

Explicit and implicit confidence judgments and developmental differences in metamemory: an eye-tracking approach

Thomas Roderer · Claudia M. Roebers

Received: 21 October 2009 / Accepted: 29 June 2010 /
Published online: 24 July 2010
© Springer Science+Business Media, LLC 2010

Abstract In the present study, primary school children's ability to give accurate confidence judgments (CJ) was addressed, with a special focus on uncertainty monitoring. In order to investigate the effects of memory retrieval processes on monitoring judgments, item difficulty in a vocabulary learning task (Japanese symbols) was manipulated. Moreover, as a first exploratory step to uncover fast and retrieval bound (implicit) monitoring processes that take place before explicit CJ are openly reported, fixation time allocation during recognition and monitoring was recorded with an eye-tracking device. Results revealed developmental progression in uncertainty (but not in certainty) monitoring between the age of 7 and 9 years. Differences in CJ across levels of item difficulty point to a substantial impact of retrieval processes on 9-yr-olds' but not on 7-yr-olds' monitoring. Eye-tracking data revealed an overall bias towards medium and high CJ, and confirmed evidence on developmental progression in monitoring skills.

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

Introduction

Assessing one's confidence is important in many everyday life situations and requires explicit memory recollections. In some contexts, when different alternative answers come to an individual's mind, monitoring and selection of the best candidate answer is crucial for accurate responding. The ability to monitor one's memory retrieval is a main component of metamemory (Flavell and Wellman 1977) and monitoring is closely linked to metacognitive

The current study was partially financed by a grant of the Swiss National Science Foundation (SNSF—Grant No. 100014–112492 to the second author). We wish to thank Françoise Nor for her help with the data collection and the teachers, children, and their parents for their cooperation and participation. Thanks are also due to the Tobii Technology support team for their technical assistance.

T. Roderer (✉) · C. M. Roebers
Institute of Psychology, University of Bern, Muesmattstr. 45, CH-3000 Bern 9, Switzerland
e-mail: thomas.roderer@psy.unibe.ch

self-regulation processes, as reflected in control processes, like allocation of study time, selection of search strategy, termination of memory search, and selection of the best candidate answer (Nelson and Narens 1990). Metacognitive monitoring skills are associated with efficient control processes and academic as well as memory performance, independent of age and cognitive task (Renner and Renner 2001; Zimmerman and Kitsantas 1999; Maki 1998; Thiede et al. 2003; Dunlosky et al. 2005; Thiede 1999; Roebers et al. 2009; Schneider and Pressley 1997; Van der Stel and Veenman 2008; Veenman and Spaans 2005). Given the importance of metacognitive monitoring for cognitive performance in many domains including school performance, issues of metacognitive development still appear underrepresented in the empirical literature (for a recent review see Schneider and Lockl 2008). The present study focuses on developmental progression in confidence judgments (CJ; item-by-item performance post-dictions in a Japanese vocabulary learning paradigm) and applies an innovative technology (eye-tracking) aiming to gain further insights into these emerging skills.

Studies investigating the development of metamemory typically reveal protracted progression of self-monitoring skills during study and test over the elementary school years (Schneider and Lockl 2008). When required to predict future memory performance, younger children overestimate their performance (Visé and Schneider 2000), and their post-dictions of performance are not as accurate as those of older children (Von der Linden and Roebers 2006). Developmental progression takes place mainly with respect to uncertainty monitoring, that is, with respect to the adequate CJ for incorrect answers and answers to unanswerable questions (Roebers 2002; Howie and Roebers 2007). In other words, while the levels of confidence for correct answers is high in all age groups, older children and adults report lower confidence for answers that turn out to be incorrect. This results in overall better metacognitive discrimination between correct and incorrect answers with increasing age. Such improvements in uncertainty monitoring are mirrored by Pillow's studies on children's ability to differentiate between different cognitive states like guesses or drawing inferences that also improve substantially during middle childhood (Pillow et al. 2000; Pillow 2002; Pillow and Anderson 2006). Older children are more likely to identify guesses and to rate them as uncertain, while first graders are often very sure that their guess was correct. Thus, there are important advances in children's conceptual understanding and monitoring of mental processes over the elementary school years. Little, however, is known about what factors fuel these improvements.

The adult literature, in contrast, offers ample evidence for different situational influences on metacognitive judgments, with a strong and consistent impact of memory retrieval processes or item difficulty (i.e., easy *versus* difficult items) on judgments of certainty. For example, the degree of detailedness of a memory content (Arbuthnott et al. 2008), the experienced memory vividness (Robinson et al. 2000), cue familiarity and target retrievability (Metcalfe and Finn 2008), issues of consensuality (Koriat 2008), and trace accessibility (Koriat 1993, 1995) have been discussed to influence monitoring judgments via the involved memory retrieval processes. In children, however, only very few studies focusing on performance predictions have documented an influence of item difficulty and thus strength of underlying memory representation on monitoring (Koriat et al. 2009; Koriat and Shitzer-Reichert 2002; Lockl and Schneider 2003). The present study aims to extend these findings by systematically comparing CJ for easy, difficult, and unanswerable items. These comparisons allow investigating the relationship between monitoring and retrieval, exploring the impact of retrieval for developmental progression in monitoring, as well as looking at metacognitive development in relation to the underlying knowledge base. In fact, because gains in knowledge base have been shown to be the most important factor fuelling memory development during the primary school years (Schneider and Pressley

1997), evidence on a substantial impact of retrieval processes on metamemorial development would add to our understanding of how and through which processes the development of memory and monitoring comes about.

Based on the assumption that monitoring processes are retrieval-based cognitive operations (Koriat 1993, 1995; Metcalfe and Finn 2008; Schraw and Roedel 1994), Koriat (1998, 2000) locates metacognitive judgments at the interface between implicit and explicit processes. Other authors even claim that metacognition and implicit memory can hardly be separated (Kelley and Jacoby 1996). Following Schacter's (1987) distinction of implicit-explicit memory, 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. Such a definition is in accordance with an implicit-automatic mode of metacognitive judgments based on mnemonic cues like fluency of processing or the ease with which information is accessed (Koriat 2007). As Koriat has pointed out, calculations of these cues are performed below awareness and may influence behavior directly and automatically, but can also give rise to metacognitive experiences. These mnemonic cues can be used for explicit metacognitive judgments if the metacognitive experiences exceed the threshold for conscious awareness.

From this conceptual point of view the question arises whether or to what extent the classical measures of metacognitive monitoring capture such implicit processes. The development of measurements tapping metacognitive monitoring processes that are more closely related to retrieval processes than the classical judgments thus appears promising and challenging, both from a theoretical and methodological point of view. A comparison between classical metacognitive judgments, those that participants openly report, and indicators of implicit metacognitive processes, those that are closely linked to ongoing retrieval processes, may increase our understanding about the involved cognitive processes, the interplay between retrieval and monitoring, and factors fuelling developmental progression in this domain.

