Metacognition Learning (2011) 6:3–23
DOI 10.1007/s11409-010-9063-3

The effects of summary production and encoding condition on children's metacognitive monitoring

Nicole von der Linden · Wolfgang Schneider · Claudia M. Roebers

Received: 27 January 2009 / Accepted: 22 October 2010 /
Published online: 24 November 2010

Abstract Two studies were conducted to investigate whether context variations were suitable to improve metacognitive judgments in children in a complex, everyday memory task. In the first phase of each experiment, participants were shown a short event (video) and gave judgments-of-learning (JOLs), that is, rated their certainty that they would later be able to recall specific details correctly. In the second phase of the experiments, participants took part in a memory interview about the memory event and gave confidence judgments (CJs), that is, rated their certainty that the provided answers to the memory questions were correct. Study 1 specifically investigated the potential positive influence of giving a verbal summary before the JOL-interview on metacognitive monitoring, whereas Study 2 had a closer look on the effect of intentional versus non-intentional encoding on JOL and CJ accuracy. Results revealed no significant influence of giving a summary and hardly any effect of encoding condition on metamemory monitoring although children from age 6 on showed adequate monitoring performance. JOL accuracy appears to be a complex process, which is even more difficult to influence in children than in adults.

Keywords Metacognition · Monitoring · Judgments-of-learning · Confidence judgments · Children

Metacognitive monitoring has been the focus of a large number of research studies in the past 30 years (Nelson and Narens 1990, 1994; W. Schneider and Lockl 2008). The main research motivation stems from the fact that subjective metacognitive judgments influence self-paced

This research was conducted as part of a research project on metacognitive factors involved in children's event recall and financed by the German Research Foundation (DFG-Gz. SCHN 315/34-2). We wish to thank all participating children and adults as well as teachers, principals and parents for their cooperation.

N. von der Linden (✉) · W. Schneider
Department of Psychology, University of Würzburg, Röntgenring 10, 97070 Würzburg, Germany
e-mail: linden@psychologie.uni-wuerzburg.de

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

learning behavior and its outcomes (for a review, see Son and Metcalfe 2000). In the current article two studies are presented in which participants' monitoring competencies were explored in the context of a complex everyday memory task. One major purpose was to improve monitoring accuracy by producing an active reproduction of the to-be-judged material (Study 1) and thus making JOL-relevant cues like ease of retrieval or accessibility of the material more salient (Benjamin and Bjork 1996; Koriat 1993, 1997). A second aim was to explore the influence of intentional vs. non-intentional encoding on monitoring accuracy (Study 2), thereby using a direct instruction to memorize and monitor the learning material as a means to reduce persistent overconfidence in young children (Howie and Roebers 2007; Pressley et al. 1987).

In recent years there has been a growing body of literature examining metamemory monitoring judgments in children (see Brown et al. 1983; Flavell et al. 2002; W. Schneider and Lockl 2008 for overviews). Of special interest are judgments-of-learning (JOLs), which provide subjective information about the degree to which encoded information is mastered and can potentially be recalled during a future memory test (Nelson and Narens 1990, 1994). Several studies have documented that children are able to monitor their learning process from the time they enter school (Koriat et al. 2009; Roebers et al. 2007). More specifically, it has repeatedly been shown for pair-associate learning tasks, that children between 7 and 12 years adequately differentiate between easy (high associative) and hard (low associative) item pairs. Children typically give higher JOLs for easy than for hard pairs, adequately predicting that their recall of the easy item pairs will be superior than the one of the hard item pairs (Koriat et al. 2009; Koriat and Shitzer-Reichert 2002; Lockl and W. Schneider 2003). Hardly any developmental progression was found for this kind of task during the primary school years (Lockl and W. Schneider 2003). Accuracy of metacognitive judgments ranges from low to moderate for pair-associate tasks (Koriat and Shitzer-Reichert 2002; W. Schneider et al. 2000), indicating that precise performance prediction poses a problem.

Developmental literature of monitoring processes has often focused on quite simple memory material, like pair-associates or pictures (Butterfield et al. 1988; Dufresne and Kobasigawa 1989; Koriat and Shitzer-Reichert 2002; Lockl and W. Schneider 2003; Pressley and Ghatala 1989). This is a sharp contrast to everyday life where children are generally confronted with and required to memorize more complex information than word or picture pairs. This is true in nearly every aspect of life: especially in school contexts but also in leisure time, for example, when recalling the content of a film or a social event. In two studies Roebers and colleagues (2007) used a short fictional film as to-be-remembered event instead of pair-associates. They showed that children from age 8 on were able to differentiate later correct and incorrect answers as well as different question types in a JOL interview (Roebers et al. 2007). JOL accuracy for the film material was moderate for participants from 8 years on and similar to pair-associate tasks hardly any developmental progression was detected in judgment accuracy with primary school children (Roebers et al. 2007). As clearly more research is needed with other material than the commonly used pair-associates, in the two studies reported here monitoring processes were explored in relation to a fictional children's film rather than to more simple material.

In sum, the literature on the development of children's metacognitive skills shows that although children possess monitoring abilities for both pair-associate tasks and more complex material, these are far from perfect. Absolute levels of discriminability between different kind of answers (e.g. correct and incorrect) tend to be small (Roebers et al. 2007) and gamma correlations indicating metacognitive accuracy are often only moderate (Koriat and Shitzer-Reichert 2002; Koriat et al. 2009).

One major bias occurring in monitoring processes is the fact that individuals tend to overestimate their performance. This phenomenon exists in adults but is even more pronounced in children (Howie and Roebers 2007; Pressley et al. 1987). Flavell (1979) described this overoptimistic view of performance in young children. In some situations overconfidence can be adaptive (Bjorklund et al. 1993; Shin et al. 2007) as it may protect children's self-esteem and encourage them to be persistent in new and difficult tasks. Yet the heavy overestimation of personal abilities especially found in young children can also lead to serious problems. For example, overestimation of physical abilities has been shown to be connected with unintentional injuries (Plummert and Schwebel 1997). Concerning cognitive tasks, overestimation of abilities is disadvantageous in every situation where accuracy of memory is crucial, e.g., in achievement tests or eyewitness situations. Realistic judgments of personal knowledge are especially desirable in academic contexts. If students realize which aspects of to-be-learned materials they have not yet mastered, they can allocate study time accordingly (Metcalfe and Cornell 2005) and thus improve learning results. Monitoring and regulation of learning behavior have not only been shown to improve memory performance (Benjamin and Bjork 1996; W. Schneider and Lockl 2008) but also for example mathematical abilities (Cohors-Fresenbort et al. 2010).

Efforts have been made to identify conditions that help reduce overconfidence in prospective memory judgments in a variety of age groups: preschoolers (Lipko et al. 2008), primary school children (Roebers et al. 2007), middle school children (Lipko et al. 2009) and adults (Anderson and Thiede 2008; Koriat and Bjork 2006; Nelson and Dunlosky 1991). While the adult literature succeeded to detect several important factors influencing JOL accuracy (e.g. Nelson and Dunlosky 1991; Shaughnessy and Zechmeister 1992; Vesonder and Voss 1985), the research on children's JOLs is less clear. For pair-associate tasks a significant improvement of JOL accuracy could be achieved by delaying the JOL interview (Koriat and Shitzer-Reichert 2002; W. Schneider et al. 2000). In contrast, this did not have a positive effect on JOLs for complex material (Roebers et al. 2007). Yet, this finding is in line with the adult literature reporting that JOL timing has no impact on monitoring of complex materials (Dunlosky et al. 2005; Maki 1998), but significantly augmented JOL accuracy for pair-associates (Kimbal and Metcalfe 2003; Nelson and Dunlosky 1991, 1992;). Monitoring accuracy of middle school children could be successfully improved by prompting them to do a pre-judgment recall and compare their answer with a given standard (Lipko et al. 2009). Yet, for primary school age children no strategy to improve JOLs for complex material has been successfully tested, although this could be favorable in a number of daily learning situations. Study 1 was designed as a first step to fill this gap in the literature and to explore whether monitoring performances of primary school children in the context of event recall could be improved with an active reproduction task.

