Metacognition Learning (2014) 9:1–23
DOI 10.1007/s11409-013-9109-4

Can you see me thinking (about my answers)? Using eye-tracking to illuminate developmental differences in monitoring and control skills and their relation to performance

Thomas Roderer · Claudia M. Roebers

Received: 1 October 2012 / Accepted: 5 December 2013 /
Published online: 18 January 2014

Abstract This study focuses on relations between 7- and 9-year-old children's and adults' metacognitive monitoring and control processes. In addition to explicit confidence judgments (CJ), data for participants' control behavior during learning and recall as well as implicit CJs were collected with an eye-tracking device (Tobii 1750). Results revealed developmental progression in both accuracy of implicit and explicit monitoring across age groups. In addition, efficiency of learning and recall strategies increases with age, as older participants allocate more fixation time to critical information and less time to peripheral or potentially interfering information. Correlational analyses, recall performance, metacognitive monitoring, and controlling indicate significant interrelations between all of these measures, with varying patterns of correlations within age groups. Results are discussed in regard to the intricate relationship between monitoring and recall and their relation to performance.

Keywords Development · Performance · Metacognition · Metamemory · Confidence judgments · Implicit · Eye-tracking

Introduction

While it is generally advisable to answer questions correctly, sometimes this is not possible in an absolute and final way. However, in many every-day situations or in an educational context, it is important that knowledge is accurately assessed and that subsequent actions are adequate and based on these memory evaluations. Thus, memory related metacognitive skills encompass not only the ability to *monitor* learning and retrieval, but also the *control* of memory related information processing (Nelson and Narens 1990). Control behaviors may encompass looking for more information when uncertain, increasing learning and recall effort, and/or withholding an answer. Such and similar behaviors affect performance in educational and test settings (e.g., Renner and Renner 2001; Schneider and Pressley 1997; Thiede et al. 2003; Veenman and Spaans 2005).

T. Roderer (✉) · C. M. Roebers
Department of Psychology, University of Bern, Fabrikstrasse 8, CH-3012 Bern, Switzerland
e-mail: thomas.roderer@phsg.ch

Monitoring

In general, younger children's metacognitive skills deviate from those of older children and adults because the estimation of their own knowledge or the quality of their answers is highly optimistic. This overconfidence arises from too positive estimations of correctness when the given answers are in fact incorrect or even unanswerable. As the level of confidence for correct answers is high in all age groups, developmental progression is mainly related to uncertainty monitoring of difficult, misleading, and incorrectly answered questions (Howie and Roebers 2007; Roderer and Roebers 2010; Roebers 2002). However, since university students' uncertainty monitoring is also inadequate in some tasks, progression of metacognitive monitoring is protracted over the elementary school years into young adulthood (Allwood et al. 2006; Schneider and Lockl 2008; Visé and Schneider 2000; Von der Linden and Roebers 2006).

Many of these findings are based on explicit judgments about the participant's confidence in the accuracy of his or her answer to a memory task. Such confidence judgments (CJ) generally stand at the end of sometimes quite complex memory retrieval and memory monitoring processes. Often, monitoring does not follow retrieval in a straight forward way. Instead, these processes may be intertwined with, for example, one or more retrieval attempts preceding and following unfavorable evaluations of the predominant answer. Therefore, when focusing only on the final and explicit CJ, the continuous nature of the monitoring process is lost. In this vein, the overconfidence bias often found in children's explicit CJ may be related to a method of gathering data which focuses on end products and thereby misses monitoring processes occurring concordantly with retrieval processes. Their explicit judgments at the end of the task are more likely to be influenced by social or motivational processes, such as a social-desirability-response-set or wishful thinking, than those of older participants (Schneider 1998). As these processes may induce overconfidence, accurate monitoring based on cognitive-evaluative processes may be masked in the explicit judgments. Measuring monitoring processes during retrieval allows comparing them with explicit CJ in order to uncover the beginning of young children's overconfidence.

Monitoring and controlling: an intertwined relationship

Monitoring processes preceding explicit judgments are closely associated with retrieval process (Koriat 1993, 1995; Metcalfe and Finn 2008; Schraw and Roedel 1994). As such, they are difficult to capture as they are located at the interface between implicit and explicit processes (Koriat 1998, 2000). Following Schacter's (1987) distinction of implicit and explicit memory, such early monitoring processes can be characterized as implicit, as they are influenced by experience but do not require conscious or intentional recollection of those experiences.

As assumed in the accessibility model (Koriat 1993, 1995), access to more or higher quality information about the recall process, like recall or recognition latency, should promote more accurate judgments. The opposite is to be expected for low quality information. This model is relevant for monitoring processes in general. Thus, differences in access to memory cues are likely to affect implicit monitoring in a way similar to explicit monitoring judgments. However, its close ties to retrieval processes makes implicit monitoring an even better candidate for influences through experiences related to accessibility, quantity, and intensity of information than explicit judgments. With regard to young children's overconfidence, there are two possible outcomes when comparing implicit and explicit monitoring measures. Following the wishful thinking hypothesis, young children are less apt at differentiating between expectations and wishes concerning their own performance (Schneider 1998).

Wishful thinking, therefore, should overshadow the whole monitoring process. On the other hand, implicit monitoring could be quite accurate even in young children. Their self-reports of judgments may be distorted by a social desirability bias resulting in overconfident explicit CJ. In this case, a difference in the accuracy of implicit and explicit CJ would be expected.

In an earlier study, we followed Veenman's (2005) suggestion to use eye-tracking technology as a promising tool for studying metacognitive processes (Roderer and Roebers 2010). Implicit metacognitive judgments were operationalized by measuring the cumulated maximum fixation time on one of five icons before they were used by the participants to indicate their explicit metacognitive judgments. Similar developmental progressions as for explicit memory and performance evaluation were confirmed for these measures of implicit monitoring. The developmental progression for implicit CJs paralleled those for explicit judgments, with adequate differentiation between correct and incorrect answers even in young children (Roderer and Roebers 2010). The difference was more pronounced in 9-year-old children, while 7-year-olds' implicit CJs for incorrect answers were only adequate when item difficulty was low but not when items were more difficult. In addition, analyses of fixation time allocated to the different areas of the CJ scale revealed that younger children, especially, looked longer at medium to high than low CJs even when the answer was incorrect, implying a bias towards high CJs. These results imply that young children started off with an already very high opinion of their performance indicative for wishful thinking (Schneider 1998). The present study aims at replicating the results of the Roderer and Roebers (2010) study and extending it with respect to implicit monitoring in adults: In a memory task, first Japanese symbols (Kanji) and their meaning are presented in a learning trial. In a multiple-choice recognition task the Kanji is then presented as recall cue, and participants are required to choose one picture out of four as the correct translation for the Kanji. In addition, they rate their certainty that their solution was correct on a 5-point scale (Inspection of Fig. 1 may facilitate the comprehension of the task).

Monitoring: explicit and implicit CJ

As some monitoring skills are already apparent at a young age, significant differences between CJ for correct and incorrect answers, for both implicit and explicit monitoring, are expected in all age groups. Again, for implicit and explicit monitoring alike, and due to developmental trends in uncertainty monitoring, lower mean CJ for incorrect answers and increasing differences between mean CJ for correct and incorrect answers are expected in older participants. Introducing unanswerable items allows for an additional comparison of uncertainty monitoring

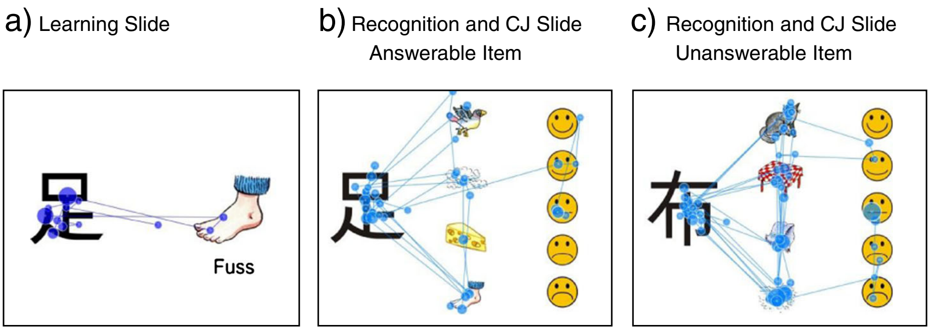

Fig. 1 Examples of a third grader's gaze paths and fixations on a learning slide and slides for retrieval and CJs for answerable and unanswerable items. *Note*. Fuss = Foot

skill between age groups. Kanji that had not been present during learning trials were introduced in the recognition task. As many Kanji share the same features and the complexity of answerable and unanswerable items was similar, using Kanji allowed introducing items which should elicit uncertain judgments in all age groups. Still, due to developing uncertainty monitoring, differences between age groups with lower CJ for unanswerable answers in older participants are expected.

