witnesses' metamemory accuracy. *Applied Cognitive Psychology, 19*, 331–344.

Bandura, A. (1969). *Principles of behavior modification*. Oxford: Holt, Rinehart, & Winston.

Bjorklund, D. F., & Bering, J. M. (2002). The evolved child: Applying evolutionary developmental psychology to modern schooling. *Learning and Individual Differences, 12*, 347–373.

Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology, 24*, 654–663.

DeMarie, D., & Ferron, J. (2003). Capacity, strategies, and metamemory: Test of a three-factor model of memory development. *Journal of Experimental Child Psychology, 84*, 167–193.

Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language, 52*, 551–565.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906–911.

Granhag, P. A., Strömwall, L. A., & Allwood, C. M. (2000). Effects of reiteration, hindsight bias, and memory on realism in eyewitness confidence. *Applied Cognitive Psychology, 14*, 397–420.

Howie, P., & Roebers, C. M. (2006). Developmental progression in the confidence-accuracy relationship in event recall: New insights provided by a calibration perspective. *Applied Cognitive Psychology* (in press).

Ibabe, I., & Sporer, S. L. (2004). How you ask is what you get: On the influence of question form on accuracy and confidence. *Applied Cognitive Psychology, 18*, 711–726.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.

Koriat, A., & Goldsmith, M. (1998). The role of metacognitive processes in the regulation of memory performance. In G. Mazzoni & T. O. Nelson (Eds.), *Metacognition and cognitive neuropsychology* (pp. 97–118). Mahwah, NJ: Erlbaum.

Koriat, A., Ma'ayan, H., Sheffer, L., & Bjork, R. A. (2006). Exploring a mnemonic debiasing account of the underconfidence-with-practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 595–608.

Lockl, K., & Schneider, W. (2002). Developmental trends in children's FOK judgments. *International Journal of Behavioral Development, 26*, 327–333.

Lockl, K., & Schneider, W. (2003). Metakognitive Überwachungs- und Kontrollprozesse bei der Lernzeiteinteilung von Kindern. *Zeitschrift für Pädagogische Psychologie, 17*, 173–183.

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology, 97*, 723–731.

Maki, R. H., & Swett, S. (1987). Metamemory for narrative text. *Memory & Cognition, 15*, 72–83.

Migueles, M., & Garcia-Bajos, E. (1999). Recall, recognition and confidence patterns in eyewitness testimony. *Applied Cognitive Psychology, 13*, 257–268.

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

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The Psychology of Learning and Motivation: Advance in Research and Theory* (Vol. 26, pp. 125–173). New York: Academic.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In A. P. Shimamura & J. Metcalfe (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge, MA: MIT.

Nolan, J., & Markham, R. (1998). The accuracy–confidence relationship in an eyewitness task: Anxiety as a modifier. *Applied Cognitive Psychology, 12*, 43–54.

Perfect, T., & Hollins, T. (1999). Feeling-of-knowing judgments do not predict subsequent recognition performance for eyewitness memory. *Journal of Experimental Psychology: Applied, 5*, 250–264.

Pierce, S. H., & Lange, G. (2000). Relationships among metamemory, motivation and memory performance in young school-age children. *British Journal of Developmental Psychology, 18*, 121–135.

Pressley, M., Levin, J. R., Ghatala, E. S., & Ahmad, M. (1987). Test monitoring in young grade school children. *Journal of Experimental Child Psychology, 43*, 96–111.

Renner, C. H., & Renner, M. J. (2001). But I thought I knew that: Using confidence estimation as a debiasing technique to improve classroom performance. *Applied Cognitive Psychology, 15*, 23–32.

Robinson, M. D., Johnson, J. T., & Robertson, D. A. (2000). Process versus content in eyewitness metamemory monitoring. *Journal of Experimental Psychology: Applied, 6*, 207–221.

Roebers, C. M. (2002). Confidence judgments in children's and adults' event recall and suggestibility. *Developmental Psychology, 38*, 1052–1067.

Roebers, C. M., & Fernandez, O. (2002). The effects of accuracy motivation on children's and adults' event recall, suggestibility, and their answers to unanswerable questions. *Journal of Cognition and Development, 3*, 415–443.

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.

Roebers, C. M., von der Linden, N., & Howie, P. (2006). Favourable and unfavourable conditions for children's confidence judgments. *British Journal of Developmental Psychology* (in press).

Roebers, C. M., von der Linden, N., Howie, P., & Schneider, W. (2006). Children's metamemorial judgments in an event recall task. *Journal of Experimental Child Psychology* (in press).

Schneider, W. (1998). The development of procedural metamemory in childhood and adolescence. In G. Mazzoni & T. O. Nelson (Eds.), *Metacognition and cognitive neuropsychology: Monitoring and control processes* (pp. 1–21). Mahwah, NJ: Erlbaum.

Schneider, W., & Bjorklund, D. F. (1998). Memory. In W. Damon (Ed.), *Handbook of child psychology: Volume 2: Cognition, perception, and language* (pp. 467–521). New York: Wiley.

Schneider, W., Schlagmüller, M., & Visé, M. (1998). The impact of metamemory and domain-specific knowledge on memory performance. *European Journal of Education, 13*, 91–103.

Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology, 18*, 455–463.

Schreblowski, S., & Hasselhorn, M. (2001). Zur Wirkung zusätzlicher Motivveränderungskomponenten bei einem metakognitivem Textverarbeitungstraining. *Zeitschrift für Pädagogische Psychologie, 15*, 145–154.

Schwanenflugel, P. J., Henderson, R. L., & Fabricius, W. V. (1998). Developing organization of mental verbs and theory of mind in middle childhood: Evidence from extensions. *Developmental Psychology, 34*, 512–524.

Schwarz, S., & Roebers, C. M. (2006). Developmental progression in the effects of social influence on children's eyewitness performance and their metacognitive monitoring. *Journal of Experimental Child Psychology, 94*, 229–248.

Schwarz, S., Roebers, C. M., & Schneider, W. (2004). Entwicklungsveränderungen in Konformität und in kognitiven Folgen sozialer Beeinflussung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 36*, 173–181.

Waterman, A. H., Blades, M., & Spencer, C. (2001). Interviewing children and adults: The effect of question format on the tendency to speculate. *Applied Cognitive Psychology, 15*, 521–531.

Waterman, A. H., Blades, M., & Spencer, C. (2004). Indicating when you do not know the answer: The effect of question format and interviewer knowledge on children's 'don't know' responses. *British Journal of Developmental Psychology, 22*, 335–348.

Wellman, H. M. (1977). Preschoolers' understanding of memory-relevant variables. *Child Development, 48*, 1720–1723.

Zabrucky, K., & Ratner, H. H. (1986). Children's comprehension monitoring and recall of inconsistent stories. *Child Development, 57*, 1401–1418.