In this paper, we report an effort to measure implicit metacognitive processes and compare these with classical, openly reported CJ. Thereby, we define implicit metacognitive processes as fast and retrieval-bound cognitive operations that take place before individuals openly report how sure they are that their answer is correct. These implicit processes take place early in the course of information processing, in a close and reciprocal relationship with retrieval processes but may influence later reported classical metacognitive judgments. Assuming that these implicit metacognitive processes are closely linked to memory retrieval (Koriat 2000, 2007), assessment attempts have to be realized while individuals engage in cognitive processes closely linked to retrieval. Exploration of implicit metamemory thus requires fine-grained assessment methods in order to capture these fast and retrieval-bound monitoring processes (Rhodes and Kelley 2005).

Reviewing the literature on measurements of implicit cognitive processes, several fields of child development research come into sight in which emerging, albeit still implicit skills are assessed before measurement of the fully developed skills is possible. For example, in the domain of understanding false belief (Clements and Perner 1994) and causality (Leslie and Keeble 1987), object permanence and identity (Baillargeon 1987; Spelke and Kestenbaum 1986), category discrimination (Pauen 2002), and simple mathematical calculations (Wynn 1992) implicit skills are assumed to be present before explicit and more fully developed proficiencies are measurable. In all these studies, young children's gaze patterns have been used as indicators of implicit, emerging abilities. Applying this approach for measuring implicit and retrieval-bound metacognitive monitoring appears promising for the field of metacognitive development. For the present study, we therefore drew inspiration from these research fields on emerging cognitive

abilities using gaze patterns (fixation time) as indicator for implicit cognitive processes, following Veenman's suggestion (2005). He proposed the use of eye-tracking technology as a promising tool for studying metacognitive processes.

Eye-tracking allows gathering information on the allocation of attention in terms of what, where, and how long individuals attend to. This technology can thus provide fine-grained data on cognitive processes that may otherwise not easily be accessible (e.g., through open reporting; Scheiter and Van Gog 2009). It is therefore not surprising that eye-tracking approaches are increasingly being used in various domains of cognitive psychology (Aslin and McMurray 2004). Because a comprehensive review of the eye-tracking literature and measures (saccades, papillary dilation, tracking, and fixations) is far beyond the scope of this paper (see Karatekin 2007, for a detailed review), we will focus on allocation of fixation time as classical measure during scene perception and for attentional and other cognitive processes as being used in the present study.

Although an individual's fixation itself does not uniquely specify the cognitive operation being performed, there is a close and consistent link between an individual's fixation and the occurrence of cognitive processes (Hayhoe 2004). In other words, fixations can be considered as reliable indicators of the focus of active information processing with the advantage that they are now being accepted as a measure of implicit abilities and performance not necessarily involving conscious processes (Hayhoe et al. 1998; Karatekin 2007). Moreover, there is good correspondence between eye-tracking data and behavioral measures of attention allocation, not only in adults but also in children (Corbetta 1998; Karatekin 2004). In contrast to saccades, for example, allocation of fixation time mirrors top-down processes that are controlled by the individual's ongoing cognitive processes, rather than by the visual characteristics of the presented stimuli (Hayhoe 2004; Karatekin 2007). Importantly and in support of our approach, eye-tracking technology has already been used as an indicator of monitoring processes in Kinnunen and Vauras' studies while reading (i.e., changes in reading speed, look-backs; Kinnunen and Vauras 1995; Kinnunen et al. 1998). Therefore, we expect to gain interesting insights into monitoring processes by applying an eye-tracking approach (with a focus on fixation time allocation) in a vocabulary learning task. By analyzing the distribution and duration of fixations we aim to explore implicit monitoring processes taking place during retrieval before individuals express explicit CJ and to compare these measures of implicit monitoring with classical, explicit, openly reported CJ.

Precisely, by looking at an early stage of retrieval-bound monitoring processes we expect to detect implicit metamemorial abilities in primary school children. As results about early cognitive abilities in infants and young children suggest (Hayhoe 2004), we expect these basic and fast monitoring processes to be more comparable between younger and older children than are explicit, openly-reported CJ for which younger children's judgments are more likely to be influenced by motivational, social, and situational demands and processes compared to older children (Schneider 1998; Schneider and Lockl 2008; Roebers et al. 2007). Our decision to use a vocabulary learning task in this study on implicit and explicit metamemory monitoring was influenced by several aspects related to metamemory and learning: Metamemory has been shown to affect word learning strategies and transfer of word knowledge (Nagy and Scott 2000; Wang et al. 2009). Japanese symbols paired with an illustration of the meaning were used in the present study in order to avoid prior knowledge effects (Körkel 1987), to reduce the effect of reading skills, and to render the task motivating and interesting for the participating children.

Taken together, aims of the present study are twofold: First, we want to continue the study of metacognitive monitoring abilities in children and the developmental progression

thereof with a special emphasis on uncertainty monitoring. This is done by systematically comparing monitoring for easy, difficult and unanswerable (not learned) Japanese symbols in order to assess the impact of retrieval characteristics for monitoring. Second, by applying eye-tracking technology while children give their CJ after recognition, we intend to make a first step in exploring assessment methods for implicit monitoring processes. We anticipate that measures of implicit metacognitive monitoring in children may mirror relatively adequate monitoring skills, especially in the younger age group, and that they are related, albeit not identical to explicit, openly reported CJ.

Method

Participants

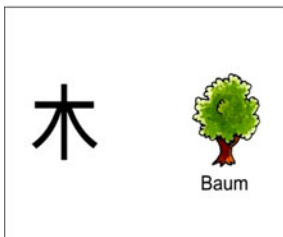
Seventy-four children (42% female) completed the study. Due to missing data caused by technical failures or the structure of the children's answers (e.g. not using the whole range of confidence judgments, no correct/ incorrect answers) sample sizes in the analyses may vary slightly. For the analyses reported below, data of $N=37$ 7-year-olds ($SD=4.7$ months) and $N=37$ 9-year-olds ($SD=4.3$ months) were included. Participants were recruited from public schools in the region of Bern and Fribourg, Switzerland. For all participants written informed consent was obtained from the parents.

Material

In the memory task which was used for assessing recognition and procedural metamemory, children were asked to learn Japanese symbols (Kanji), therefore we will subsequently refer to this task as the Kanji-Task. The task consisted of two sets, each containing 10 different Kanji—picture associations.

In the learning phase, each symbol was presented on the left hand side of a computer screen together with a pictorial representation of its meaning on the right hand side. The pictures consisted of colored line-drawings including a label of the object's name written out beneath each picture (see Fig. 1a; see Appendix for the list of items). A large number of Kanji—picture associations were piloted previously with 25 children (14 girls, 11 boys) in third grade. For the present study Kanji symbols with an item difficulty of .80 (easy items) or .51 (difficult items) were selected from this pool. In the learning phase both sets of items consisted of 10 items, half of which were easy and the other half difficult items.