Monitoring: correlations between implicit and explicit CJ

In general, the two measures of monitoring should be correlated, as implicit monitoring processes are likely to have an impact on later metacognitive judgments. As the Roderer and Roebers (2010) study revealed, children's fixations are biased towards high confidence judgments, which should translate into overconfidence in implicit CJ resulting in a high correlation with the expected overconfident explicit CJ. A significant correlation is also expected in the adult sample, as their better metacognitive abilities (e.g., more elaborate use of memory traces) should be mirrored both during monitoring and in the explicitly expressed CJ.

Monitoring: correlations with recall performance

As the claim that metacognitive skills aid task performance lies at the very basis of metacognition research, explicit and implicit monitoring variables are expected to be correlated with recognition performance. Uncertainty monitoring skills should be related to good performance, as they enable identification and revision of unsatisfactory solutions. Participants assigning low CJ to incorrect answers during recall (implicit monitoring) or after the answers have been given (explicit CJ) are more likely to find flawed answers, enabling error correction. However, as young children's uncertainty monitoring is not as established as that of older participants, a translation of uncertainty monitoring skills into better performance is less likely; therefore the correlations are assumed to be weaker.

Controlling

Like monitoring skills, on-task metacognitive control behavior develops during the primary school years. Older children are more apt at withholding incorrect answers or adjusting study-time allocation to item-difficulty (Dufresne and Kobasigawa 1989; Krebs and Roebers 2010, 2011; Lockl and Schneider 2004; Roebers et al. 2009). In many of these studies, younger children's monitoring accuracy was substantial and close to that of older children. It seems that they were not able to translate information from monitoring processes into adequate control strategies (Schneider and Lockl 2008). Thus, it appears as if the development of control skills lags somewhat behind the development of monitoring skills.

An important goal of the present study is to gain insight into the development of control processes and their interplay with task performance and monitoring. With this in mind, connecting metacognition research with research on strategy use during problem solving may help the understanding of the participants' control behavior and its relation to developing monitoring skills and task performance. Control processes are not only important for recall monitoring in memory tasks, but also have an impact on monitoring of performance in multiple choice problem solving tasks (Bethell-Fox et al. 1984; Mitchum and Kelley 2010).

Mitchum and Kelley (2010), using the Raven's Progressive Matrices test of fluid intelligence (Raven et al. 1998), found that students looking longer at the problem matrices before inspection of the answer alternatives (constructive matching strategy) performed better and monitored their answers more accurately compared to those following a strategy more concerned with a stepwise elimination of response choices (response elimination strategy). In these tasks several solution alternatives are available to complete patterns of abstract shape analogies. This setup is similar to the one used in the present study, where together with the recall cue, several pictures are presented as possible translations for the Kanji. Since to our knowledge, similar analyses of strategic behavior are not available in the domain of learning and memory, we would like to transfer the method of strategic behavior analyses from a problem-solving to a recognition context. Consequently, detailed information is needed in order to identify which strategies were used. Applying an eye tracking device during learning and recognition provides such information in the form of eye movements and fixation times. Preferred processing of certain information specified by longer fixation times during learning and retrieval indicates strategic behavior as an expression of control processes.

However, in previous research on metacognitive control processes the investigated control-related behaviors were generally explicit, like pressing a button, turning a page, or indicating a decision (e.g., Dufresne and Kobasigawa 1989; Koriat et al. 2009b; Krebs and Roebers 2010; Mitchum and Kelley 2010). Measuring control processes via eye tracking, in contrast, allows measuring time used for learning and recall without explicit responses. This method is likely to be less intrusive and to tap on more natural behavior. This technology thus enables recording a variety of very subtle control behaviors during learning and recall. Bethell-Fox and colleagues (1984) recorded their participants' eye-movements in a multiple-choice geometric problem solving task. Increasing item difficulty and number of alternatives had an impact on eye movements. Results showed more reinspection within the problem when item difficulty was high, indicating that the subjects prepared themselves with greater care before looking at the answer alternatives (constructive matching strategy). However, there were also more look-backs from the answer alternatives to the problem with higher item difficulty, thus implying a strategy change towards response elimination when constructive matching failed. Relating scores of fluid intelligence to eye-movement patterns or fixation time on problem terms and answer alternatives revealed that higher ability subjects more often used constructive matching, while lower ability subjects more often employed a response elimination strategy (Bethell-Fox et al. 1984; Snow 1980; Vigneau et al. 2006).

On the other hand, two studies in our eye-tracking lab on children's development of attention allocation in a memory task revealed strategic learning behavior in young children. Across the age groups 6-year-old to 10-year-old children displayed control over their learning by allocating more fixation time to critical as opposed to irrelevant or distracting information (Roderer et al. 2012; Roebers et al. 2010). Differences in the proportion of fixation time allocation can be interpreted as developmental progression in the use of adequate learning strategies. In all age groups, aspects of learning strategy use were significantly related to recall performance whereas the correlations were highest for the oldest children. In the Roderer and Roebers (2010) study, using a very similar design as in the present study, we found strategic control behavior in 7-, and 9-year-old children: Both age groups looked longer at the critical information, the Kanji (recall cue), than the answer alternatives. This difference, however, was generally larger for the older children and even more pronounced for more difficult items as compared to the 7-year-olds who did not adjust allocation of fixation time to item difficulty. At the same time, recall performance was higher and metacognitive monitoring was more accurate in the 9-year-olds (Roderer and Roebers 2010).

Controlling: age related differences

With these results in mind, developmental differences in control behavior can be expected as older participants are more likely to display strategic learning and retrieval behavior, mirroring additional experience with memory tasks. Therefore, during learning trials they are expected to allocate more time to encoding the shape configuration of the Kanji sign. Also during recognition, they are expected to invest more time into analyzing the provided recall cue and comparing it with the memorized shape in comparison with younger participants. With respect to unanswerable items, adaptation of a response elimination strategy is expected in all age groups, thus unanswerable Kanji should receive more attention than answerable Kanji.

In addition, due to their more pronounced and efficient use of encoding and constructive matching strategies, older participants should have access to more memory cues relevant for monitoring. As availability of such cues supports faster decisions for explicit CJ, older participants are expected to allocate less attention to the icons representing the CJ than younger participants.

Controlling: correlations with recall performance

Application of these learning and recall strategies is supposed to positively correlate with performance and this more so in older participant samples who are more likely to use them.

Also, response elimination is a helpful strategy to increase recognition performance. It has its pitfalls, however, particularly when the incorrect alternatives get too much attention. Longer fixation times on an incorrect as compared to the correct answer may give rise to the impression that this incorrect alternative is related to the displayed Kanji. This new relation may interfere with the learned connection between the Kanji and the correct answer. Thus, longer fixation times on incorrect alternatives may distract from the correct alternatives and contribute to low recognition performance. Based on the results of the Roderer and Roebers (2010) study, we expect younger children to be more challenged by this abstract memory task and to rely more on a recognition strategy in line with response elimination rather than constructive matching, much like the lower ability subjects in problem solving tasks (Bethell-Fox et al. 1984; Snow 1980; Vigneau et al. 2006). As response elimination comes with longer fixation times on incorrect answers, we expect the negative correlation between this variable and recognition performance to be higher in younger participants.