JOLs for pair-associates can either be collected immediately after learning an item or with a time delay which on average lasts several minutes (Keleman 2000; Nelson and Dunlosky 1991). Empirical studies have successfully shown that an average time delay of two minutes (W. Schneider et al. 2000) or slightly more (Koriat and Shitzer-Reichert 2002) was sufficient to boost children's JOL accuracy compared to immediate JOLs. Considering this fact, JOLs in both studies reported here have to be considered delayed because participants watched a whole film before the JOL interview was conducted. Even for questions on the last scenes of the film, the time interval between information presentation and JOL questions exceeded two minutes as instructions of how to give JOLs, training questions and JOLs on other details of the film took more time (Study 2). In Study 1 the time delay was even greater due to the collection of an individual summary before the JOL interview.

Delayed-JOLs are mainly based on retrieval-based cues like retrieval fluency (Benjamin and Bjork 1996; Dunlosky et al. 2005; Koriat and Ma'ayan 2005) or accessibility (Koriat 1997; Nelson and Narens 1990) of the to-be-judged material. This is in accordance with Koriat's cue-utilization view according to which delayed-JOLs are primarily based on mnemonic cues which are subjective feelings like ease of recall (Koriat 1997). Corresponding evidence also exists for children (Koriat and Shitzer-Reichert 2002; Roebers et al. 2007). In Study 1 participants were asked to do an active reproduction of the to-be-remembered material in order to provide them with a salient basis for forming a JOL. Without explicit instruction it is unlikely that all children do a complete pre-judgment recall as a basis for their JOL as this is an effortful task. To ensure that the reproduction task was suitable for children, the design was kept as simple as possible and participants had to monitor only one event. A delayed summary of this event was collected verbally instead of a in written form in order to spare cognitive resources in young children and to allow them to fully concentrate on recall and monitoring processes. The interview was done as a free recall task with unspecific prompts. This procedure should maximize the recall of correct information (Elischberger and Roebers 2001) and at the same time augment the amount of recalled information compared to free recall without prompts (Saywitz and Geiselman 1998). Yet, generally free recall tasks lead to low percentages of recalled information in young children (Cassel et al. 1996; Peterson and Whalen 2001). In sum, the recalled information should provide children with a diagnostic basis for predicting later memory test performance. During the free recall they should experience relevant retrieval-based cues like accessibility of information (or lack thereof) and ease of retrieval. To ensure that children were aware of the quality of their summary a tape-recorded version was played to them after they finished the task, and they were asked to evaluate their performance.

It could be argued that pointing out memory gaps, even if it is done implicitly as in Study 1, could have negative effects on children's self-esteem. Yet, results from existing studies lead to the conclusion that although more realistic monitoring judgments are possible, children remain fairly overconfident of their memory performance despite external interventions (Lipko et al. 2008; 2009). Additionally, adult participants also show persistent overconfidence in prospective monitoring tasks (Anderson and Thiede 2008; Roebers et al. 2007). In sum, it can be assumed that the positive effects of more realistic JOLs on self-regulated learning should outweigh the little, if any, negative effects that giving a summary could have.

Additional evidence that an active reproduction of to-be-remembered contents can improve monitoring accuracy comes from the adult literature. Thiede and colleagues showed that monitoring accuracy for complex materials could be significantly improved if participants gave a delayed summary (Anderson and Thiede 2008; Thiede and Anderson 2003) or created keywords (Thiede et al. 2003) after a time delay of the to-be-remembered contents. Although it must be noted that Thiede and colleagues (Thiede and Anderson 2003; Thiede et al. 2003) included not only text-based information as in Study 1 but also inference questions.

Instead of a global JOL on the whole film, content item-specific JOLs were collected separately for each detail in Study 1. This lead to higher accuracy than global JOLs in an adult sample (Dunlosky et al. 2005) and generated good monitoring results in children as well (Roebers et al. 2007). Additionally, it has been proposed to be an effective way to improve monitoring accuracy for complex materials (Dunlosky and Lipko 2007).

In accordance with the literature on JOLs little age effects for JOLs were expected both for complex and simple encoding material.

In order to consider the effects of an active reproduction on the whole confidence continuum from answers which participants should be very sure of to answers about which they should be very uncertain, non-answerable questions were included in Study 1. This allowed the common comparison of confidence in correct and incorrect answers to answerable questions. Additionally, it was possible to examine if participants adequately showed even lower confidence in answers to non-answerable questions of which they could have absolutely no memory. Non-answerable questions correspond for example to questions on materials learners have not studied at all in an academic context. Furthermore, the inclusion of two question types allowed a direct comparison with existing studies that investigated JOLs in children on complex materials (Roebers et al. 2007).

Additionally to exploring JOLs, confidence judgments (CJs) were collected on all questions answered in the memory interview. This was done immediately after the memory interview was completed. Possibly, a positive effect of giving a summary on JOLs may lead to generally improved monitoring. If a person judged his or her knowledge more realistically before the memory interview, this judgment might be stored and accessed again after the memory interview. Thus, higher CJs accuracy should result.

Study 1

Method

Overview A 2 (age group: 7-year-olds, 9-year-olds) × 2 (experimental condition: summary, no summary) × 2 (question type: answerable, non-answerable) factorial design was utilized in this study, with age and experimental condition being between-subject factors, and question type being a within-subject factor. Dependent measures were accuracy of event recall, JOLs and CJs.

Sample A total of $N=70$ participants (35 female and 35 male) from two age groups took part in the study: $N=35$ 7-year-olds (17 girls) with a mean age of 7 years and 9 months ($SD=4$ months) and $N=35$ 9-year-olds (18 girls) with a mean age of 9 years and 11 months ($SD=6$ months). In both age groups, half of the participants were randomly assigned to each experimental condition (summary vs. no summary) with the only constraint that male and female participants were equally distributed across conditions. Children were recruited from six primary schools in Switzerland near a town of about 130.000 inhabitants in order to ensure that sufficient children participated in the study. They came primarily from middle class families. Written consent was obtained from parents prior to the study.

Procedure In the first phase of the study all participants watched a seven minute video in small groups. The video depicted a group of children who visited a farmer's family. During their stay, the children learned that a castle once stood on the farm but was destroyed by a fire. They searched for the ruins of the castle, found the former entrance to the dungeons, climbed down into underground tunnels and in the end discovered a treasure. Participants were ignorant of the following memory interview.

On average approximately 70 min after watching the video half of the participants were individually asked to give a verbal summary of the film content (experimental condition: summary). This was done as a Free Recall with unspecific prompts ("What else has happened in the film?, ""Do you remember anything else?"). Summaries were recorded on a tape and afterwards played to the participants. To initiate metacognitive monitoring processes children were asked to evaluate their performance, for example "Do you think that other children know what happened in the film when they listen to your summary? "The other half of participants only watched the film without giving a summary or answering metacognitive questions (experimental condition: no summary).