1a) Learning Slide



1b) Recognition Slide



1c) CJ Slide



Fig. 1 Examples of slides used in the three phases of the Kanji-task: Learning, recognition and CJ

In the recognition phase, the Kanjis from the learning phase were presented on the left side, together with four alternative answers on the right side. Items were presented on a remote non-invasive eye-tracking system (Tobii, 1750) with a 17 in (43 cm) computer screen. The four alternatives for the easy and difficult Kanjis consisted of two new pictures of objects and two pictures that had been shown during learning, including the correct answer (see Fig. 1b). Alternatives of new items were included in order to increase the total number of possible answers and to reduce inflation of correct answers through guessing. The position on the slide and the combination of learned and new pictures as alternatives was counterbalanced. All alternatives, new ones or from the learning trial, were used twice. In the recognition phase, only four out of five easy and five difficult Kanjis from the learning phase were included. Additionally, four new and therefore unanswerable Kanjis were introduced, resulting in a total of 12 items for recognition. When asked at the end of the experiment if they had noticed something odd about the Kanji-Task none of the participants claimed to have realized the difference in the number of the 10 Kanjis in the learning and two additional Kanjis in the recognition phase. In addition, only few participants stated to have noticed that some recognition items were unanswerable, but only after they had been debriefed by the experimenter.

In the metamemory phase, the 12 items (Kanjis) from recognition were again presented on the left side of the screen. On the right side, a confidence scale consisting of five yellow smiley faces was shown in an upright line (see Fig. 1c).

Procedure

Children were individually tested by the author and a trained research assistant in a quiet room at their school. After a short rapport building phase, children were asked if they had some experience with Japanese Kanji symbols and were then introduced to the task using a laptop computer. It was explained that the Japanese write in symbols that 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 of that symbol accompanied with a label were shown to the children (see Fig. 1a).

Practice trial

After these explanations, children were given the opportunity to practice learning and recognition in a practice trial. Children were shown four slides with Kanji symbols together with their illustrated meaning. Each slide was presented for 5 s. After learning, 4 recognition slides were presented. Children answered by naming one out of four alternative objects they assumed to be the correct translation for the Kanji (see Fig. 1b). If they were not able to recall the correct answer, they were instructed to guess (forced report).

After practice, children were told that the experimenter was interested in finding out which of the symbols were easy and which were hard to learn and that therefore, they would later be asked how confident they are regarding the correctness of their answers. For these confidence judgments, a five point scale was introduced (see Fig. 1c). Each of the five faces was introduced as standing for an indication of how sure or unsure one is that the given answer is correct. The experimenter demonstrated how to use the scale on the basis of easy and hard to answer questions. If an answer had to be guessed, the experimenter suggested choosing the confidence rating at the lower end of the scale to indicate uncertainty. After these explanations, children were trained to the use of the confidence scale on the basis of a short practice: The four Kanjis from the learning and recognition

phase were shown together with the confidence scale. The children's answer from the recognition phase was repeated and they were asked to indicate how sure they are that this particular answer was correct by using the confidence scale. Children who did not use the whole range of the confidence scale during practice were asked where they would have pointed to if they were either sure or unsure that their answer was right and corrected if necessary. All children learned the use of the confidence scale with ease. They were also trained in a cautious way of pointing by only moving their arm and not leaning forward in order to minimize eye-tracking problems later in the test trials, as eye-tracking is impaired by head movements, especially if the eyes move out of range of the eye-tracking camera. (In case that the choice of one of the smileys was not clearly recognizable, children were given the opportunity to clarify before the next confidence judgment by pointing at a sheet put forward by the experimenter.)

Test trials

Before starting with each learning phase, a plate containing 20 pictures was presented. These 20 pictures were composed of the 10 translations of the learned Kanjis and 10 new pictures. Children 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 Kanjis on the laptop computer, each presented for 5 s. Presentation time of the recognition slides was not fixed, as each participant was given the time he or she needed to select an answer. After each answer the experimenter briefly paused the presentation and marked the given answer on a protocol sheet. Presentation time of the confidence judgment slides was also not fixed in order to give participants enough time to decide how confident they felt concerning their answer. The presentation of the confidence judgment slides was stopped as soon as children decisively pointed at one of the smileys. In each of the two trials, a backward digit task (Pickering and Gathercole 2001) was inserted as a filler task between learning and recognition to prevent rehearsal or the use of other recall strategies.

After the filler task, children were seated in a comfortable armchair (leaning slightly backwards with the head leaning on the back-rest of the chair) in front of the eye-tracking system hidden in a 17 in computer screen. They were not aware of the eye-tracking device. An adjustable mounting allowed for optimal placement of the screen with a distance of 60 cm from the children's eyes. Next, a five-point eye tracker calibration was done disguised as a very simple task, while keeping the child unaware of the task's purpose. With very few exceptions, calibration was successful within less than 20 s.

After successful calibration, 12 recognition slides were presented and children chose one of the four objects as correct translation for the item (easy, difficult and unanswerable Kanjis). Between presentations of the recognition slides, a picture of a star was shown in order to ensure fixation on the center of the screen before presentation of the next slide. Immediately after recognition, the metamemory phase and a short repetition of confidence judgment instructions followed. The same Kanjis as in recognition were shown, the experimenter reported the child's answer for each Kanji, and children indicated how sure they were that their answer was correct by pointing at one smiley. After completing the first test trial, a second filler task lasting approximately 5 min was introduced where children played a short computer game of feeding a fish (in fact an adapted form of the Eriksen-Flanker task (Eriksen and Eriksen 1974) in order to minimize interference from the first trial. For the second test trial, the second set of Kanji symbols and their translation was applied. The

order of the two Kanji sets was counterbalanced across children. After completing the two trials of the Kanji-Task, children were debriefed about the unanswerable Kanjis, praised for the good performance, thanked for their participation, given a small gift, and escorted back to their classroom.

Results

In the following sections, we will first describe results for recognition and explicit metacognitive judgments. Then, we will present analyses on the eye-tracking data for both recognition and metamemory. Finally, we will report about measurements of implicit metamemorial judgments and analyze relations between measurements of implicit and explicit metamemory judgments. Preliminary analyses did not reveal sex differences in any of the variables addressed; data were therefore collapsed across gender. As estimators of effect sizes partial η^2 - values are reported to allow direct comparisons of effects across measurements.

Recall

Overall, children chose the correct answer for 52% of the Kanjis ($SD=18\%$; 7-year-olds: $M=44\%$, $SD=14\%$; 9-year-olds: $M=60\%$, $SD=17\%$). Recognition rates for easy Kanjis were 68% ($SD=21\%$), and 35% ($SD=22\%$) for difficult Kanjis. The mean numbers of correct answers for both trials (max. 8 for easy and difficult items, respectively) were entered into an analysis of variance (ANOVA) with item difficulty (easy vs. difficult items) as within-subject and age as between-subjects factor. Results revealed significant main effects of item difficulty, $F(1, 72)=154.25$, $p<.001$, $\eta_p^2=.68$, [easy items > difficult items], and age $F(1, 73)=18.22$, $p<.001$, $\eta_p^2=.20$ [9-year-olds >7-year-olds]. There was no significant interaction between item difficulty and age.