Correlations between monitoring and controlling

Research on interaction of monitoring and control abilities in children revealed that primary school students can rely on metacognitive evaluations of their answers in order to control their behavior and thereby improve recall accuracy. In these studies using cloze tests in imitation of actual school tests, increases in recall accuracy were more pronounced in 11-year-old compared to 9-year-old children when they were allowed to withhold answers based on their CJs (Krebs and Roebers 2010, 2011; Roderer and Roebers 2009; Roebers et al. 2009). Conversely, monitoring can in turn be influenced by control processes like time and effort invested in a retrieval attempt or selection of problem solving strategy (Koriat and Goldsmith 1996; Mitchum and Kelley 2010). Children adjust their estimates for future retrieval according to the study time or number of repetitions necessary to learn the test content, with older children relying more on learning related experiences and cues in order to increase monitoring accuracy (Koriat et al. 2009a, b).

Regarding the relation between control behavior and monitoring, following the accessibility model, the cognitive processes leading up to the production of responses are important sources of cues for metacognitive judgments (Koriat 1993, 1995). Processes controlling the encoding of information, like directing more attention towards some content than other, may contribute useful cues to memory monitoring. If more of these processes are necessary during learning or retrieval of an item in comparison to other items, the latency to solve the task increases, and it is likely to be perceived as more difficult. The latency during learning or the attempt to produce a response has been found to be a valid basis for CJ, with longer latencies being related to less confident metacognitive judgments (Koriat et al. 2006). This relation is most likely to be true for controlling behavior directed at problem relevant information during learning and retrieval, the Kanji. As developmental differences in adequate control behavior are expected, older participants should have access to more memory cues relevant for monitoring. As they are also more apt at using information from control processes (Koriat et al. 2009a, b), the relation between fixation time on the Kanji and uncertainty monitoring should be stronger in older participants. Using eye-tracking technology also allows gathering information about control behavior directed at distracting and possibly misleading information: the incorrect answer alternatives. Just as concerning oneself with problem related information increases uncertainty, extended fixation time on putative solutions may give rise to the impression that the examined (but in fact incorrect) answer is correct. In such a case, the associated judgment of the answers' correctness is likely to be overconfident. As younger participants are expected to direct more attention to the alternatives, including the incorrect ones, they are also more likely to get misleading cues. Therefore, we expect increasingly higher positive correlations between CJ for, and fixation times on, incorrect answers with decreasing age of participants.

While implicit monitoring processes are deeply intertwined with recall related control processes, it's the explicit CJ which stand at the end of the control and implicit monitoring processes. Thus, the significant correlations between monitoring and controlling variables are expected for both implicit and explicit CJ.

The following list shows a summary of our specific hypotheses:

Monitoring

All age groups:

- Higher Explicit and Implicit CJ for correct than incorrect answers.
- Explicit and implicit CJ are correlated.

Age related differences:

- Lower explicit and implicit CJ for incorrect answers in older age groups than in younger age groups
- Lower explicit and implicit CJ for unanswerable items in older age groups than in younger age groups

Controlling

All age groups:

- Recognition: More fixation time allocated to the Kanji of unanswerable items compared to answerable items

Age related differences:

– Learning: More fixation time allocated to the Kanji in older age groups than in younger age groups
– Recognition: More fixation time allocated to the Kanji in older age groups than in younger age groups
– Recognition: More fixation time allocated to the correct answer in older age groups than in younger age groups
– Recognition: More fixation time allocated to the incorrect answer in younger age groups than in older age groups
– Monitoring: More fixation time allocated to CJ in younger age groups than in older age groups

Correlations between monitoring, controlling and recall performance

All age groups:

– Negative correlation between recall performance and explicit and implicit CJ for incorrect answers
– Positive correlation between recall performance and fixation time allocated to the Kanji during learning and recall
– Negative correlation between recall performance and fixation time allocated to incorrect answer alternatives during recall

Correlations between monitoring and controlling

All age groups:

– Negative correlation between fixation time allocated to the Kanji during learning and recall and explicit and implicit CJ for incorrect answers
– Positive correlation between fixation time allocated to the incorrect alternative during recall and explicit and implicit CJ for incorrect answers

Taken together, we aim to extend our knowledge on the development of monitoring, control, and performance in several aspects. First, replicating the results for implicit and explicit metacognitive skills as found in the predecessor study (Roderer and Roebers 2010) would confirm the reliability of the findings and including an adult sample allows extended inferences on the developmental trajectory of metacognitive monitoring. Further advances in our understanding of uncertainty monitoring for incorrect and unanswerable items in adults compared to children are expected. Second, measuring controlling by fixation time allocation during learning and retrieval allows for acquisition of behavioral data free of explicit control responses. The participant's behavior is therefore more likely to resemble natural learning and test situations without response demands which may interrupt or induce awareness of ongoing information processing. Concerning the third goal of the study, interrelations between controlling, monitoring, and performance will be explored.

To reach all of these goals, the present study is inspired by, and shares several features with earlier research. These include the use of CJs (Mitchum and Kelley 2010), comparison of on task controlling behavior and performance (Vigneau et al. 2006), and the use of eye-tracking technology (Bethell-Fox et al. 1984; Roderer and Roebers 2010; Snow 1980; Vigneau et al. 2006).

Method

Participants

A total of $N=148$ children and adults (50 % female) participated in the study. Technical failures or incomplete or monotonous responses resulted in missing data and therefore, sample sizes may vary slightly between different analyses as cases with missing data were excluded for each analysis separately. For the analyses reported below, data of $N=54$ 7-year-olds ($SD=$ 4.2 months; 46 % female), $N=44$ 9-year-olds ($SD=5.5$ months; 43 % female) as well as a sample of $N=43$ adults ($M_{age}=25$ years, $SD=30.3$ months; 60 % female) were included. Adult participants were recruited on the campus of the University of Bern, Switzerland. For the children's samples, school administrations and teachers of public schools in the region of Bern were contacted. Written informed consent was obtained from the parents. The adults were asked for allowance to use the data.

Material

The main task for this study is a modification of the task used in Roderer and Roebers (2010). In order to assess recognition and procedural metamemory, the participants are required to learn Japanese symbols (Kanji); therefore we will subsequently refer to this task as the Kanji-Task. Kanji were used as test material as they have several advantages compared to words or pictures, among which are: learning is independent of reading ability; learning real Kanji is very motivating and interesting for children; the abstract configuration as well as the similarity in surface features which allows introducing unanswerable items without them standing out. Learning, recognition, and CJs were divided into two separate sets in order to reduce task demands. The number of different Kanji - picture associations in the learning phase of each set differed across the age groups. The 7-year-olds learned 10, the 9-year-olds 12, and the adults 18 associations. More items were introduced for older participants in order to adjust task difficulty to the different age groups[1].

The images used in the learning phase consisted of a Kanji on the left hand side together with a pictorial representation of its meaning serving as translation on the right hand side. The pictures showed colored line-drawings including the object's name beneath each picture (see Fig. 1a). The Kanji were drawn from a pool of items covering the whole difficulty continuum from easy to hard. Each Kanji – translation slide was displayed for 5 s.

In the recognition and CJ phase, the Kanji from the learning phase were presented on the left side; the picture of the translation was replaced with four possible answers in the middle and the CJ scale consisting of five smiley-faces on the right side of the screen. Answer alternatives and smiley-faces were presented in an upright line (see Fig. 1b). The four alternatives for the easy and difficult Kanji included two new pictures of objects and two pictures associated with the Kanji during learning, including the correct answer. New pictures were included in order to increase the total number of possible answers and to decrease the repetition of each picture. In addition, as unanswerable items were blended in with the other items during recognition, adding new answer alternatives was necessary. The answer

[1] However, with a correct recall of $M=45$ % in 7-year-olds ($SD=15$ %), $M=54$ % in 9-year-olds ($SD=12$ %) and $M=83$ % in adults ($SD=13$ %), the differences in recall performance were highly significant, $F\,(2, 138)=103.48$, $p<.001$, $\eta_p^2=.60$.

alternatives' position on the screen and the combination of learned and new pictures was counterbalanced. All alternatives, the new ones as well as those from the learning trial were used twice. In each set, only 8 out of all Kanji learned were used in the recognition phase. In addition, four new and therefore unanswerable Kanji were introduced, resulting in a total of 12 items presented for recognition in each set. The recognition task was identical for all participants.