Next, all participants took part in a JOL-interview including 24 questions appropriate for children about the film content (see Appendix). They judged how sure they were that they would remember a specific detail from the film in 2 weeks on a seven-point scale (Fig. 1). These details were identical to those targeted in the later memory interview (see below). For the JOLs, the interviewer introduced the 7-point scale, indicated what each point on the scale meant and gave examples for each category. Children were then asked to give JOL-ratings for a set of training questions that required judgment and that were unrelated to the recall questions, e.g., *"How sure are you that you will remember in 2 weeks that you watched a video today?" and "How sure are you that you will remember in 2 weeks what kind of clothes the other kids in your class wore today?"* Children who did not give an appropriate JOL were corrected, given a rationale for the appropriate rating and asked another training question. The 24 JOL questions about the video details were asked only after three appropriate training JOLs had been given. Similar training and practice for scale use had been successfully utilized in previous studies (Roebers 2002; Roebers and Howie 2003; Roebers et al. 2007). Participating children learned the use of the scale quickly and with ease. For the JOL interview itself, participants were prompted to remember the video and asked to indicate for each of the details in question, how sure they were that they were able to correctly recall this specific aspect of the video in 2 weeks. For example, the interviewer asked *"How sure are you that you will remember how the boys got to the farmhouse?"* or, "*How sure are you that you will remember who found the entrance to the tunnel?"* and the participant then gave his or her metacognitive judgment using the 7-point scale. Participants were retained from recalling the to-be-remembered information openly.

Two weeks after the video presentation and the JOL interview, participants in both experimental conditions were individually questioned about the contents of the video. Participants were instructed as follows: *"I want you to tell me as much as you can about what was in the film. But I also want you to try the best you can to give me only correct answers."* The memory interview contained 24 questions that matched exactly the questions of the JOL interview 2 weeks earlier. As was done with the JOLs, answers to the memory questions were recorded on standard protocol sheets for later coding. Correct answers to the answerable questions consisted mostly of one or several words and were only scored as correct if they exactly mirrored the expected answer to minimize the possibility of guessing. For the non-answerable questions, only "I don't know" answers were scored as appropriate and included in the analyses. After asking all the recall questions, the interviewer introduced the CJ task. To minimize socially desirable CJs, it was explained that some of the questions were easy, some hard and some in between. At that point, the interviewer displayed the seven-point scale (identical to that used for the JOL-interview, see Fig. 1) and asked the participant to show how sure she or he was about the correctness of each answer. Children were also asked a minimum of three training questions

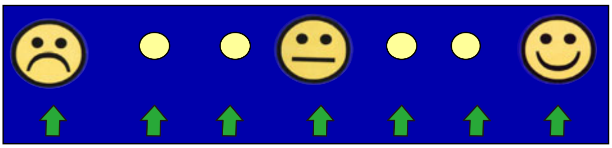

Fig. 1 Seven-point scale used to measure JOLs and CJs. (Note: Higher values indicate higher confidence)

about facts that were unrelated to the event to be recalled and that reflected a range of confidence levels, e.g., "*How old are you*", *"How old am I"* and *"What are you going to do this afternoon?"* After each answer, children were asked *"and how sure are you about that?"* and asked to point to the smiley or dot that best matched their degree of confidence. If a child did not indicate an appropriate level of confidence for a question, she or he was corrected, given a rationale for the appropriate rating and asked another training question. Participants generally learned the use of the confidence scale with ease; nevertheless, the procedures were only started when the child had given 3 appropriate training CJs. If a participant answered with "I don't know" during the interview, no CJ was requested.

When the questioning was finished, participants were praised for their good performance, given a small gift, and thanked for their help.

Materials The memory test consisted of 24 target questions, of which 14 were answerable and 10 were non-answerable. All questions were asked in an unbiased question format (for example, "Where did the boys spent the night?" or "What were the girls carrying when they got off the bus?"), were successfully used in studies with primary school children (Roebers 2002; Roebers et al. 2007) and were of moderate difficulty based on data of previous studies for the age groups included in the Study 1 (Roebers and Howie 2003; Roebers et al. 2007).

Results

First, JOLs will be presented as a function of age, summary and correctness, for both answerable and non-answerable questions. Then, the metacognitive predictions (JOLs) will be related to the corresponding memory performance. Third, memory performance will be presented in terms of correct answers to the answerable questions as a function of age, as well as presence vs. absence of a summary of the film content. Next, frequencies of appropriate "don't know" responses to the non-answerable questions will be examined as a function of summary vs. no summary. Finally, accuracy of the postdictions (CJs) will be reported. Preliminary analysis assessing the effect of gender did not reveal any systematic differences between male and female participants. Thus, data was collapsed across this variable. As a post-hoc follow-up on main effects, Student-Newman-Keuls tests were used. A probability level of $p<.05$ was used as the criterion for statistical analysis.

Judgments-of-Learning before answerable questions Table 1 presents the mean JOLs before answerable questions as a function of correctness of later recall, experimental condition and age.

Table 1 Mean JOLs before answerable questions as a function of correctness of later recall, experimental condition, and age (standard deviations in parenthesis)

	Correct answer	Incorrect answer
Summary		
7-year-olds	5.1 (1.1)	3.1 (1.6)
9-year-olds	5.8 (1.0)	3.8 (1.4)
Control group		
7-year-olds	5.2 (1.5)	3.4 (2.1)
9-year-olds	5.9 (.8)	4.2 (1.8)

An ANOVA with correctness of answer as inner-subject factor, and experimental condition and age as between-subject factors revealed a significant main effect of correctness of answers, $F(1, 58)=82.42$, $\eta^2=.59$, and a significant main effect of age, $F(1, 58)=4.87$, $\eta^2=.08$. More confident JOLs were given before correct answers [5.5] than before incorrect answers [3.6], and 9-year-olds [4.9] gave more confident JOLs than 7-year-olds [4.2].

Judgments-of-Learning before non-answerable questions Table 2 presents the mean JOLs before incorrect and "don't know" answers to non-answerable questions as a function of later recall, experimental condition, and age.

An ANOVA with appropriateness of answer as inner-subject factor, and experimental condition and age as between-subject factors revealed a significant main effect of appropriateness of answer, $F(1, 60)=52.52$, $\eta^2=.47$. Before incorrect answers [3.2] more confident JOLs were given than before "don't know" answers [2.2]. No other main effects or interactions were significant.

An additional analysis was performed in order to explore differences between JOLs before incorrect answers to answerable questions and JOLs before invented answers to non-answerable questions. ANOVAs with question type (answerable vs. non-answerable) as a within-subject factor, and age and experimental condition as between-subject factors was conducted. No significant effects or interactions were found.

Judgment-of-Learning accuracy in relation to answerable questions To estimate JOLs accuracy, gamma correlations were computed between JOLs before answerable questions and recall for answerable questions. Mean gamma correlations in the summary group were .79 (*SD*=.21) for 7-year-olds and .73 (*SD*=.31) for 9-year-olds. In the control group mean gamma correlations were .79 (*SD*=.21) for 7-year-olds and .81 (*SD*=.24) for 9-year-olds. One-sided t-tests revealed that all correlations differed significantly from zero. An ANOVA with age and experimental condition as between-subject factors revealed no significant main effects or interactions.