Confidence judgments (CJ)

The upper half of Table 1 shows the mean CJ openly and explicitly given by participants, as a function of correctness of recognition, item difficulty and age. An ANOVA on the CJ for

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

Item difficulty	7-year-olds			9-year-olds		
	Easy	Difficult	Unanswerable	Easy	Difficult	Unanswerable
Explicit confidence judgments						
Correct answers	4.68 (0.38)	4.02 (1.21)		4.47 (0.39)	4.10 (0.97)	
Incorrect answers	3.75 (0.79)	3.78 (0.87)		3.10 (1.08)	2.96 (0.91)	
Total	4.28 (0.51)	3.76 (0.91)	3.66 (0.96)	4.20 (0.51)	3.54 (0.77)	2.83 (0.71)
Implicit confidence judgments						
Correct answers	3.84 (0.73)	3.37 (1.09)		3.89 (0.43)	3.78 (0.80)	
Incorrect answers	3.66 (0.84)	3.71 (0.60)		3.50 (0.65)	3.36 (0.51)	
Total	3.74 (0.62)	3.54 (0.61)	3.43 (0.65)	3.77 (0.48)	3.41 (0.64)	3.15 (0.50)

Standard deviations are in parentheses

answerable (easy and difficult) items with correctness of answer and item difficulty as within-subject factors and age as between-subjects factor was conducted and revealed significant main effects for correctness of answer, $F(1, 58)=112.52, p<.001, \eta_p^2=.66$ [correct answers > incorrect answers], item difficulty, $F(1, 58)=7.61, p<.01, \eta_p^2=.12$ [easy items > difficult items], and age, $F(1, 58)=6.16, p<.02, \eta_p^2=.02$ [7-year-olds > 9-year-olds], as well as significant interaction effects between correctness of answer and item difficulty, $F(1, 58)=8.60, p<.01, \eta_p^2=.13$, and correctness of answer and age, $F(1, 58)=14.96, p<.001, \eta_p^2=.21$. Follow-up analyses on the interaction between correctness of answer and age revealed that it was due to the fact that children's CJ for correct answers did not differ between the two age groups, while 9-year-olds' CJ for incorrect answers were lower than the 7-year-olds', indicating improved monitoring of uncertainty with increasing age. Detailed analyses on the interaction between correctness of answer and item difficulty revealed that while item difficulty had no impact on CJ for incorrect answers, confidence for correct answers of difficult items were lower than for easy items. In an additional analysis, we integrated the CJ of the unanswerable questions and compared all three levels of item difficulty as within-subject factor and age groups as between-subjects factor in a repeated measure ANOVA. Since there are no correct answers for unanswerable questions there was no within-subject factor of correctness of answer and judgments for correct and incorrect answers were collapsed for the easy and difficult items. Results revealed significant main effects of item difficulty, $F(2, 71)=60.50, p<.001, \eta_p^2=.63$ [CJ for easy items > CJ for difficult items > CJ for unanswerable items], and a main effect of age, $F(1, 72)=6.63, p=.01, \eta_p^2=.08$ [7-year-olds > 9-year-olds]. However, the interaction between item difficulty and age, $F(2, 71)=12.10, p<.001, \eta_p^2=.25$, mainly shows a differential impact of unanswerable questions on the CJ in the two age groups. As can be seen in Table 1, there was no difference between the age groups in the CJ for easy and difficult items. However, 9-year-olds gave even lower CJ for unanswerable items while the 7-year-olds' judgments for difficult and unanswerable are similar. Therefore, the interaction originates from lower CJ for unanswerable items and thus from sufficient uncertainty monitoring in 9-year-old children only.

In order to gain further insights into children's emerging uncertainty monitoring skills, we additionally compared children's CJ for incorrect answerable and unanswerable items. Results indicated no difference between CJ for incorrect answerable and unanswerable items in 7-year-olds but significantly lower CJ for unanswerable items than for incorrectly answered answerable items in 9-year-old children, $t(35)=2.07, p<.05$. Thus, 9-year-olds but not 7-year-olds were able to differentiate between incorrectly recalled items and guesses.

Eye-tracking

In order to identify the location and duration of fixations as indicator for ongoing processing of information we applied a filter to the gaze data as is done in other studies (Henderson and Hollingworth 1998; Hauland 2003; Kinnunen and Vauras 1995). In the present study, all gaze data in an area of 2.5° of visual angle with a minimal duration of 100 ms was counted as 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 cumulating the fixation times for the different areas of interest over all relevant trials and then dividing by the total number trials (24 for

analyses of all trials, 8 for analyses of separate levels of item difficulty). On average, 7-year-olds fixation times were 1.95 s ($SD=0.70$) for the Kanji, 0.79 s ($SD=0.28$) for the correct answer alternative, and 0.72 s ($SD=0.25$) for each incorrect answer alternative on the recognition slides. Seven-year-olds fixation times on the CJ slides were on average 2.33 s ($SD=0.72$) for the Kanji part and 1.02 s ($SD=0.38$) for the confidence scale part of the slide. Nine-year-olds' average fixation times were 2.35 s ($SD=0.60$) for the Kanji, 0.65 s ($SD=0.19$) for the correct answer alternative and 0.65 s ($SD=0.27$) for each incorrect answer alternative during recognition. During monitoring, that is, on the CJ slides, 9-year-olds' fixations lasted 2.16 s ($SD=0.51$) for the Kanji part and 0.82 s ($SD=0.34$) for the confidence scale.

As the presentation time of the recognition and confidence judgment slides was adjusted to the time the participants needed to produce a response, percentages of the fixation time for each area of interest in relation to the total fixation time on the slides were calculated as dependent variable. See Fig. 2 for examples of gaze paths and fixations of one 9-year old during recall (recognition slides) and monitoring (CJ slides) for all three levels of item difficulty.