The whole procedure, including instruction, practice, and test trials was presented on a remote non-invasive eye-tracking system (Tobii 1750, Tobii Technology AB, Danderyd, Sweden) with a 17 in. (43 cm) computer screen. Item presentation during the learning and recognition/CJ phase was in random order.

Eye-tracking data

As in many studies using fixation data as indicator for information processing, a fixation filter was applied to the gaze data in order to identify the location and duration of fixations (Hauland 2003; Henderson and Hollingworth 1998; Kinnunen and Vauras 1995). In the present study, the fixation filter as implemented in the TobiiStudio, 2.0 software was used (Olsen 2012). This filter is based on a Velocity-Threshold Identification fixation classification algorithm which is used to classify eye movements below a velocity of 30°/s (visual degrees per second) as part of a fixation. Only fixations falling within predefined areas of interest on the slides containing relevant segments of the pictures (e.g., Kanji, answer alternatives, confidence scale) were included in the analyses. Average allocation of fixation time was calculated by summing the fixation times for the different areas of interest across relevant items and then dividing by the total number of items. The items included for the collection of eye-tracking data were the same for all participants. In the case of the learning phase, fixation data of 16 items across both trials was used. Data for items learned but not presented in the recognition phase (4 for 7-year-olds, 8 for 9-year-olds and 20 for adults) was omitted. For the recognition and CJ phase, data for all 16 answerable and 8 unanswerable items were included.

In order to prevent any impact of explicit judgments on the data used for calculating implicit judgments, and learning and recognition behavior, the eye-tracking data for each recognition and judgment-slide was carefully analyzed. Having the participants pointing at the screen when indicating a CJ results in eye movements that are anchored to hand movement which is typical for eye-hand coordination (Hayhoe et al. 2003; Neggers and Bekkering 2002). Pointing at the screen in order to provide explicit CJs results in fixation times on CJs which are longer than average fixations or fixation sequences. They differ from eye-movements towards the smiley-faces related to implicit judgments. Based on the unique characteristic of the eye-movements and the records on the protocol sheet, fixations corresponding to explicit CJs were identified. The implicit CJs were then calculated based on fixations on the CJ scale prior to fixations corresponding to explicit CJs.

Procedure

Children were tested individually by the first author and a trained research assistant in a quiet room at their school. Adults were also tested individually at the research lab at the University of Bern. The research assistant walked the children from their classroom to the test room, thereby building rapport as well as asking them if they had some experience with Japanese Kanji symbols. Adults were welcomed in a similar way at the university.

All participants were seated in a comfortable armchair in front of the eye-tracking system hidden in a computer screen. An adjustable mounting allowed for optimal placement of the screen with a distance of 60 cm from the participants' eyes. Next, a 5-point eye tracker calibration was performed. Thereafter, the participants were introduced to the task. They were told that the Japanese write in symbols that sometimes stand for entire words and that their task would be to learn some of these symbols and their meaning. An example of a Kanji symbol and the depicted translation including the objects' name were shown to the participants (see Fig. 1a). Next, the recognition and CJ slide was displayed. The participants were instructed to have a good look at the Kanji, and then choose the correct answer. If they were not able to recall the correct answer, they were instructed to guess (forced report). The experimenter continued by explaining that not only correct recall but also correct judgment of the recall was important in this task and that, therefore, they would later be asked how confident they are regarding the correctness of their answers. For these CJs, a 5-point scale was introduced (see Fig. 1b). Each of the five smiley-faces was declared to the participants as standing for an indication of how sure or unsure they are that the previously given answer is correct. If an answer had to be guessed, the experimenter suggested choosing the confidence rating at the lower end of the scale to indicate uncertainty. Subjects were instructed to provide their judgments by pointing with the finger at the chosen CJ. The experimenter demonstrated how to use the scale and tested the participants' understanding of the instructions. In general, the participants learned the use of the confidence scale with ease. After these explanations, participants were given the opportunity to practice learning and recognition of four Kanji and giving CJs. This practice trial also served as test for the participants' understanding of the task.

Before starting with each learning phase, a plate containing all answer alternatives the participants would see during the trial was presented. The plates were matched to the age groups and the number of items they had to learn. The participants were asked to name all illustrations to ensure correct labeling, to control for idiosyncratic labels, and to increase familiarity with all depicted objects.

Following the practice trial and the labeling, the experimenter moved on to the test trials. In both of these trials children learned a set of 10 to 18 Kanji, depending on age group. For recognition and metacognitive judgments, 12 recognition and CJ slides were presented (8 answerable, 4 unanswerable items). The experimenter noted the answer and CJ while it was on display. The next slide was shown as soon as the answer was recorded. For the second test trial, the second set of Kanji symbols and their translation was applied. The order of the two trials was counterbalanced across participants. A computer game with a duration of about 4 to 6 min was inserted as a filler task between trials as well as within each trial between learning and recognition to prevent rehearsal.

After completing the two trials of the Kanji-Task, participants were debriefed about the unanswerable Kanji and thanked for their participation. The children were praised for the good performance, given a small gift, and escorted back to their classroom. Adults received 10 Swiss Francs and questions were answered.

Results

Preliminary analyses did not reveal any systematic differences between male and female participants. Data were therefore collapsed across gender. Partial eta^2- values are reported as estimators of effect size.

Table 1 Mean explicit and implicit confidence judgments as a function of age

Item difficulty	7-year-olds	9-year-olds	Adults
Explicit CJs			
Correct answers	4.34 (0.61)	4.09 (0.67)	4.25 (0.44)
Incorrect answers	3.66 (0.94)	2.97 (0.99)	2.27 (0.81)
Unanswerable	3.45 (1.02)	2.83 (1.04)	1.62 (0.41)
Implicit CJs			
Correct answers	3.84 (0.52)	3.76 (0.52)	3.77 (0.51)
Incorrect answers	3.70 (0.57)	3.23 (0.70)	2.73 (0.62)
Unanswerable	3.78 (0.61)	3.33 (0.84)	2.41 (0.52)

Note. Standard deviations are in parentheses

Monitoring: explicit CJ

The upper half of Table 1 contains the mean explicit CJ, divided in CJs for correct and incorrect answers and unanswerable Kanji, separated for the three age groups. Several analyses were computed in order to look at age differences in metacognitive monitoring. First, an ANOVA on the CJs for answerable items with correctness of answer as within-subject factor and age as between-subject factor was conducted and revealed significant main effects for correctness of answer, F (1, 130)=360.96, p<.001, η_p^2=.74 and age, F (2, 130)=29.15, p<.001, η_p^2=.21, as well as a significant interaction effect between correctness of answer and age, F (2, 130)=32.92, p<.001, η_p^2=.33. Follow-up ANOVAs on the interaction between correctness of answer and age revealed that it was due to the fact that the 9-year-olds' CJs for correct answers were lower than those of the 7-year-olds' [Tukey HSD post-hoc, p=.03] and did not differ from the adults' [p=.44], while for incorrect answers significant decreases in CJs between groups indicate improved monitoring of uncertainty with increasing age F (2, 130)= 28.10, p<.001, η_p^2=.30 [Tukey HSD post-hoc tests; 7-year-olds>9-year-olds>adults; all p<.001]. In an additional analysis of uncertainty monitoring skills, CJs of the unanswerable questions were compared across age groups. Results revealed a significant main effect of age, F (1, 136)=55.75, p<.001, η_p^2=.45 [Tukey HSD post-hoc tests; 7-year-olds>9-year-olds> adults; all p<.001].

Monitoring: implicit CJ

As explained earlier, the dependent variables for implicit CJs were calculated based on the eye-tracking data *before* an explicit CJ was expressed. For the following analyses, we used the smiley-face of the CJ scale that attracted a maximum of fixation time during presentation of a recognition and CJ slide as a measure for the implicit CJ[2]. However, this approach resulted in the loss of some implicit CJ, as for some items (e.g., very easy items) participants looked directly at only one CJ (the explicit CJ). As the missing values are likely to be biased towards high CJ, they were not substituted (with e.g. the overall or subgroup mean). On average, there were M=6.22 (SD=6.01) values of the 20 items missing per participant.