Recall: answerable questions Table 3 presents the mean percentages of correct answers to answerable questions as a function of age and experimental condition.

An ANOVA with experimental condition and age as between-subject factors and percent of correct answers as dependent variable was conducted. It revealed a significant main effect of age, $F(1, 65)=12.23$, $\eta^2=.16$. 7-year-olds [50.8%] gave fewer correct answers than 9-year-olds [60.8%].

Table 2 Mean JOLs before non-answerable questions as a function of correctness of later recall, experimental condition, and age (standard deviations in parenthesis)

	Incorrect answers	"Don't know" answers
Summary		
7-year-olds	3.3 (1.8)	1.9 (.7)
9-year-olds	3.3 (1.1)	2.0 (.9)
Control group		
7-year-olds	3.5 (1.9)	2.6 (1.6)
9-year-olds	2.9 (1.3)	2.2 (1.3)

Table 3 Mean percentages of correct answers to answerable questions as a function of age and experimental condition (standard deviations in parenthesis)

	7-year-olds	9-year-olds
Summary	51.6 (10.0)	62.3 (12.0)
Control group	50.0 (14.5)	59.2 (10.9)

Non-answerable questions Table 4 presents the mean percentages of "don't know" answers to non-answerable questions as a function of age and experimental condition.

An ANOVA with experimental condition and age as between-subject factors and the percentage of "don't know" answers as dependent variable was conducted. It revealed a significant main effect of age, $F(1, 66)=6.05$, $\eta^2=08$. 7-year-olds [75.3%] gave more "don't know" answers than 9-year-olds [64.7%].

Confidence Judgment accuracy in relation to answerable questions To estimate CJ accuracy, gamma correlations were computed between CJs for answerable questions and recall for answerable questions. Mean gamma correlations in the summary group were $M=.69$ ($SD=.67$) for 7-year-olds and $M=.77$ ($SD=.32$) for 9-year-olds. In the control group, mean gamma correlations were $M=.57$ ($SD=.69$) for 7-year-olds and $M=.77$ ($SD=.32$) for 9-year-olds. One-sided t-tests revealed that all correlations differed significantly from zero. An ANOVA with age and experimental condition as between-subject factors revealed no significant main effects or interactions.

Discussion

First, the present study replicated previous findings concerning monitoring performances in the context of a complex recall task over the primary school years. For one, no age differences were found in discriminability of JOLs or in gamma correlations as indicators of JOL accuracy. This finding is consistent with former research on both JOLs collected in event recall (Roebers et al. 2007) and in pair-associates tasks (Koriat and Shitzer-Reichert 2002; Lockl and W. Schneider 2003). Second, levels of performance were also very similar to those reported in studies using complex materials: Participants showed adequate yet still extendable monitoring activities. JOL accuracy was moderate (Dunlosky et al. 2005; Keleman 2000; Roebers et al. 2007). Participants differentiated successfully between later correct and incorrect answers to answerable questions and incorrect and "don't know" answers to non-answerable questions. Only discrimination of answerable and non-answerable questions just failed significance. In sum, distinctions of different certainty levels were done successfully but with small absolute differences (cf. Roebers et al. 2007)

Table 4 Mean percentages of "don't know" answers to non-answerable questions as a function of age and experimental condition (standard deviations in parenthesis)

	7-year-olds	9-year-olds
Summary	78.9 (14.1)	61.1 (17.1)
Control group	71.7 (21.6)	68.2 (19.1)

Overall, children from 7 years of age on showed adequate performance in all collected measures and results corresponded with former findings, which speaks for the reliability of our data.

The major aim of this study was to analyze a method to improve monitoring performances in primary school children by giving a verbal summary before the JOL interview. The recall process was supposed to reduce overconfidence by forming a realistic information base for monitoring processes and offering retrieval-based cues (Benjamin and Bjork 1996; Koriat 1997; Nelson and Narens 1990). The design was also influenced by studies with adults which had demonstrated that monitoring accuracy could be significantly augmented by a delayed summary or keyword generation of contents which were relevant for the memory test (Thiede and Anderson 2003; Thiede et al. 2003; Thiede et al. 2005). In our study this task did not lead to comparable positive effects. Giving a summary of the film content influenced neither differentiation of later correctness of answers nor differentiation of question types. The same was true for accuracy of JOLs.

It was speculated that giving a summary might improve monitoring process by creating enduring realistic assessment of personal knowledge and reducing overconfidence. If this was the case, CJs collected after the memory interview should also be positively influenced. Yet, as could be expected with no effect on JOL accuracy, the summary group could not benefit in their postdictions of memory processes.

One possible reason for the missing effect could be the extended forgetting interval in Study 1. In the present study the memory test took place 2 weeks after the JOL interview, while most other existing studies used considerably shorter forgetting intervals generally lasting only minutes. However, this thesis appears unlikely considering the fact that other findings concerning JOL performances could be documented for short and longer periods of forgetting. This is, for example, the case for developmental progression (or better the lack of it) and the improvement of memory performance after a JOL interview (Koriat et al. 2006; Lockl and W. Schneider 2003; Roebers et al. 2007).

Another cause for the lack of effect of the summary task could be the very good monitoring performances found in Study 1. Monitoring accuracy with gammas between .70 and .80 in both experimental conditions are nearly as good as the most accurate gamma correlations ever reported for pair-associates in children (W. Schneider et al. 2000) and adults (Dunlosky and Nelson 1992). In the adult literature, metacomprehension accuracy for texts hardly ever exceeded .70, even under favourable conditions (Thiede and Anderson 2003; Thiede et al. 2003). These good predictive accuracies make it difficult to detect further improvements as generally much more effort is necessary to improve results at the upper end of the performance range than middle or low performances. Maybe a more significant change in the experimental design is necessary for children to improve their monitoring accuracy.

One major change in the experimental design and at the same time a third reason why giving a summary might have had no effect in Study 1 is the encoding condition. In Study 1 encoding was non-intentional, while it is more common in the literature on JOLs to do encoding intentionally (e.g. Dunlosky et al. 2005; Koriat and Bjork 2006; W. Schneider et al. 2000; Thiede et al. 2005). Maybe a conscious decision or active concentration is necessary to achieve better monitoring. Although both kinds of encoding are important in everyday life, under non-intentional encoding conditions individuals might not be able to make the most of their monitoring abilities. This might have been the case in Study 1 where participants were not aware of the following memory test. Thus, one possibility for improving monitoring performance could be to give an explicit warning that a memory interview will follow, that is, to induce an intentional encoding situation. To explore this possibility in Study 2, intentional and non-intentional encoding were contrasted and thus another means to improve monitoring accuracy in children was investigated.

If participants are aware of the subsequent memory interview, they might be able to monitor information more specifically during encoding (Mazzoni and Nelson 1995). Generally, intentional encoding should lead people to monitor their degree of learning (Nelson and Narens 1990, 1994), which is not a necessary activity in non-intentional encoding situations. This should on the one hand result in better metacognitive knowledge of the learning process and thus more accurate monitoring judgments. Findings with adult participants hint to the fact that intentional encoding leads to higher item-by-item JOL accuracy than non-intentional encoding (Mazzoni and Nelson 1995). On the other hand, realizing which content is hard and which one is easy to remember already during encoding might enable participants to use specific strategies to memorize the more difficult pieces of information: for example, they could use "on-line"-rehearsal processes during encoding. This might in turn foster accuracy of monitoring because participants thus spend more time concentrating on the difficulty of different learning materials. Additionally, it might improve memory performance and lead to a greater correspondence with monitoring judgments, in which participants generally overestimate their subjective memory performance (Howie and Roebers 2007).