Fixation time allocation during recall

In Table 2, the percentages of fixation time allocated to Kanjis, the correct answers and the incorrect alternatives are presented. Repeated measure ANOVAS were performed to compare percentages of fixation time allocation across all three levels of item difficulty as a

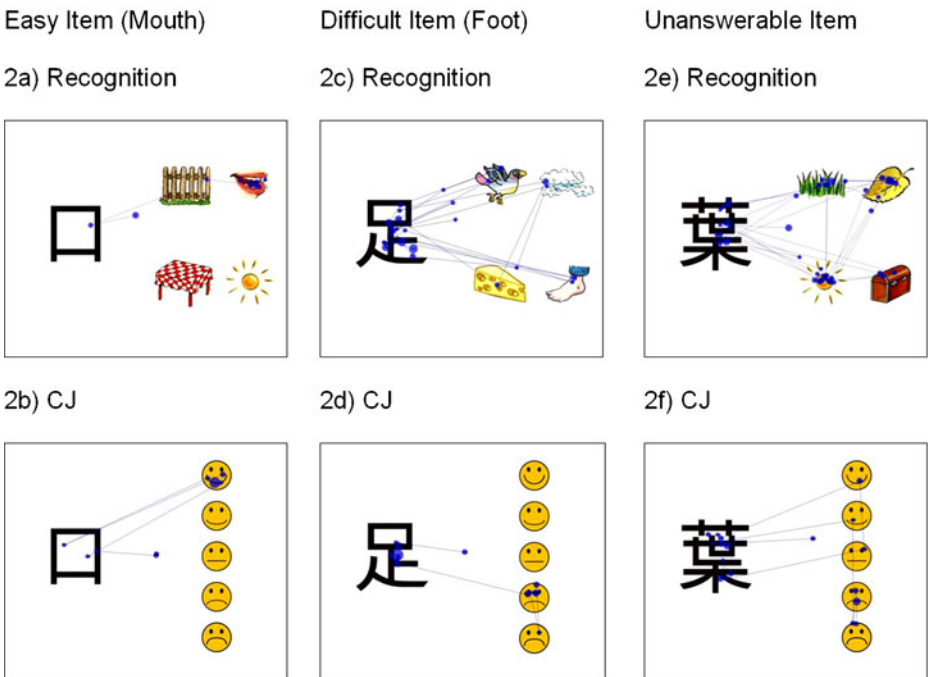


Fig. 2 Examples of a third grader's gaze paths and fixations for recognition and CJ for three levels of item difficulty

Table 2 Percentages of fixation time for recognition and confidence judgment slides as a function of age and item difficulty

Difficulty	7-year-olds			9-year-olds		
	Easy	Difficult	Unanswerable	Easy	Difficult	Unanswerable
Recognition slides						
Kanji	39.39 (8.93)	37.74 (8.87)	34.60 (9.97)	44.84 (9.48)	47.25 (10.14)	42.58 (8.65)
Correct answer	18.30 (3.87)	13.03 (4.26)		15.93 (3.71)	11.82 (2.81)	
Incorrect alternatives	11.78 (2.96)	14.33 (2.27)	14.36 (2.65)	10.76 (3.23)	11.71 (3.03)	12.45 (2.39)
Confidence judgment slides						
Kanji	58.69 (12.27)	64.76 (8.91)	65.06 (12.34)	62.39 (16.39)	70.52 (8.43)	65.90 (7.67)
Confidence scale	30.80 (10.86)	26.12 (8.20)	26.02 (11.35)	25.46 (8.48)	21.61 (8.07)	26.50 (7.73)

Standard deviations are in parentheses; percentages of fixation time do not add up to 100% as some fixations are located on the slide outside of the specified areas of interest

function of participants' age. For the fixation time on the Kanji symbols, the significant main effects of item difficulty, $F(2, 71)=14.11, p<.001, \eta_p^2=.28$ [easy items, difficult items > unanswerable items], and age, $F(1, 72)=14.90, p<.001, \eta_p^2=.17$ [9-year-olds > 7-year-olds], were qualified by a significant interaction between item difficulty and age, $F(2, 71)=3.7, p=.03, \eta_p^2=.09$. While the children of both age groups allocated relatively more fixation time to easy than to unanswerable items, the older children's relative fixation time was larger for difficult than for easy items, in contrast to the younger children's relative fixation time for difficult items, which was lower compared to that of easy items.

Together with the Kanjis, four alternative answers were presented (one correct and three incorrect alternatives; see Fig. 1). The analysis of variance with percentages of fixation time allocated to correct answers during recall (easy and difficult items) as dependent measure revealed a significant effect of item difficulty, $F(2, 71)=71.28, p>.001, \eta_p^2=.50$ [easy items > difficult items], and age, $F(1, 72)=6.57, p=.01, \eta_p^2=.08$ [7-year-olds > 9-year-olds], but no interaction between item difficulty and age.

For the analysis of differences between the percentage of fixation time allocated to the incorrect alternatives of easy, difficult, and unanswerable items, the average fixation time across the three incorrect alternatives were used. Besides a significant main effect of item difficulty, $F(2, 71)=28.00, p<.001, \eta_p^2=.44$ [easy items < difficult items < unanswerable items], results also revealed an age effect, $F(1, 72)=11.30, p=.001, \eta_p^2=.14$ [7-year-olds > 9-year-olds], as well as an interaction between item difficulty and age, $F(2, 71)=4.03, p=.02, \eta_p^2=.10$. The interaction between age and item difficulty was due to the fact that in 9-year-olds, relative fixation time to the alternatives differed between all three levels of item difficulty whereas younger children's fixation times only increased from easy to difficult items but did not differ between difficult and unanswerable items.

Fixation time allocation during monitoring

One major goal of using the eye-tracking device during individuals' recall and monitoring was to explore the possibility that individuals engage in monitoring before openly giving CJ to the experimenter and that such early, fast, and retrieval-bound monitoring processes

differ from later, openly given CJ. In order to gain such eye-tracking data, fixations that could be attributed to either orientation towards the confidence scale or that were part of explicit judgments given at the end of the slide presentation were deleted. For example, eye-tracking data on the confidence scale before the first fixation on the Kanji on the left side were counted as orientation on the screen and not considered further, as CJ require at least a reminder of the to be recalled and monitored stimulus. In the present study, pointing at the confidence scale associated with giving explicit judgments has a major impact on eye-movement patterns (e.g. longer fixations). To obtain fixation data associated with monitoring processes and not contaminated by eye-hand coordination, eye-tracking data for explicit judgments were excluded from the analyses. This was done based on records of explicit judgments and based on data patterns indicative for explicit judgments. In the lower part of Table 2 the mean percentages of fixation time allocated to the Kanji part and the confidence scale part are presented. As was done with the recognition slides, ANOVAS with repeated measures were calculated for the Kanji and the confidence scale separately with item difficulty as within-subject factor and age as between-subjects factor. For the Kanji, a main effect of item difficulty, $F(2, 71)=12.26, p<.001, \eta_p^2=.26$ [easy items < unanswerable items < difficult items], as well as an interaction between item difficulty and age, $F(2, 71)=3.79, p=.03, \eta_p^2=.10$ was found, but no main effect of age. The interaction implies that while the relative fixation time was comparable for easy and unanswerable items between the two age groups, 9-year-old's allocation of fixation time to difficult Kanjis was longer than 7-year-old's.