In a first step, we present basic analyses that parallel those for explicit CJs (see Table 1 for a comparison of explicit and implicit CJs). The first ANOVA included correctness of answer as

[2] Other measures of implicit CJs were calculated as well (e.g., the last fixation on a confidence judgment before the explicit judgment was given during recording). These analyses revealed the same or very similar results.

within-subject factor and age as between-subjects factor. Significant main effects for both, correctness, F (1, 124)=93.14, p<.001, η_p^2=.43, and age, F (2, 124)=12.37, p<.001, η_p^2=.17, were found as well as a significant interaction between correctness of answer and age, F (2, 124)= 18.46, p<.001, η_p^2=.23. To follow up on the significant interaction, two separate ANOVAs on CJs for correct and incorrect items were calculated. The ANOVA on CJs for correct items showed no age effect, F (2, 124)=0.36, p=.70, η_p^2=.01. For incorrect answers, however, CJs decreased significantly from younger to older age groups confirming improved monitoring of uncertainty with increasing age as found for the explicit CJ, F (2, 124)=23.75, p<.001, η_p^2=.28 [Tukey HSD post-hoc tests; 7-year-olds>9-year-olds>adults; all p<.01]. Also, implicit CJs for unanswerable items were compared across age groups in order to get insights into fast and early uncertainty monitoring skills, revealing a significant main effect of age, F (1, 124)=40.72, p<.001, η_p^2=.40 [Tukey HSD post-hoc tests; 7-year-olds>9-year-olds>adults; all p<.01].

Monitoring: correlations between implicit and explicit CJ

In order to get further insights into the relation between implicit and explicit CJs, several one-tailed correlations were computed. Collapsing data across answerable and unanswerable items and age groups, there was a high correlation between implicit and explicit CJs, r=.78, p<.001. Correlations between implicit and explicit CJs tended to be very high for 7-year-olds, r=.75, p<.001, and 9-year-olds, r=.86, p<.001, compared to medium correlations in adults, r=.44, p<.01, a pattern that was confirmed when the answerable and unanswerable items were examined separately.

Controlling: allocation of fixation time

During the learning trials, where presentation time was fixed to 5 s for all participants, total fixation time on the screen differed significantly between age groups, F (1, 136)=55.75, p<.001, η_p^2=.45. Tukey HSD post-hoc tests revealed, that average fixation times of 7-year-olds (M=4.54 s, SD=0.24) was significantly lower, p<.01, than those of adults (M=4.70 s, SD=0.21), while 9-year-olds' fixation times (M=4.59 s, SD=0.23) were in between, only marginally significant lower than the adults', p=.08. In contrast, there were no significant differences between the 7-year-olds' (M=6.92 s, SD=1.98), 9-year-olds' (M=7.14 s, SD= 1.60) and adults' (M=7.44 s, SD=2.01) mean fixation times for the recognition and CJ screens, F (1, 133)=0.87, p=.42, η_p^2=.01, where presentation time was not restricted.

For the following analyses the eye-tracking data was transformed in percentages of the fixation time in order to analyze the distribution of attention across the different areas of interest whereas the total sum of fixation time on the whole slide amounts to 100 %.

Controlling: fixation time allocation during learning

The percentages of fixation time allocated to the Kanji and the picture of the translation are presented in Fig. 2. Due to the fixed presentation time, fixation time allocation to these areas of interest is not independent. In order to look into age effects, therefore, only an ANOVA concerning the allocation of fixation time for the Kanji to be learned was computed. It revealed that older participants directed more attention at the Kanji with significant differences between all age groups, F (2, 133)=74.75, p<.001, η_p^2=.53 [Tukey HSD post-hoc tests; adults>9-year-olds>7-year-olds; all p<.001].

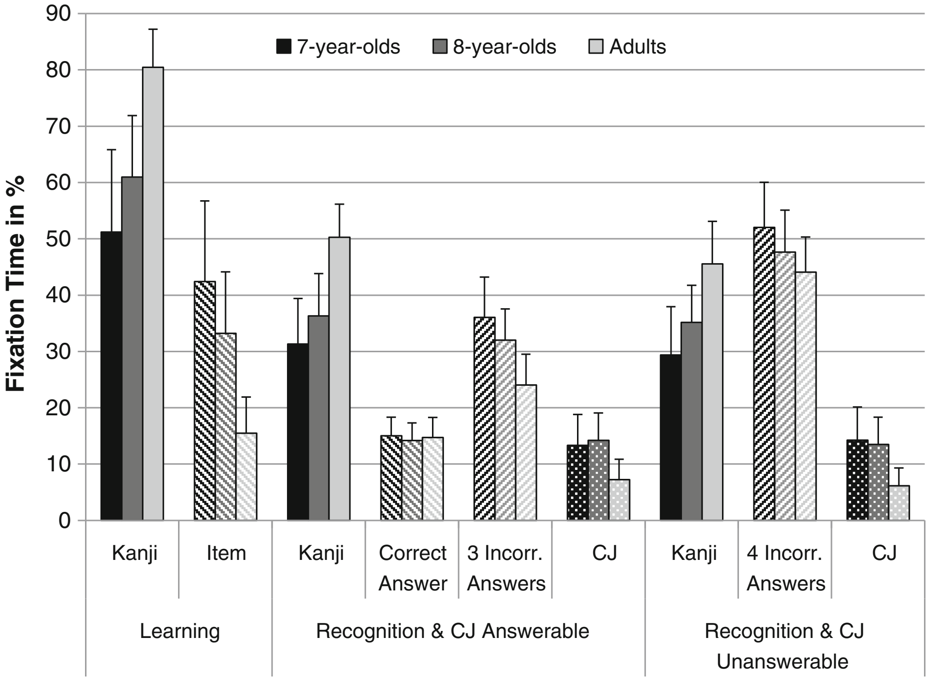

Fig. 2 Controlling: percent of fixation time for learning and recognition slides as a function of age and item answerability. *Note*. *Error bars* represent the standard deviation of the mean

Controlling: fixation time allocation during recognition

In Fig. 2, the mean percentages of fixation time allocated to the Kanji, the correct answer, the sum of incorrect answer alternatives, and the sum of all CJs are presented. ANOVAS were performed to compare percentages of fixation time for the different areas of interests as a function of participants' age.

Controlling: fixation time allocation to the kanji

For the fixation time on the Kanji symbols, a repeated measures ANOVA revealed significant main effects of item answerability, F (1, 132)=28.11, p<.001, η_p^2=.06 [answerable items>unanswerable items], and age, F (2, 132)=75.56, p<.001, η_p^2=.53 [Tukey HSD post-hoc tests; adults>9-year-olds>7-year-olds], which were qualified by a significant interaction between item answerability and age, F (2, 132)=4.55, p=.01, η_p^2=.18. To follow up on the significant interaction, in this case paired samples T-tests were computed in order find out if the Kanji were considered different as a function of item answerability. The results indicate that 7-year-olds, t (48)=2.61, p=.01, and adults, t (41)=4.72, p<.001 but not 9-year-olds, t (43)=1.43, p=.16, allocated more fixation time to the Kanji of answerable items. But while the 7-year-old children allocated somewhat less fixation time to the Kanji for unanswerable compared to answerable items, the same adjustment is more pronounced in adults. In addition, adults allocate still much more fixation time to the Kanji independent of item answerability (see Fig. 2).

Controlling: fixation time allocation to answer alternatives

Together with the Kanji, four alternative answers were presented (one correct and three incorrect alternatives).