As a consequence of the suggested processes, age effects in metacognitive monitoring might occur. Older children and adults generally use strategies more efficiently than very young children, especially for rehearsal strategies age effects could be shown during the primary school years (Kunzinger 1985). Metacognitive control processes like on-line strategies also improve consistently with age (Dufresne and Kobasigawa 1989; Lockl and W. Schneider 2004). Thus, older children and adults might profit more from the possibility to actively monitor the encoding process than younger children. In order to explore possible age effects, four different age groups from 6-year-olds to adults were included in Study 2. At the same time, this extended the age range explored in Study 1 in order to obtain more general results.

Besides being a possibility to improve JOL accuracy, a direct comparison of monitoring processes after non-intentional and intentional encoding in one study is also of theoretical interest. In the existing literature children's JOLs have either been studied with simple material and intentional encoding (Koriat and Shitzer-Reichert 2002; Koriat et al. 2009) or complex material and non-intentional encoding (Roebers et al. 2007). Additionally, studies differ concerning other variables like forgetting interval. Therefore, Study 2 allows for the first time a direct comparison of the effects of encoding condition on monitoring competencies within one study and thus without any confounding variables. This systematically extends existing studies and will help to clarify the question whether encoding condition is a critical factor for children's monitoring processes or not. Findings with adults suggested that encoding condition had significant effects on metacognitive monitoring quality (Mazzoni and Nelson 1995). In contrast, studies with children revealed that for both intentional and non-intentional encoding hardly any age effects and adequate performances can be detected. This might suggest that results of both encoding conditions are comparable.

To assess a broad range of monitoring activities and in order to allow comparisons with Study 1 and other former findings non-answerable questions and CJs were included in Study 2.

Study 2

Method

Overview A 4 (age group: 6-year-olds, 8-year-olds, 10-year-olds, adults) × 2 (experimental design: intentional, non-intentional encoding) × 2 (question type: answerable, non-answerable) design was utilized in this study, with age and experimental condition being between-subject factors, and question type being a within-subject factor. Dependent measures were accuracy of event recall, JOLs and CJs.

Sample A total of $N=200$ participants from four age groups took part in the study: $N=50$ 6-year-olds (25 girls) with a mean age of 6 years and 2 months ($SD=4$ months), 50 8-year-olds (26 girls) with a mean age of 8 years and 0 months ($SD=4$ months), $N=50$ 10-year-olds (25 girls) with a mean age of 10 years and 2 months ($SD=4$ months) and $N=50$ adults (20 women) with a mean age of 20 years ($SD=4$ years). In all age groups, half of the participants were randomly assigned to each experimental condition (non-intentional vs. intentional encoding) with the only constraint that male and female participants were equally distributed across conditions. Children were recruited from four kindergartens and three primary schools near a town of approximately 130.000 inhabitants in Southern Germany and primarily came from middle class families. Adults were in the last year of a vocational training and university students with different majors (psychology, education, business studies). Written consent was obtained from children's parents, and adults gave oral consent prior to the study.

Procedure In the first phase of the study all participants watched the same video as in Study 1 in small groups. Before the video started, half of the participants were told that the experimenter would later be interested in their opinion of it. They were asked to judge the suitability of the video for use in another study with children their age in the children group and with primary school children in the adult group. After the presentation, the group was briefly asked for their opinion of the video and thanked for their helpful comments. In this experimental condition (non-intentional encoding), the upcoming memory test was not mentioned.

In the intentional encoding condition, the other half of the participants were told to pay close attention to the film because a memory test would take place in 2 weeks time. It was stressed that it was very important to remember as much as possible about the film content and to remember everything correctly.

Participants in both experimental conditions were asked individually to give JOLs concerning 14 details of the video immediately after watching the video. The 7-point JOL scale and the training questions for the JOL interview as well as the interview procedure were identical to Study 1.

Two weeks after the video presentation and the JOL interview, participants in both experimental conditions were individually questioned about the contents of the video. Again, instruction, scale and procedure of the memory interview were identical to Study 1. When the questioning was finished, participants were praised for their good performance, given a small gift (children) or a monetary award (adults), and thanked for their help.

Materials The memory interview consisted of 14 questions, of which 8 were answerable and 6 were non-answerable (see Appendix). Questions were successfully utilized in former studies with primary school children (Roebers and Howie 2003; Roebers et al. 2007).

Results

First, JOLs will be presented as a function of age, encoding condition and correctness, for both answerable and non-answerable questions. Then, the metacognitive predictions (JOLs) will be related to the corresponding memory performance. Third, memory performance will be presented in terms of correct answers to the answerable questions as a function of age, and encoding condition. Then, frequencies of appropriate “don’t know” responses to the non-answerable questions will be examined as a function of encoding condition. Finally, accuracy of CJs will be reported. Preliminary analysis assessing the effect of gender did not reveal any systematic differences between male and female participants. Thus, data was collapsed across this variable. As a post-hoc follow-up on main effects, Student-Newman Keuls tests were used. A probability level of $p<.05$ was used as the criterion for statistical analysis.

Judgments-of-Learning before answerable questions Table 5 presents the mean JOLs before answerable questions as a function of age, correctness of later recall, and experimental condition. An ANOVA with correctness of answers as inner-subject factor and experimental condition and age as between-subject factors revealed a significant main effect of correctness of answer, $F(1, 164)=62.57$, $\eta^2=.28$, a significant main effect of experimental condition, $F(1, 164)=4.70$, $\eta^2=.03$, a significant interaction between experimental condition and age, $F(3, 164)=3.10$, $\eta^2=.05$, as well as a significant interaction between correctness of answer and age, $F(3, 164)=3.06$, $\eta^2=.05$. JOLs before correct answers [5.5] were more confident than before incorrect answers [4.3], independent of age and experimental condition. The interaction between correctness of answer and age was caused by the fact that JOLs before incorrect answers differed significantly between the age groups (Adults: 3.6< 8-year-olds: 4.8; 10-year-olds: 4.2 and 6-year-olds: 4.5 fell in between and differed neither from each other nor from 8-year-olds or adults), while for JOLs before correct answers there was no age effect, $F(3, 196)=.65$, n.s.. The interaction between experimental condition and age group was caused by the fact that for 6-year-olds JOLs after intentional and non-intentional encoding differed significantly [intentional: 3.6 < non-intentional: 5.4, $F(1, 43)=10.02$] while this was not the case for the other three age groups [intentional vs. non-intentional: 8-year-olds: 4.7 vs. 5.0, $F(1, 41)=.46$, n.s.; 10-year-olds: 4.3 vs. 4.2, $F(1, 40)=.03$, n.s.; adults: 3.2 vs. 4.0, $F(1, 40)=1.95$, n.s.]