Next, an ANOVA for the fixation time on the confidence scale part of the slides was conducted with item difficulty as well as the five separate CJ (1–5 as categories) as within-subject factors and age as between-subjects factor. The main effects of item difficulty, $F(2, 71)=9.79, p<.001, \eta_p^2=.22$, [easy items, unanswerable items > difficult items], of CJ, $F(4, 69)=106.00, p<.001, \eta_p^2=.86$, [4, 3>5, 2>1] but not of age were significant. However, there was a significant interaction between item difficulty and age, $F(2, 71)=3.44, p=.04, \eta_p^2=.09$. While there were no differences between the age groups for unanswerable items in percentages of fixation time allocated to the CJ scale, the fixation time children allocated in case of easy and difficult items was significantly lower in 7- than in 9-year-olds. The significant interaction between item difficulty and CJ, $F(8, 65)=6.73, p<.001, \eta_p^2=.45$, is a result of relative more fixation time allocated to the higher CJ (4 and 5) for easy as compared difficult and unanswerable items (see Fig. 3).

Implicit CJ

Based on the eye-tracking data, measurements for implicit CJ were calculated. For the following analyses, we used the confidence judgment “category” that attracted a maximum of fixation time during confidence scale presentation as a measure for implicit metacognitive judgments.¹ Such a confidence judgment is interpreted as a fast and implicit metamemorial evaluation of an answer during or only few milliseconds after retrieval and before explicit CJ are given. However, this approach resulted in the loss of some implicit CJ, as for some items (e.g. very easy items), children directly looked only at one CJ, the one they also openly reported immediately. In such cases no data indicative of any implicit

¹ Other measures of implicit CJ 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.

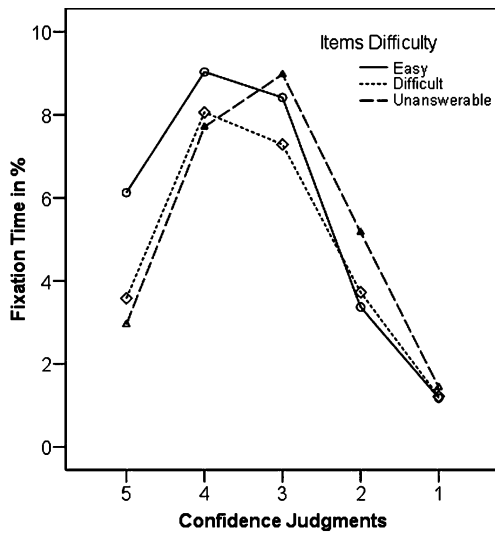


Fig. 3 Percent of fixation time for the confidence scale: Confidence judgments as a function of item difficulty

CJ is available. On average, there were $M=5.53$ ($SD=3.22$) missing values of the 24 items per participant.

In a first step, we present basic analyses that parallel those for explicit CJ (see Table 1 for a comparison of values for explicit and implicit CJ). The first ANOVA included correctness of answer and item difficulty as within-subject factors and age as between-subjects factor. No main effects but a significant interaction between correctness of answer and age, $F(1, 44)=4.20$, $p<.05$, $\eta_p^2=.09$, a non-significant trend for an interaction between correctness of answer and item difficulty, $F(1, 44)=3.41$, $p=.07$, $\eta_p^2=.07$, and a three-way interaction between correctness of answer, item difficulty and age, $F(1, 44)=4.26$, $p<.05$, $\eta_p^2=.09$, were found.

To follow up on the significant three-way interaction, two separate ANOVAS on CJ for easy and difficult items were calculated. The ANOVA on CJ for easy items showed a main effect of correctness of answer, $F(1, 57)=8.48$, $p<.01$, $\eta_p^2=.13$ [CJ of correct answers > CJ of incorrect answers]. For the difficult items, however, there was no main effect but an interaction between correctness of answer and age, $F(1, 55)=5.08$, $p=.03$, $\eta_p^2=.09$. Seven-year-olds' implicit CJ were higher for incorrect than for correct answers. In the 9-year-olds, implicit judgments were still in the expected direction, that is, implicit judgments on recognition of difficult items were significantly lower for incorrect than for correct answers, mirroring more advanced implicit monitoring skills in 9-year-olds compared to 7-year-olds when the items were difficult.

When comparing implicit CJ between the two age groups with all three levels of item difficulty as a within-subject factor, results revealed a significant main effect of item difficulty, $F(1, 69)=27.10$, $p<.001$, $\eta_p^2=.44$ [CJ for easy items > CJ for difficult items > CJ for unanswerable items]. The significant interaction between item difficulty and age points to a stronger effect of item difficulty on 9-year-olds' compared to 7-year-olds' implicit CJ with more pronounced declines in confidence between easy and difficult items in older compared to younger children.

Correlations between implicit and explicit CJ

In the next step, one-tailed correlations between implicit and explicit judgments were calculated. Collapsing across all levels of item difficulty and age, there was a high correlation between implicit and explicit CJ, $r=.70, p<.001$. Correlations were highest for unanswerable items, $r=.70, p<.001$, followed by easy items, $r=.60, p<.001$, and were lowest for difficult items, $r=.52, p<.001$. Separate correlations for correct and incorrect answers of easy and difficult items revealed higher correlations for incorrect answers $r=.52, p<.001$, than for correct answers within easy items, $r=.30, p<.01$, while there was no difference between correlations for incorrect, $r=.53, p<.001$, and correct, $r=.53, p<.001$, answers within the difficult items. Correlations between implicit and explicit CJ tended to be stronger for 7-year-olds, $r=.77, p<.001$, than for 9-year-olds, $r=.55, p<.001$, a pattern that was confirmed when the different levels of item difficulty were examined separately.

Discussion

The present study aimed to investigate children's metacognitive monitoring skills and to explore factors fuelling developmental progression in this domain. Special emphasis was put on uncertainty monitoring and it was hypothesized that retrieval processes and characteristics thereof substantially influence primary school children's monitoring judgments, especially when retrieval is difficult or ambiguous (Koriat and Ackerman 2010; Koriat et al. 2009; Koriat and Shitzer-Reichert 2002; Lockl and Schneider 2003). Therefore, easy, difficult, and unanswerable items were included in a vocabulary learning task and 7- and 9-year-olds' monitoring judgments across these 3 classes of stimuli were contrasted with each other.

Importantly, recall data revealed that our task was suitable for both age groups with 44% to 60% correct answers across the two learning trails. For reliably studying monitoring processes, both sufficient correct and incorrect answers have to result from the paradigm. The frequencies prove that necessary prerequisites for subsequent data analyses were given. Additionally, significant recall differences between easy and difficult items in absence of an age-by-item-difficulty interaction underline the fact that the task was suitable for the age range covered by the present study.