A repeated measures ANOVA was performed. In the within subjects variable, percentages for fixation time allocated to the correct alternative and the mean of the three incorrect alternatives during recognition of answerable items were compared. The age factor constituted the between subjects variable. The results revealed a significant effect of correctness, F (1, 132)=192.95, $p<.001$, $\eta_p^2=.59$ [correct>incorrect], and age, F (2, 132)=12.63, $p<.001$, $\eta_p^2=.16$, as well as a significant interaction, F (2, 132)=12.93, $p<.001$, $\eta_p^2=.16$. In order to follow up on the significant interaction, separate analyses for fixation times on correct and incorrect answers were calculated. While there was no age effect for correct alternatives, F (2, 132)=0.76, $p=.47$, $\eta_p^2=.01$, the result for incorrect alternatives was highly significant, F (2, 132)=44.16, $p<.001$, $\eta_p^2=.40$ and Tukey HSD between group post-hoc comparisons revealed highly significant differences between all age groups [7-year-olds>9-year-olds>adults, all $p<.01$]. Thus, while there was no difference in how much fixation time was allocated to the correct answer, younger participants considered incorrect alternatives longer before expressing a CJ than older participants did.

An ANOVA for unanswerable items with percentages of fixation time allocated to the mean of all four alternatives revealed a significant effect of age group, F (2, 130)=13.25, $p<.001$, $\eta_p^2=.17$. Tukey HSD comparisons revealed significantly higher percentages of fixation time allocated to the answer alternatives in 7-year-olds in comparison to older participants [$p<.01$], and no difference between 9-year-olds and adults.

Monitoring: fixation time allocation to CJs

We computed an ANOVA for the total fixation time on the confidence scale with answerable vs. unanswerable items as within-subject factors and age as between-subjects factor (see Fig. 2). There was no significant main effect of item answerability, F (1, 131)=0.64, $p=.43$, $\eta_p^2<.01$, and no significant interaction, F (2, 131)=2.90, $p=.06$, $\eta_p^2=.04$, but a significant main effect of age [Tukey HSD post-hoc tests; 7-year-olds, 9-year-olds>adults], F (2, 131)=39.40, $p<.001$, $\eta_p^2=.38$.

Table 2 Correlation between retrieval performance, monitoring and controlling variables

Sample	7-year-olds	9-year-olds	Adults	Total sample
Monitoring				
Explicit CJs incorrect answers	-0.43**	0.02	0.23	-0.53**
Implicit CJs Incorrect Answers	-.20[t]	-0.06	-0.45**	-0.50**
Controlling learning				
% fixation time Kanji	0.48**	0.06	0.03	0.70**
Controlling Recall				
% fixation time Kanji	0.34**	0.40**	0.11	0.69**
% fixation time incorrect alternative	-0.54**	-0.32*	-0.50**	-0.72**

[t] <0.10; * $p<.05$; ** $p<.01$ (one-tailed)

Correlations between monitoring, controlling and recall performance

Finally, we turn to the correlational analyses between the most important controlling, monitoring, and performance variables (see Table 2). First, all monitoring and all controlling variables were significantly related to recognition performance when the whole sample is included in the analyses. The significant correlations were all of medium to high magnitude. The results for adequate monitoring imply that correctly detecting uncertainty and ascribing low CJs to incorrect answers were related to high performance for both explicit as well as implicit monitoring processes. Correlations for controlling are even higher, where allocating more fixation time to the Kanji is positively related to recognition performance both during learning and recognition. Longer fixation time for incorrect alternatives in the recognition phase, in contrast, correlated negatively with performance.

Separate analyses for the age groups showed a more differentiated picture. Both for monitoring and recognition, the youngest children's pattern of correlations was closest to that of the total sample. For the 9-year-olds, we found no significant correlation for monitoring variables with recognition and significant correlations of controlling only during recognition, but not during learning behavior. Interestingly for adults, only implicit monitoring and, in terms of controlling only fixation times on incorrect alternatives, are related to recognition.

Correlations between monitoring and controlling

As an intricate relationship between monitoring and recall is assumed, correlations between several variables of these metacognitive abilities were computed (see Table 3). For analyses based on the total sample, all correlations are significant and of medium magnitude. Importantly, the correlational patterns with controlling variables are similar for implicit and explicit monitoring measures, although they are based on very different data. Thus, both explicit and implicit uncertainty monitoring is related to better controlling as indicated by more fixation time allocated to critical information and less fixation time allocated to potentially interfering information. Separate analyses for the age groups again reveal that the pattern of correlations for the total sample is closest to that of the youngest children. For 9-year-olds we find only relations between controlling during recognition and implicit measures of uncertainty monitoring. In the case of adults, no correlations are significant.

Table 3 Correlation between monitoring and controlling variables

	Explicit CJs incorrect answers				Implicit CJs incorrect answers			
Sample	7-year-olds	9-year-olds	Adults	Total sample	7-year-olds	9-year-olds	Adults	Total sample
Controlling learning								
Fixation time Kanji	-0.24*	-0.01	0.03	-0.49**	-0.16	-0.12	0.18	-0.41**
Controlling recall								
Fixation time Kanji	-0.32*	-0.10	0.15	-0.50**	-0.31*	-0.31*	0.16	-0.48**
Fixation time incorrect alternative	0.48**	0.23^{t}	-0.08	0.52**	0.30*	0.34*	0.12	0.48**

^{t}<0.10; * p<.05; ** p<.01 (one-tailed)

Discussion

Monitoring

As in an earlier study (Roderer and Roebers 2010), in the present study the development of explicit and implicit uncertainty monitoring was investigated. By using a similar paradigm and adding an adult sample the replication and extension of earlier results allowed confirmation of an increasingly critical evaluation of incorrect answers between 7-years of age and adulthood.

Monitoring: explicit CJ

Thus, the 7-year-olds' CJs for incorrect answers and even unanswerable items are above the "neither sure nor unsure" judgment while the 9-year-olds' CJs oscillate around this middle category and only the adults' CJs clearly indicate uncertainty. However, the youngest children are also obviously able to differentiate between correct and incorrect answers. All in all, the results concerning developing ability to monitor the correctness of answers with CJ corresponds to those of other studies in this field (e.g., Schneider and Lockl 2008).

Monitoring: implicit CJ

When turning to implicit metacognitive judgments as measured by the eye-movements directed at the CJ-scale just before the CJs were stated, several points need to be highlighted; first and most important, the repeated successful implementation of a new method to measure early metacognitive processes strengthens the notion that precursors of explicit metacognitive judgments can be measured reliably via an eye-tracking device. The similarity of the data concerning the development of implicit metacognitive monitoring obtained in the two studies hints at the commensurateness of the results. As in the Roderer and Roebers (2010) study and paralleling the explicit CJs, the differences between the uncertainty monitoring abilities of the three age groups are evident for implicit CJs also. As can be inferred from the eye-tracking data, the children in both samples considered higher CJs longer than those below the middle category, even when their answers were incorrect or when the items were unanswerable. This pattern of results concords best with a wishful thinking hypothesis (Schneider 1998). Young children's monitoring is already biased during the recall process. Thus a motivational process seems to be active before and during the whole task: the desire to do well. It overshadows the subtle hints from implicit monitoring processes and thereby biases self-evaluation.

Nevertheless, a shift towards attention allocation to somewhat lower CJs in older children points to a development in the early phase of uncertainty monitoring, away from wishful thinking towards monitoring processes based on memory cues (Koriat et al. 2009a, b). Mean values for the adults' implicit CJs are also close to the middle category for incorrect answers, but clearly in the region of uncertainty and thus far off wishful thinking when it comes to unanswerable items. These results are more in line with the accessibility model, where (missing) memory cues informed the participants, that a given answer is not likely to be correct. Therefore, observing how participants look at the confidence scale can tell us a lot about developmental differences between age groups and when those differences appear. Also, adding an adult sample enabled verification that development of monitoring skills continues well into adulthood and allows comparing the children's current skill level against a realistic potential for future development.

Monitoring: correlations between implicit and explicit CJ

The method used to assess implicit metacognition has the additional advantage that both explicit and implicit judgments are measured using the same CJ scale although the methods for gathering the data are quite different. This approach allows direct comparisons of implicit and explicit judgments and the corresponding correlations show that, in general, convergence between implicit and explicit CJs is considerable. Interestingly, the correlations were highest in the children's samples. This may reflect wishful thinking with a bias towards high CJs both before and during the explicit judgment of their answer. As seen in the CJs, the impact of wishful thinking on the adults' overall monitoring is small. Moreover, lower correlations between the two monitoring measures are to be expected. However, the correlation is still substantial, highly significant, and likely to be based on early retrieval-related monitoring processes via their impact on adults' explicit monitoring judgments.