Table 5 Mean JOLs before answerable questions as a function of age, correctness of later recall, and experimental condition (standard deviations in parenthesis)

		Correct answer	Incorrect answer
Intentional encoding	6-year-olds	4.8 (1.9)	3.6 (2.1)
	8-year-olds	5.6 (1.4)	4.7 (1.8)
	10-year-olds	5.6 (1.1)	4.3 (1.7)
	Adults	5.9 (1.2)	3.2 (1.9)
Non-intentional encoding	6-year-olds	5.8 (1.5)	5.4 (1.9)
	8-year-olds	5.7 (1.3)	5.0 (1.9)
	10-year-olds	5.5 (1.1)	4.2 (2.0)
	Adults	5.5 (1.3)	4.0 (1.8)

Judgments-of-Learning before non-answerable questions Table 6 presents the mean JOLs before non-answerable questions as a function of correctness of later recall, experimental condition, and age. An ANOVA with appropriateness of answer as inner-subject factor and experimental condition and age as between-subject factors revealed a main effect of age, $F(3, 148)=3.87$, $\eta^2=.07$, and a main effect of appropriateness of answer, $F(1, 148)=90.65$, $\eta^2=.38$. More confident JOLs were given for incorrect answers [3.4] than for "don't know" answers [2.0]. Adults [2.2] gave less confident JOLs than 6-year-olds [2.9], 8-year-olds [3.0] and 10-year-olds [2.8].

An additional analysis was performed in order to explore differences between JOLs before incorrect answers to answerable questions and JOLs for invented answers to non-answerable questions. An ANOVA with question type (answerable vs. non-answerable) as a within-subject factor, and age and experimental condition as between-subject factors were conducted. It revealed significant main effects of question type, $F(3, 135)=24.24$, $\eta^2=.15$, age, $F(3, 135)=4.16$, $\eta^2=.09$, and experimental condition, $F(1, 135)=5.19$, $\eta^2=.04$. Before incorrect answers to answerable questions more confident JOLs [4.3] were given than before incorrect answers to non-answerable questions [3.4]. Adults [3.2] gave less confident JOLs than 8-year-olds [4.4], 6-year-olds [4.0] and 10-year-olds [3.6] fell in between and differed significantly from neither age group. In the intentional encoding situation [3.6] lower JOLs were given than in the non-intentional condition [4.1].

Judgment-of-Learning accuracy in relation to answerable questions To estimate JOLs accuracy, gamma correlations were computed between JOLs for answerable questions and recall for answerable questions. Mean gamma correlations in the intentional encoding situation were .55 (SD=.54), .49 (SD=.61), .79 (SD=.61), .61 (SD=.38) in ascending age order. In the non-intentional encoding situations gamma correlations were .63 (SD=.42), .53 (SD=.54), .62 (SD=.45) and .60 (SD=.61) in ascending age order. One-sided t-tests revealed that all correlations differed significantly from zero. An ANOVA with age and experimental condition as between-subject factors revealed no significant main effects or interactions.

Recall: answerable questions Table 7 presents the mean percentages of correct answers to answerable questions as a function of age and experimental condition.

An ANOVA with experimental condition and age as between-subject factors and percent of correct answers as dependent variable was conducted. It revealed a significant main effect of age, $F(3,192)=15.15$, $\eta^2=.19$. 6-year-olds [38.5%] gave fewer correct answers than 8-year-olds [54.8%], 10-year-olds [62.0%] and adults [57.0%].

Table 6 Mean JOLs before non-answerable questions as a function of correctness of later recall, experimental condition, and age (standard deviations in parenthesis)

		Incorrect answer	"Don't know" answer
Intentional encoding	6-year-olds	3.2 (1.8)	2.1 (1.3)
	8-year-olds	4.2 (2.1)	2.1 (1.1)
	10-year-olds	3.0 (1.6)	2.1 (.9)
	adults	2.5 (1.8)	1.5 (.7)
Non-intentional encoding	6-year-olds	3.7 (1.9)	2.5 (1.3)
	8-year-olds	3.7 (1.8)	2.0 (1.2)
	10-year-olds	3.8 (1.8)	2.3 (1.1)
	adults	3.0 (1.9)	1.6 (.6)

Table 7 Mean percentages of correct answers to answerable questions as a function of age and experimental condition (standard deviations in parenthesis)

	6-year-olds	8-year-olds	10-year-olds	Adults
Intentional encoding	37.0 (16.7)	54.0 (18.7)	63.5 (18.7)	55.5 (18.4)
Non-intentional encoding	40.0 (14.4)	55.5 (24.0)	60.5 (14.7)	58.5 (20.3)

Non-answerable questions Table 8 presents the mean percentages of "don't know" answers to non-answerable questions as a function of experimental condition and age.

An ANOVA with experimental condition and age as between-subject factors and the percentage of "don't know" answers as dependent variable was conducted. It revealed a significant main effect of age, $F(3, 192)=5.59$, $\eta^2=.08$. 6-year-olds [74.7%] and adults [73.3%] answered more often with "don't know" than 8-year-olds [57.7%] and 10-year-olds [60.7%].

Confidence Judgment accuracy in relation to answerable questions To estimate CJ accuracy, gamma correlations were computed between CJs for answerable questions and recall for answerable questions. Mean gamma correlations in the non-intentional encoding group were .57 (*SD*=.67), .33 (*SD*=.85), .81 (*SD*=.24) and .62 (*SD*=.67) in ascending age order. In the intentional encoding group mean gamma correlations were .69 (*SD*=.57), .54 (*SD*=.73), .79 (*SD*=.51) and .79 (*SD*=.48) in ascending age order. One-sided t-tests revealed that all correlations differed significantly from zero. An ANOVA with age and experimental condition as between-subject factors revealed no significant main effects or interactions.

Discussion of Study 2 and general discussion

A first research question in Study 2 concerned possible age effects. In accordance with Study 1 and literature findings, Study 2 did not reveal any age effects in JOL accuracy or the ability to discriminate different confidence levels. This speaks for adequate but not perfect monitoring competencies of very young children and over a broad age range. This finding was shown for different encoding materials, encoding conditions and different context variations (Study 1; Koriat and Shitzer-Reichert 2003; Roebers et al. 2007; W. Schneider et al. 2000).

Yet, the absolute levels of JOLs were lower for adults than for younger children before incorrect answers for answerable questions and JOLs before non-answerable questions. It was assumed that age effects might result from the influence of encoding condition. Yet, the age effects did not occur specifically in the intentional encoding condition and thus cannot be solely explained by the fact that older participants in this experimental condition made better use of strategies and thereby improved their monitoring. It appears more likely that

Table 8 Mean percentages of "don't know" answers to non-answerable questions as a function of age and experimental condition (standard deviations in parenthesis)

	6-year-olds	8-year-olds	10-year-olds	Adults
Intentional encoding	75.3 (26.8)	56.0 (30.0)	61.3 (23.9)	76.1 (25.6)
Non-intentional encoding	74.0 (27.2)	59.3 (27.3)	60.0 (25.0)	76.0 (20.5)

the youngest age group (6-year-olds), which was only included in Study 2, overestimated their memory performance to a greater extent than primary school children and adults. This is in correspondence with the fact that the overconfidence effect is more pronounced in young children than in adults and can explain why findings in Study 1 and 2 differ from each other in this regard.