Confirming our expectations, we found evidence for improvements in children's monitoring skills with increasing age and these improvements were—without exception—located in the domain of uncertainty monitoring. In other words, older children proved to be better at metacognitively distinguishing between correct and incorrect recall compared to younger children. Moreover, it was only in the group of 9-yr-olds that monitoring judgments substantially differentiated between difficult and unanswerable items. At the same time, no age differences were found concerning participants' monitoring of correct recall confirming previous studies on metacognitive development (Roebers 2002; Howie and Roebers 2007; Von der Linden and Roebers 2006). These findings are important as accurate monitoring of uncertainty promotes efficient control of learning and correction of errors. Uncovering and precisely describing existing metacognitive deficits in younger compared to older children is important from both a theoretical and an applied perspective. Theoretically, knowledge about the developmental timetables of different aspects of monitoring may lead to a more profound

understanding of the underlying concept. For applications in the educational field, suggestions for fostering metacognitive development or interventions can be inferred from these results.

Literature on adults' metacognition consistently substantiates an impact of memory retrieval on monitoring accuracy. Assuming a bidirectional and close relationship between monitoring and retrieval processes (Dunlosky and Bjork 2008; Koriat 2007; Rhodes and Kelley 2005) studying the developmental origins or the emergence of this relationship appears to be important. Our study revealed a significant influence of retrieval processes on children's openly reported confidence judgments. In general, CJ for easy items were higher than for difficult items suggesting that children took memory and retrieval characteristics, for example, recall latency, fluency, ease of access, or memory vividness into account (Koriat and Ackerman 2010; Koriat et al. 2009; Robinson et al. 2000). However, this effect was limited to monitoring correct recall. That is, children gave lower CJ for difficult compared to easy items, but only when their recall had been correct. When recall was incorrect, children were not able to take item difficulty further into account for their monitoring. Possibly, recognition of difficult items is more demanding than recognition of easy items, and thus cognitive resources for accurate monitoring difficult items may have been limited.

Children's fixation time allocation during recognition mirrored retrieval processes in that differences in fixation times were found for the different classes of items. During recognition, age-dependent patterns of fixation time allocation were found for easy versus difficult items. Older children spent more fixation time on the target Kanji when the item was difficult, possibly while engaging in more effortful retrieval attempts, compared to easy items. This finding could be interpreted as showing that older children relied more on their own memories and retrieval efforts than on the offered alternatives. Insofar, 9-year-olds' fixation time allocation correspond nicely to their advantage in uncertainty monitoring when the openly reported CJ are considered. Younger children, in contrast, looked longer both at correct and incorrect alternatives for the difficult items compared to the older children. Their approach may be interpreted as a less-than-optimal retrieval strategy either because of an untimely termination of memory search or because of an increased risk for interference. In fact, younger children's difficulties with monitoring recall of difficult items as discussed in the previous paragraph may be rooted in their less favourable allocation of fixation time during recognition. Support for this interpretation can be also drawn from fixation times measured during monitoring: While giving explicit CJ, older but not younger children looked longer at difficult target Kanjis compared to easy Kanjis. Therefore, age differences in fixation time allocation uncovered through eye-tracking may indicate developmental progression in the ability to perceive the need for and to make an additional effort when retrieval is difficult. These increasing abilities may influence simultaneous and subsequent monitoring.

As to implicit monitoring judgments, we found the same age-dependent overall decrease in confidence from easy to difficult items as is typically found in explicit judgments (Roebbers 2002). Interestingly, 9-yr-olds showed the expected pattern with lower confidence for incorrect than for correct answers even before explicitly judging the correctness of answer, regardless of item difficulty. Seven-year-olds, in contrast, showed the expected pattern in the implicit judgments only when considering the easy items. For difficult items, however, they gave even higher implicit CJ for incorrect than for correct answers. These very positive early implicit judgments of the younger participants when monitoring difficult items were unexpected and may reflect children's overoptimistic view of their memory

skills as well as their wishful thinking, at least to some extent (Schneider 1998). Older children's more accurate early judgments may be based on assessments of memory strength. Again and similar to the interpretation offered concerning fixation time allocation during retrieval and explicit monitoring, it appears from the implicit judgments that younger children have more difficulties in using retrieval characteristics for monitoring than older children.

When implicit and explicit judgments were related to each other, substantial associations were found. The substantial link between implicit and explicit CJ suggests that explicit judgments are built on implicit judgments. At the same time, associations between implicit and explicit CJ tended to be closer in younger compared to older children, possibly indicating a stronger reliance on implicit monitoring processes when reporting confidence. For the older children, conversely, more conscious consideration of social desirability and/or theory-based heuristics may play a role (Koriat 1997, 2007). Both, increasing use of more conscious information as well as more accurate implicit monitoring processes are likely to contribute to better explicit monitoring in the 9-year-old compared to the 7-year-old children.

The fact that the correspondence of implicit and explicit judgments was largest for unanswerable items may be interpreted as showing that successful (in the case of easy and difficult items) and unsuccessful (in the case of unanswerable items) retrieval efforts have differentially influenced these judgments.


The question of whether eye-tracking approaches can provide new and interesting insights for the study of metacognition can clearly be answered positively based on the present study. On the one side, a current and important discussion among researchers in the field is related to the question whether monitoring is at all separable from long-term memory and its retrieval processes. In this respect, our eye-tracking data reveals that monitoring and retrieval are very closely related if not intertwined. When children were asked to give confidence judgments, they allocated more fixation time on the retrieval cues, that is, on the target Kanjis than on the CJ scale itself. Possibly, by looking at the retrieval cues participants tried to gather information useful for metacognitive monitoring. This pattern of fixation time allocation might therefore indicate efforts to estimate familiarity with the Kanji, correctness of the association, or an attempt to recall earlier feelings of correctness during the recognition task, especially in the older children. In the case of difficult items, investing more time for using heuristics may result in more accurate monitoring of recall, as was found for the 9-, but not for the 7-year-olds. Eye-tracking data may also enhance our understanding, as to why, even with relatively good uncertainty monitoring, CJ are still far higher than the lowest possible value. As can be seen from Fig. 3, the lowest CJ are hardly ever looked at, independent of item difficulty and age. Possibly, this finding indicates a reluctance to seriously consider uncertainty, or at least, pronounced degrees of uncertainty. The reasons for this general bias towards higher certainty even in the implicit CJ are unclear but should be targeted in future research. Thus, eye-tracking appears to be a promising research tool for uncovering implicit and retrieval-bound monitoring processes as well as age differences in these cognitive activities.


Taken together, the present study yielded evidence for improvements in monitoring during the primary school years in the domain of uncertainty monitoring and was successful in showing that even in children, retrieval processes are closely connected with and have a significant impact on monitoring. Furthermore, improvements in uncertainty monitoring seem due to more sophisticated skills in older compared to younger children when information from retrieval processes are to be taken into account for accurate monitoring.