Controlling

Besides metacognitive monitoring, the development of metacognitive control skills is a central topic of this study.

Controlling: fixation time allocation during learning

Analyzing fixation time during learning and recall allowed drawing inferences from allocation of attention on the learning and recall strategies used by the participants. During the learning trials, allocating more fixation time to the critical information (Kanji) than their translations (the pretty and colorful drawings) is assumed to be the more adequate learning strategy. The Kanji is insofar critical for correct recognition, as its exact appearance is the cue to the correct answer while the answer itself can be chosen from an array of alternatives and only requires recognition. Inspection of Fig. 2 shows that this strategy was used in all age groups indicating that the youngest children's learning behavior is also strategic. However, there are obvious differences between the age groups as to what extent the strategy is followed, from a small difference between allocation of fixation time for Kanji and their translations in 7-year-olds to a much longer focus on the Kanji in adults.

Controlling: fixation time allocation during recognition

When turning to recall, the age-dependent differences in fixation time allocated to the Kanji is not only obvious, but also points to more complex differences in strategy use between younger and older participants. Allocating most attention to the Kanji during recall may be interpreted as a preference for a recognition strategy paralleling the constructive matching strategy as introduced in problem solving research. Looking relatively longer at the four answer alternatives, on the other hand, may be taken as an indication for a response elimination strategy (Bethell-Fox et al. 1984; Snow 1980; Vigneau et al. 2006). When comparing fixation times for the Kanji as the recognition cue and the sum of all answer alternatives (correct and incorrect for answerable items) 7-, and 9-year-old children allocated more fixation time to answer alternatives than the Kanji. Adults, on the other hand, allocated more attention to the Kanji when the items are answerable and about the same amount of time to both Kanji and answer alternatives for unanswerable items. When paralleling these results with those of Mitchum and Kelley (2010), the adults' behavior seems to indicate a constructive matching strategy, where first a recall attempt was made ("constructed") independent of retrieval cues and the result then was then compared to available answer alternatives. Checking the alternatives requires little time in this case. Children, on the other hand, adopted a strategy where checking the answer alternatives took more time than actually looking at the recognition cue. Such fixation time allocation would be expected when participants use a response elimination strategy. While for unanswerable items such an interpretation is straight forward, the results for answerable items provide additional support. The elimination part of the children's' strategy is accentuated, as there is no age difference in fixation time allocated to *correct* answers. This fixation time likely corresponds with the time needed in order to check and (in the case of correct recognition) verify the correct answers. The additional time children use to look at the alternatives in comparison to adults is therefore exclusively directed at *incorrect* answers, the candidates for response elimination.

When the difficulty of recall is increased by introducing unanswerable items, a change in the adults' allocation of fixation time indicates a shift towards a response elimination strategy. Contrary to results in the problem solving research, adult participants did not use more time before giving the answer. Instead, they reduced the time allocated to the Kanji in comparison to answerable items to a greater extent than children. And while fixation time allocation to the answer alternatives slightly increased, the reduction in fixation time on the Kanji results in about an equal amount of fixation time distributed to recognition cue and possible answers. Increasing the task difficulty seems to eliminate the adults' preference for a constructive matching strategy while the children's preference for response elimination is unchanged.

Mitchum and Kelley (2010) found that participants using a constructive matching strategy showed better confidence calibration. In our study, this strategy can only be attributed to adults, whose uncertainty monitoring is more accurate in comparison to younger participants. Following accessibility model a more serious attempt to recall an answer based on longer inspection of the Kanji is likely to generate more and better cues for metacognitive judgments (Koriat 1993, 1995). Additional information about the recall process would not only facilitate giving accurate CJs but is also likely to speed up the process of judging the chosen answer. The eye-tracking data supports this interpretation, as the adults' fixation time allocated toward the CJ scale before the explicit judgment is shorter than the children's.

Correlations between monitoring and controlling

Besides the finding of developmental differences for monitoring and controlling, correlations were computed to look into the reciprocal relationship between the two aspects of metacognition. In general, across all age groups, for learning and recognition and for implicit and explicit uncertainty monitoring alike, medium size correlations indicate that better monitoring is related to better controlling. However, within age groups this relation is only consistently evident in the youngest sample. For the 9-year-olds, only implicit CJs are related to controlling while for adults no such correlations hold whatsoever. These correlational patterns are inconsistent with our predictions. The adults' more adequate control behavior was expected to provide them with more high quality cues about recall processes. As the accessibility of these cues should facilitate the monitoring of progress and performance in the recognition task, the correlation between monitoring and control was expected to be highest in the adult sample. Based on these results, the impression arises, that, especially during early development of metacognitive skills, monitoring and control processes are closely connected. Thus 7- and 9-year-old children's implicit CJ are lower, when more fixation time is directed at the Kanji. Feedback from using this adequate strategy for recalling not currently accessible information is therefore related to more accurate uncertainty monitoring. On the other hand, and as predicted, longer inspection of incorrect alternatives, probably by inducing inappropriate memory cues, is related to children's overconfidence. Thus, the children's reliance on memory cues for monitoring processes seems to be a double-edged sword depending on which recall strategy is used. Behavior in line with constructive matching and response elimination provides access to cues for metacognitive monitoring. However, not all of these cues are helpful for accurate uncertainty monitoring. As the processes are thought to be intertwined, it may also be that accurate or inaccurate monitoring processes may influence the children's control behavior.

Moreover, in the 7-year-olds there is a close link to explicit monitoring as well. As the correlation between implicit and explicit CJs is very high, high correlations between explicit CJs and controlling variables could also be expected in the 9-year-olds. However, non-significant correlations, or a trend at best, may indicate a change as to how monitoring and controlling are connected when each of these metacognitive skills develop further.

For adults who constitute the group with the highest metacognitive skill, the weak or non-existent correlations, but nevertheless most adequate application of controlling strategies and uncertainty monitoring, does not concur with the hypotheses put forward in the introduction. Explanations about the basis of the adults' adequate recall related monitoring and controlling remain speculative. The results imply that other sources of information may have been tapped. The protracted growth of declarative metacognitive knowledge from childhood into adulthood may constitute such a resource and thus an intervening variable. With increasing age, explicit knowledge grows regarding what makes a task easy or difficult, when and how learning and retrieval strategies are advocated, and under what circumstances good or not so good performance can be expected (Schneider 2008). This knowledge may influence task related behavior and metacognitive judgments independent of each other. Therefore, while metacognitive processes may still profit from feedback concerning adjustments of attention allocation and vice versa, decisions and interpretations based on explicit knowledge may overshadow these more basic item dependent behaviors and evaluations. If such an interpretation were valid, weaker correlations between basic strategies of attention allocation and metacognitive monitoring would be expected for older participants. To verify these assumptions, information about declarative metacognitive knowledge would be needed, which may be taken as an interesting clue for future research.

Correlations between monitoring, controlling and recall performance

When turning to the central question of the relation between metacognitive variables and recognition, a similar pattern is discernible. Again, when looking at the correlations across the total sample, uncertainty monitoring and control, both during learning and recognition, are consistently and strongly related to recognition performance. Analyses within age groups underline the importance of metacognitive skills for young children's performance, while in the older samples only two out of three correlations are significant. For the youngest children, adequate control behavior during learning and recognition is related to performance. Contrary to our expectations, the correlation between explicit uncertainty monitoring and performance was significant only in this age group. Thus, the young children who are able to explicitly and accurately identify incorrect answers seem to be at an advantage in the Kanji task. More realistic evaluation of answers is related to the 7-year-olds learning and recall behavior, which seems to carry through to better recognition, whereas less realistic evaluation corresponds to worse recognition. Alternatively, good performance may make the few incorrect answers stand out, thus facilitating assignment of low CJ.