The main focus of Study 2 was on the effect of encoding condition on metacognitive monitoring processes and a potential positive effect of intentional encoding compared to non-intentional encoding. In sum, no different effects of non-intentional and intentional encoding on JOL accuracy and discrimination could be detected but an effect on absolute levels of JOLs was found. For one, 6-year-olds gave lower JOLs before answerable questions in the intentional encoding condition than in the non-intentional encoding condition. Second, JOLs before incorrect answers to both answerable and non-answerable questions were lower in the intentional than in the non-intentional condition. In these cases, participants in the intentional encoding condition were less confident of remembering a certain detail than participants in the non-intentional encoding condition. To some extent a more conscious encoding and monitoring process improved monitoring performances in the desired direction by reducing unjustified overconfidence. The positive effect seems to affect areas were overconfidence is most pronounced: in the youngest age group and incorrect answers. Yet, even then participants hardly used the lower end of the confidence scale with a medium JOL of 3.59 for incorrect answers. This reflects the fact that the effect of encoding condition was rather small and not very systematic. Most importantly, differentiation and accuracy of JOLs as well as accuracy of CJs were not affected by the encoding condition. The improvement was not pronounced enough to influence the important measures of metacognitive monitoring processes. This is in contrast to prior studies in which intentional encoding improved JOL accuracy in adults (Mazzoni and Nelson 1995). However, it is in line with comparable effects on performance level in different age groups reported for monitoring processes in different studies with either intentional or non-intentional encoding condition (Koriat and Shitzer-Reichert 2002; Lockl and W. Schneider 2003; Roebers et al. 2007). Considering the results of Study 2 in sum, encoding condition does not seem to be an important factor for children's metacognitive monitoring.

One explanation could be that participants suspected the importance of paying attention to the film even in the non-intentional condition. However, observed comments and behaviour of participants in non-intentional encoding conditions in both Studies 1 and 2 lead to the conclusion that participants believed that their main task was to rate the suitability of the film for young children. Additionally, the differences between both encoding conditions, albeit small, speak against this assumption as well.

Presumably, people are able to monitor everyday information they encode non-intentionally quite well. This makes sense considering the fact that in our everyday life, we frequently recall information that was encoded un-intentionally. This is the fact in a broad range of contexts from leisure situations like listening to the radio or watching television to eyewitness situations. Empirically, the assumption that everyday life trains non-intentional encoding abilities is supported by the fact that individuals achieved quite high gamma correlations (around .75) for non-intentional listening to the news (S. Schneider and Laurion 1993). This task is very similar to the one used in Studies 1 and 2 where participants had to recall facts of a film presentation. Additionally, every day individuals refer to events experienced in their past which were not consciously memorized for later retrieval. Thus, individuals consistently have to monitor their memories. This is true for all age groups but especially for young children where personal experiences form most of the information in memory and intentional learning situations like those experienced later in school are rare.

Non-intentional and intentional encoding do not seem to have a great influence on metacognitive monitoring processes in children and thus are not likely to be the reason for the lack of monitoring improvement found in Study 1. Another aspect that might have been important here is the timing of the JOL interview which has been shown to be the most influential factor for JOLs in pair-associate paradigms. As explained above, all JOLs collected in Study 1 are to be considered as delayed because the time delay between information presentation and JOL collection exceeded two minutes for all items (Koriat and Shitzer-Reichert 2002; W. Schneider et al. 2000). Yet, in Study 1 the average time delay due to giving a delayed summary was far greater than two minutes. This might have boosted JOL accuracy even further and overshadowed a possible effect of giving a summary. This possibility cannot be ruled out completely, yet several facts let it appear unlikely: First, JOL accuracy for film material collected immediately after completion of the film was found to be as accurate as in Study 1 (Roebers et al. 2007) or only slightly less accurate (cf. Study 2). Secondly, JOL accuracy in Study 1 did not differ significantly for participants who completed the JOL interview immediately after the film compared to greater time delays. Third, no differences in accuracy could be detected between JOLs on film material collected immediately after the film and after a 24-hour-delay (Roebers et al. 2007). Additionally, in the adult literature JOL timing also had no impact on monitoring of complex materials (Dunlosky et al. 2005; Maki 1998). In sum, the existing findings suggest that timing of the JOL interview does not seem to have a great effect on JOLs collected after watching a complete film. Yet, it would be interesting to further explore the influence of immediate and delayed JOLs on complex material within one study. Immediate JOLs could be realized by stopping the film on several occasions and collect the JOLs on-line compared to delayed JOLs collected after watching the whole film. Although for immediate JOLs a summary condition would difficult to realize, it would be possible to contrast JOLs collected directly after the film and JOLs collected after a time delay, both with and without a summary. This would also help to distinguish unambiguously whether JOL timing, summarizing, both factors or none had an effect on JOL accuracy in Study 1. Equally, further research is necessary on other factors that could improve monitoring processes in children.

At present it appears that monitoring processes are complex activities which are generally hard to influence - at least in children. This is possibly why children cannot benefit from short or one-time interventions, such as the modeling of monitoring processes (McGivern et al. 1990; von der Linden and Roebers 2007) or giving a summary of the to-be-remembered event (Study 1). In the adult literature two decades of research were necessary to reveal effective possibilities to improve monitoring accuracy for complex material. Still not many successful interventions have been identified to date, which also speaks for the complexity of prospective monitoring activities (Thiede and Anderson 2003; Thiede et al. 2003, 2005). Unfortunately, they were not effective in children (Study 1). So possibly longer and more intensive trainings might be necessary to improve monitoring performances. Monitoring training and feedback over several months improved monitoring performance of students significantly (Hacker et al. 2000; Nietfield et al. 2006). Training of monitoring strategy selection in children was only successful when the guidelines for selection are detailed (Ghatala et al. 1986). More intensive training programs might also be the solution for improving JOLs and maybe CJs, which were tested in Studies 1 and 2.

Considering the results of Study 2 from a more theoretical point of view, it was the first time that JOLs under intentional and non-intentional encoding conditions were compared in one study. As little differences could be detected between both conditions, children's monitoring competencies appear to be robust in different learning settings. They generalize not only to different encoding conditions (Study 2) but also to different kinds of encoding material (Lockl and W. Schneider; Roebers et al. 2007) and to longer and shorter prediction intervals (Koriat and Shitzer-Reichert 2002; Roebers et al. 2007). Therefore, prospective metacognitive monitoring abilities appear to be a general competence which appears early in life and needs to be developed.

In sum, in the two studies reported here adequate monitoring performances, especially concerning accuracy, for children as young as 6 years old could be demonstrated. Yet, the attempts to further improve these competencies were not successful. From a practical point of view it is important to point out that even before children reach school age they are able to monitor what they learn, at least to some extent. On the other hand, the challenge to help them use these abilities more effectively remains. Metacognitive monitoring is a critical component in every self-regulated learning process and good monitoring corresponds with effective self-regulation and improved performance (Thiede et al. 2003). Additionally, it is relevant in many every-day life situations where encoding is less deliberate yet not less important, such as listening to conversations, watching television, observing an eye-witness event. Therefore, further research is needed to identify the factors that influence prospective monitoring in children including timing of JOL collection and methods to reduce overconfidence.

Appendix

JOL questions before answerable questions

1) How sure are you that you can remember in 2 weeks how the boys arrived at the farm house?
2) How sure are you that you can remember in 2 weeks what each girl carried in her hand when the girls got off the bus?
3) How sure are you that you can remember in 2 weeks the name of the dog?
4) How sure are you that you can remember in 2 weeks which other animals than the dog were in the video?
5) How sure are you that you can remember in 2 weeks who came into the kitchen while the children were sitting there and eating a meal?
6) How sure are you that you can remember in 2 weeks where the boys were supposed to spend the night?
7) How sure are you that you can remember in 2 weeks how the castle was destroyed?
8) How sure are you that you can remember in 2 weeks what kind of shoes the children had on when they walked across the field?
9) How sure are you that you can remember in 2 weeks who found the entrance to the tunnel?
10) How sure are you that you can remember in 2 weeks which clue to the treasure the children found at the entrance of the tunnel?
11) How sure are you that you can remember in 2 weeks who went down into the tunnel first?
12) How sure are you that you can remember in 2 weeks what the children used to see in the dark?
13) How sure are you that you can remember in 2 weeks what kind of room the tunnel lead to?
14) How sure are you that you can remember in 2 weeks who helped the children to get out of the tunnel?