Appendix

Easy items – learning trial

水 
Wasser

柵 
Zaun


心 
Herz

口 
Mund

兔 
Hase

山 
Berg

火 
Feuer

木 
Baum

Difficult items – learning trial

草



Gras

櫃



Truhe

猫



Katze

道



Strasse

腕



Arm

家



Haus

月



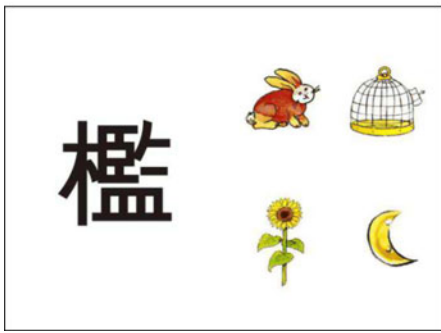
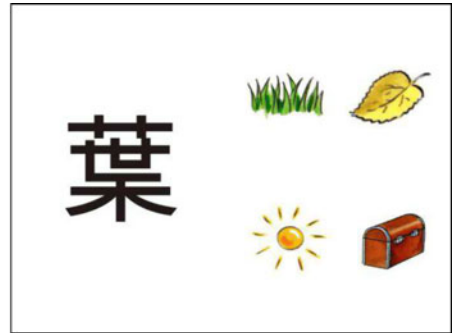
Mond

足



Fuss

Unanswerable items – recognition trial



References

- Arbuthnott, K. D., Kealy, K. L. K., & Ylloja, S. (2008). Judgement of confidence in childhood memories. *Applied Cognitive Psychology, 22*, 953–978.
- Aslin, R. N., & McMurray, B. (2004). Automated corneal-reflection eye-tracking in infancy: methodological developments and applications to cognition. *Infancy, 6*, 155–163.
- Baillargeon, R. (1987). Object permanence in 3 1/2- and 4 1/2-month-old infants. *Developmental Psychology, 23*, 655–664.
- Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377–395.
- Corbetta, M. (1998). Frontoparietal cortical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences, USA, 95*, 831–838.
- Dunlosky, J., & Bjork, R. A. (2008). The integrated nature of metamemory and memory. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 11–28). New York: Psychology.
- Dunlosky, J., Hertzog, C., Kennedy, M. R. F., & Thiede, K. W. (2005). The self-monitoring approach for effective learning. *International Journal of Cognitive Technology, 10*, 4–11.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics, 16*, 143–149.
- Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 3–34). Hillsdale: Lawrence Erlbaum Associates.
- 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. (2004). Advances in relating eye movements and cognition. *Infancy, 6*, 267–274.
- Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research, 38*, 125–137.
- 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.
- Karatekin, C. (2004). Development of attentional allocation in the dual task paradigm. *International Journal of Psychophysiology, 52*, 7–21.
- Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Developmental Review, 27*, 283–348.
- Kelley, C. M., & Jacoby, L. L. (1996). Memory attributions: Remembering, knowing, and feeling of knowing. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 287–308). Hillsdale: Erlbaum.
- 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.
- Kinnunen, R., Vauras, M., & Niemi, P. (1998). Comprehension monitoring in beginning readers. *Scientific Studies of Reading, 2*, 353–374.
- 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. (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. (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. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *Cambridge handbook of consciousness* (pp. 289–325). New York: Cambridge University Press.
- Koriat, A. (2008). Subjective confidence in one's answers: the consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 945–959.
- Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science, 13*, 441–453.
- Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy: Insights from the processes underlying judgments of learning in children. In M. Izaute, P. Chambres, & P.-J. Marescaux (Eds.), *Metacognition: Process, function, and use* (pp. 1–17). New York: Kluwer.

- Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009). The memorizing-effort heuristic in judgments of learning: a developmental perspective. *Journal of Experimental Child Psychology, 102*, 265–279.
- Körkel, J. (1987). *Die Entwicklung von Gedächtnis- und Metagedächtnisleistungen in Abhängigkeit von bereichsspezifischen Vorkenntnissen*. Frankfurt: Lang.
- Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25*, 265–288.
- Lockl, K., & Schneider, W. (2003). Metakognitive Überwachungs- und Selbstkontrollprozesse bei der Lernzeiteinteilung von Kindern. *Zeitschrift für Pädagogische Psychologie, 17*, 173–183.
- Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Mahwah: Lawrence Erlbaum Associates.
- 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.
- Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (pp. Vol. III, pp. 269–284). Mahwah: Lawrence Erlbaum Associates.
- 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*, 125–173. New York: Academic Press.
- Pauen, S. (2002). Evidence for knowledge-based category discrimination in infancy. *Child Development, 73*, 1016–1033.
- Pickering, S., & Gathercole, S. (2001). *The working memory test battery for children*. London: The Psychological Corporation.
- Pillow, B. H. (2002). Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. *Child Development, 73*, 779–792.
- Pillow, B. H., & Anderson, K. L. (2006). Children's awareness of their own certainty and understanding of deduction and guessing. *British Journal of Developmental Psychology, 24*, 823–849.
- Pillow, B. H., Hill, V., Boyce, A., & Stein, C. (2000). Understanding inference as a source of knowledge: children's ability to evaluate the certainty of deduction, perception, and guessing. *Developmental Psychology, 36*, 169–179.
- 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.
- Rhodes, M. G., & Kelley, C. M. (2005). Executive processes, memory accuracy, and memory monitoring: an aging and individual difference analysis. *Journal of Memory and Language, 52*, 578–594.
- Robinson, M. D., Johnson, J. T., & Robertson, D. A. (2000). Process versus content in eyewitness metamemory monitoring. *Journal of Experimental Psychology: Applied, 6*, 207–221.
- Roebers, C. M. (2002). Confidence judgments in children's and adult's event recall and suggestibility. *Developmental Psychology, 38*, 1052–1067.
- Roebers, C. M., Howie, P., & Beuscher, E. (2007). Can private reports enhance children's event recall, lower their suggestibility and foster their metacognitive monitoring compared to face-to-face interviews? *Computers in Human Behavior, 23*, 749–769.
- 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.
- Schacter, D. L. (1987). Implicit memory: history and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501–518.
- Scheiter, K., & Van Gog, T. (2009). Using eye tracking in applied research to study and stimulate the processing of information from multi-representational sources. *Applied Cognitive Psychology, 23*, 1209–1214.
- Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science, 1*, 291–297.
- 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.
- Spelke, E. S., & Kestenbaum, R. (1986). The origins of the concept of objects. *Psychologie Française, 31*, 67–72.
- Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin and Review, 6*, 662–667.
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66–73.
- Van der Stel, M., & Veenman, M. V. J. (2008). Relation between intellectual ability and metacognitive skillfulness as predictors of learning performance of young students performing tasks in different domains. *Learning and Individual Differences, 18*, 128–134.

- 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.
- 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.
- Wang, J., Spencer, K., & Xing, M. (2009). Metacognitive beliefs and strategies in learning Chinese as a foreign language. *System, 37*, 46–56.
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*, 749–750.
- Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: shifting from process to outcome self-regulatory goals. *Journal of Educational Psychology, 91*, 241–250.