In the 9-year-olds, although their mean values fit in with those of the other samples concerning development of monitoring and performance, correlations between these variables were not significant. Explanations of this result are speculative, as like in younger children, there are significant correlations between performance and attention allocation indicative for strategy use. However, unlike in the younger sample, these relations are restricted to strategy use during recall only. In addition, the connection between monitoring and control is much less distinct. Thus, if the relation between monitoring and performance is control behavior-mediated, the correlation is less likely to be found in our data for the 9-year-olds.

Finally, in adults, only attention allocation towards incorrect alternatives and therefore the use of a response elimination strategy is negatively related to recognition. Therefore, it seems that the importance of the basic strategies mirrored in the simple allocation of attention decreases from younger to older participants, while overall recognition performance increases. Reasons for this unforeseen pattern of correlations remain speculative: Like for the correlations between metacognitive monitoring and controlling, growth of knowledge about memory relevant variables may be responsible for the shrinking impact of attention allocation strategies. Older participants may substitute these strategies with more efficient (e.g., associative) strategies where the quality of the elaboration and not merely the time allocated to the stimuli is crucial during learning and recall. On the other hand, as performance and the fixation time allocated to Kanji during learning and recall was high in all adults, the restricted variability of the variables may have undermined the correlational analyses. Still, like in the children, adults' long fixation time on incorrect alternatives and thus "bad" strategy use is negatively correlated with performance. As performance also correlates with adults' implicit uncertainty monitoring, basic processes of metacognitive monitoring and controlling as mirrored in allocation of fixation time are closely connected to task performance. Thus, they also remain important in later stages of metacognitive development.

Conclusion

To sum up, this study provided many interesting findings, and while some of these allowed the replication of results of an earlier study, others enable new insights. Although introducing eye-tracking technology was the basis for most of the variables analyzed, it was also thanks to the inclusion of an adult sample, that pronounced changes in strategy use, metacognitive monitoring, performance and their interrelationship across age groups were uncovered. Some of these findings are hardly surprising, like the verification of better metacognitive monitoring and control skills with increasing age. Yet, when correlational analyses between these skills and with performance are calculated, interesting age related differences become apparent. While the mentioned reasons for the low or missing correlations in the adult sample remain speculative, the high correlations between all monitoring, controlling, and performance variables in the youngest sample are mostly in line with our expectations. Using eye tracking technology allowed detecting wishful thinking and a response elimination strategy in this age group. As these monitoring biases and behaviors are related to each other and performance, there seems to be room for interventions to enhance monitoring and control skills in young children and their performance (at least in memory and recognition tasks with Kanji). Thus, measuring and/or providing metacognitive knowledge useful for solving memory tasks should be included in future studies of developing metacognitive monitoring and control skills.

Acknowledgments The current study was partially financed by a grant of the Swiss Science Foundation (SNF—Grant No. 100014–112492 to the second author). We wish to thank Esther Burri and Marion Stöckli for their help with the data collection and the teachers, children, and their parents for their cooperation and participation.

References

Allwood, C. M., Knutsson, J., & Granhag, P. A. (2006). Eyewitnesses under influence: how feedback affects the realism in confidence judgments. *Psychology, Crime & Law, 12*, 25–38.

Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning - componential and eye-movement analysis of geometric analogy performance. *Intelligence, 8*, 205–238.

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

Hauland, G. (2003). Measuring team situation awareness by means of eye movement data. In *Proceedings of HCI International 2003: Vol. 3* (pp. 230-234). Mahwah, NJ: Lawrence Erlbaum Associates.

Hayhoe, M. M., Shirbatava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*, 49–63.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–293). Oxford: Elsevier.

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

Kinnunen, R., & Vauras, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction, 5*, 143–165.

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

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*, 311–333.

Koriat, A. (1998). Metamemory: The feeling of knowing and its vagaries. In M. Sabourin, F. Craik, & M. Robert (Eds.), *Advances in psychological science, Vol. 2: Biological and cognitive aspects* (pp. 461–479). Hove: Psychology Press/Erlbaum (UK) Taylor & Francis.

Koriat, A. (2000). The feeling of knowing: some metatheoretical implications for consciousness and control. *Consciousness and Cognition, 9*, 149–171.

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

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

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009a). The easily learned, easily remembered heuristic in children. *Cognitive Development, 24*, 169–182.

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

Krebs, S. S., & Roebers, C. M. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *British Journal of Educational Psychology, 80*, 325–340.

Krebs, S. S., & Roebers, C. M. (2011). The impact of retrieval processes, age, general achievement level, and test scoring scheme for children's metacognitive monitoring and controlling. *Metacognition and Learning, 7*, 75–90.

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.

Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1084–1097.

Mitchum, A. L., & Kelley, C. M. (2010). Solve the problem first: constructive solution strategies can influence the accuracy of retrospective confidence judgments. *Journal of Experimental Psychology-Learning Memory and Cognition, 36*, 699–710.

Neggers, S. F. W., & Bekkering, H. (2002). Coordinated control of eye and hand movements in dynamic reaching. *Human movement Science, 21*, 349–376.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation, 26* (pp. 125–173). New York: Academic Press.

Olsen, A. (2012). The Tobii I-VT fixation filter - algorithm description. *Tobii I-VT Fixation Filter – whitepaper [White paper]*. Retrieved from http://www.tobii.com/Global/Analysis/Training/WhitePapers/Tobii_WhitePaper_TobiiIVTFixationFilter.pdf

Raven, J. C., Raven, J., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales: Section 4. Advanced progressive matrices, sets I and II.* Oxford: Oxford Psychologists Press.

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.

Roderer, T., & Roebers, C. M. (2009). Children's strategic regulation of memory accuracy. In M. Kelley (Ed.), *Applied memory* (pp. 148–205). Hauppauge: Nova Science Publishers.
Roderer, T., & Roebers, C. M. (2010). Explicit and implicit confidence judgments and developmental differences in metamemory: an eye-tracking approach. *Metacognition and Learning, 5*, 229–250.
Roderer, T., Krebs, S., Schmid, C., & Roebers, C. M. (2012). The role of executive control of attention and selective encoding for preschoolers' learning. *Infant and Child Development, 21*, 146–159.
Roebers, C. M. (2002). Confidence judgments in children's and adult's event recall and suggestibility. *Developmental Psychology, 38*, 1052–1067.
Roebers, C. M., Schmid, C., & Roderer, T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. *British Journal of Educational Psychology, 79*, 749–767.
Roebers, C. M., Schmid, C., & Roderer, T. (2010). Encoding strategies in primary school children: insights from an eye-tracking approach and the role of individual differences in attentional control. *The Journal of Genetic Psychology, 171*, 1–21.
Schacter, D. L. (1987). Implicit memory: history and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501–518.
Schneider, W. (1998). Performance prediction in young children: effects of skill, metacognition and wishful thinking. *Developmental Science, 1*, 291–297.
Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: major trends and implications for education. *Mind, Brain, and Education, 2*, 114–121.
Schneider, W., & Lockl, K. (2008). Procedural metacognition in children: Evidence for developmental trends. In J. Dunlosky & B. Bjork (Eds.), *A handbook of memory and metamemory* (pp. 391–409). New York: Taylor & Francis.
Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty* (2nd ed.). Mahwah: Lawrence Erlbaum Associates.
Schraw, G., & Roedel, T. D. (1994). Test difficulty and judgment bias. *Memory & Cognition, 22*, 63–69.
Snow, R. E. (1980). Aptitude processes. In R. E. Snow, P. A. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction. Vol. 1. Cognitive process analyses of aptitude* (pp. 27–63). Hillsdale: Erlbaum.
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.
Veenman, M. V. J. (2005). The assessment of metacognitive skills: what can be learned from multimethod designs? In C. Artelt & B. Moschner (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 75–97). Berlin: Waxmann.
Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: age and task differences. *Learning and Individual Differences, 15*, 159–176.
Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic influences on intelligence. *Intelligence, 34*, 261–272.
Visé, M., & Schneider, W. (2000). Determinants of performance prediction in kindergarten and school children: the importance of metacognitive and motivational factors. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 32*, 51–58.
Von der Linden, N., & Roebers, C. M. (2006). Developmental changes in uncertainty monitoring during an event recall task. *Metacognition and Learning, 1*, 213–228.