JOL-questions before non-answerable questions

1) How sure are you that you can remember in 2 weeks why the children went to the farmhouse on that particular day?
2) How sure are you that you can remember in 2 weeks what the twins just finished doing when the children arrived at the farm house?
3) How sure are you that you can remember in 2 weeks who made up the twins nickname?
4) How sure are you that you can remember in 2 weeks how many rolls were on the table when the children were having their lunch?
5) How sure are you that you can remember in 2 weeks what the twins did during their last holiday?
6) How sure are you that you can remember in 2 weeks where the children got the spades from?
7) How sure are you that you can remember in 2 weeks which day of the week it was when the children got down into the tunnel?
8) How sure are you that you can remember in 2 weeks what the mother did while the children were exploring the underground tunnel?
9) How sure are you that you can remember in 2 weeks what the mother cat's name was?
10) How sure are you that you can remember in 2 weeks how many kittens the mother cat had?

References

Anderson, M. C. M., & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy*? Acta Psychologica, 128*, 110–118.

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Hillsdale: Erlbaum.

Bjorklund, D. F., Gaultney, J. F., & Green, B. L. (1993). "I watch therefore I can do:" The development of meta-imitation over the preschool years and the advantage of optimism in one's imitative skills. In R. Pasnak & M. L. Howe (Eds.), *Emerging themes in cognitive development, Vol. II: Competencies* (pp. 79–102). New York: Springer-Verlag.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markham (Eds.), *Handbook of child psychology, Vol. 3. Cognitive development* (pp. 77–166). New York: Wiley.

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

Cassel, W. S., Roebers, C. M., & Bjorklund, D. F. (1996). Developmental patterns of eyewitness responses to repeated and increasingly suggestive questions. *Journal of Experimental Child Psychology, 61*, 116–133.

Cohors-Fresenbort, E., Kramer, S., Pundsack, F., Sjuts, J., & Sommer, N. (2010). The role of metacognitive monitoring in explaining differences in mathematics achievement. *ZDM The International Journal on Mathematics Education, 42*, 231–244.

Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: differential and sufficient aspects. *Journal of Experimental Child Psychology, 47*, 274–296.

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension. A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*, 228–232.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory and Cognition, 20*, 374–380.

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.

Elischberger, H. B., & Roebers, C. M. (2001). Improving young children's free narratives about an observed event: the effects of nonspecific verbal prompts. *International Journal of Behavioral Development, 25*, 160–166.

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

Flavell, J. H., Miller, P. H., & Miller, S. A. (2002). *Cognitive development* (4th ed.). Englewood Cliffs: Prentice Hall.

Ghatala, E. S., Levin, J. R., Pressley, M., & Goodwin, D. (1986). A componential analysis of the effects of derived and supplied strategy-utility information on children's strategy selections. *Journal of Experimental Child Psychology, 41*, 76–92.

Hacker, D. G., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*, 160–170.

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

Keleman, W. L. (2000). Metamemory cues and monitoring accuracy: judging what you know and what you will know. *Journal of Educational Psychology, 92*, 800–810.

Kimbal, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory and Cognition, 31*, 918–929.

Koriat, A. (1993). How do we know what we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609–639.

Koriat, A. (1997). Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009). The memorizing-effort heuristic in judgments of memory: A developmental perspective. *Journal of Experimental Child Psychology, 102*, 265–279.

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

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language, 52*, 478–492.

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

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* (pp. 1–18). Boston: Kluwer.

Kunzinger, E. L., III. (1985). A short-term longitudinal study of memorial development during early grade school years. *Developmental Psychology, 21*, 642–646.

Lipko, A. R., Dunlosky, J., & Merriman, W. E. (2008). Persistent overconfidence despite practice: the role of task experience in preschoolers' recall predictions. *Journal of Experimental Child Psychology, 103*, 152–166.

Lipko, A. R., Dunlosky, J., Hartwig, M. K., Rawson, K. A., Swan, K., & Cook, D. (2009). Using standards to improve middle school students′ accuracy at evaluating the quality of their recall. *Journal of Experimental Psychology: Applied, 15*, 307–318.

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

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.

Maki, R. H. (1998). Predicting performance on text: delayed vs. immediate predictions and tests. *Memory and Cognition, 26*, 959–964.

Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 1263–1274.

McGivern, J. E., Levin, J. R., Pressley, M., & Ghatala, E. S. (1990). A developmental study of memory monitoring and strategy selection. *Contemporary Educational Psychology, 15*, 103–115.

Metcalfe, J., & Cornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52*, 463–477.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: the "delayed-JOL effect. *Psychological Science, 2*, 267–270.

Nelson, T. O., & Dunlosky, J. (1992). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science, 3*, 317–318.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125–141). New York: Academic.

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

Nietfield, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning, 1*, 159–179.

Peterson, C., & Whalen, N. (2001). Five years later: children's memory for medical emergencies. *Applied Cognitive Psychology, 15*, 116–128.

Plummert, J. M., & Schwebel, D. C. (1997). Social and temperamental influences on children's overestimation of their physical abilities: links to accidental injuries. *Journal of Experimental Child Psychology, 67*, 317–337.

Pressley, M., & Ghatala, E. S. (1989). Metacognitive benefits of taking a test for children and young adolescents. *Journal of Experimental Child Psychology, 47*, 430–450.

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.

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

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., Schneider, W., & Howie, P. (2007). Children's metamemorial judgments during an event recall task. *Journal of Experimental Child Psychology, 97*, 117–137.

Saywitz, K. J., & Geiselman, R. E. (1998). Interviewing the child witness: Maximizing completeness and minimizing error. In S. J. Lynn & K. M. McConkey (Eds.), *Truth in memory* (pp. 190–223). New York: Guilford.

Schneider, S. L., & Laurion, S. K. (1993). Do we know what we've learned from listening to the news? *Memory and Cognition, 21*, 198–209.

Schneider, W., & Lockl, K. (2008). Procedural metacognition in children: Evidence for developmental trends. In J. Dunlosky & B. Bjork (Eds.), *A handbook of memory and metacognition*. Mahwah: Erlbaum.

Schneider, W., Visé, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: evidence from a judgment-of-learning task. *Cognitive Development, 15*, 115–134.

Shaughnessy, J. J., & Zechmeister, E. B. (1992). Memory monitoring accuracy as influenced by the distribution of retrieval practice. *Bulletin of the Psychonomic Society, 30*, 125–128.

Shin, H., Bjorklund, D. F., & Beck, E. F. (2007). The adaptive nature of children's overestimation in a strategic memory task. *Cognitive Development, 22*, 197–212.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 204–221.

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28*, 129–160.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66–73.

Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1267–1280.

Vesonder, G. T., & Voss, J. F. (1985). On the ability to predict one's own responses while learning. *Journal of Memory and Language, 24*, 363–376.

von der Linden, N., & Roebers, C. M. (2007). Developmental changes in uncertainty monitoring during an event recall task. *Metacognition & Learning, 1*, 213–226.