

What are the metacognitive costs of young children's overconfidence?

Nesrin Destan · Claudia M. Roebers

Received: 15 May 2014 / Accepted: 28 December 2014 / Published online: 3 February 2015
© Springer Science+Business Media New York 2015

Abstract Children typically hold very optimistic views of their own skills but so far, only a few studies have investigated possible correlates of the ability to predict performance accurately. Therefore, this study examined the role of individual differences in performance estimation accuracy as a global metacognitive index for different monitoring and control skills (item-level judgments of learning [JOLs] and confidence judgments [CJs]), metacognitive control processes (allocation of study time and control of answers), and executive functions (cognitive flexibility, inhibition, working memory) in 6-year-olds ($N=93$). The three groups of under estimators, realists and over estimators differed significantly in their monitoring and control abilities: the under estimators outperformed the over estimators by showing a higher discrimination in CJs between correct and incorrect recognition. Also, the under estimators scored higher on the adequate control of incorrectly recognized items. Regarding the interplay of monitoring and control processes, under estimators spent more time studying items with low JOLs, and relied more systematically on their monitoring when controlling their recognition compared to over estimators. At the same time, the three groups did not differ significantly from each other in their executive functions. Overall, results indicate that differences in performance estimation accuracy are systematically related to other global and item-level metacognitive monitoring and control abilities in children as young as six years of age, while no meaningful association between performance estimation accuracy and executive functions was found.

Keywords Performance estimation accuracy · Metacognitive abilities · Executive functions · Children · Individual differences

Have you ever asked a 6-year-old how good he or she is at throwing a ball, counting or playing memory relatively to his or her peers? Most likely, they will tell you to outperform their friends by far. Indeed, when asked to evaluate their performance, several studies show that young children are overconfident about their own abilities in physical (e.g., Plumert and Schwebel 1997; Schneider 1998), social tasks like the imitation of a model (e.g., Bjorklund et al. 1992), but also cognitive tasks (e.g., Finn and Metcalfe 2014; Lipko et al. 2009; Shin et al. 2007). To exemplify, overconfidence persisted even when preschoolers repeatedly experienced the recall

N. Destan (✉) · C. M. Roebers
Department of Developmental Psychology and Center for Cognition, Learning, and Memory, University of
Bern, Hochschulzentrum vonRoll, Fabrikstrasse 8, CH-3012 Bern, Switzerland
e-mail: nesrin.destan@psy.unibe.ch

of less pictures than they predicted and kindergarteners remained overconfident in their performance predictions even after receiving feedback (Lipko et al. 2012; Shin et al. 2007). Although overconfidence seems to decrease gradually as children get older (e.g., Schneider et al. 2000), even adults are shown to hold overly optimistic views of their abilities (Clayson 2005; Kruger and Dunning 1999; Szpunar et al. 2014). For instance, college students were asked to estimate their exam score in relation to their peers. Based on their actual test score, they were then separated into four groups. Interestingly, the 25 % performing poorest were the least accurate in judging their competence, overestimating their own performance by approximately 30 % (Dunning et al. 2003).

The current paper addressed young children's over- and under confidence in performance in terms of potential costs for metacognitive monitoring and control skills. We aimed to investigate whether inaccurate performance predictions as a global metacognitive index are associated with costs in other global and item-level metacognitive indices as well as executive functions.

Metacognitive monitoring and control

Theoretically, the ability to estimate one's performance by making a global judgment pertains to the metacognitive process of monitoring ongoing learning and retrieval (e.g., Dunlosky and Metcalfe 2009). As such, metacognitive monitoring and control processes allow us to observe, assess and reflect on current mental or epistemic states (Flavell 1979). While monitoring processes inform the individual about an ongoing activity (i.e., how well have I learned a list of words?) in relation to a goal (i.e., obtain a good grade in an upcoming test), control behaviors are activated in response to our monitoring (i.e., spend more time learning a long compared to a short word list) (Nelson et al. 1994; Nelson and Narens 1990).

To investigate metacognitive monitoring processes, participants can be asked to estimate their performance on a global level by either anticipating how many items they will remember (performance predictions) or by retrospectively evaluating their potential test score (performance postdictions; e.g., Lipko et al. 2012; Roderer and Roebers 2013; Shin et al. 2007). Besides such global judgments, monitoring processes are assessed by evaluating the learning of individual items. Judgments of learning (JOLs; e.g., Metcalfe and Finn 2013; Schneider et al. 2000) and confidence judgments (CJs; e.g., Lyons and Ghetti 2011; Roebers et al. 2007) are two classical types of item-by-item metacognitive judgments frequently applied to study monitoring processes. For JOLs, being prospective in nature, participants are asked to judge how well they have learned a certain kind of information (i.e., word-pairs) by making judgments on the item-level, usually after a study period and before the actual test. CJs, on the other hand, are retrospective judgments for which individuals are asked to judge the certainty that their response (i.e., the answer on a recognition test) was correct for every single item.

To examine the interplay of metacognitive monitoring and control processes, JOLs and CJs can be paired with corresponding control behavior. Thus, JOLs are frequently studied in relation to subsequent study time (Dunlosky and Connor 1997; Lockl and Schneider 2003; Son 2005; Son and Metcalfe 2000), and adults are typically found to use metacognitive judgments to control their study behavior by allocating more study time to items with lower JOLs (e.g., Nelson et al. 1994). Lockl and Schneider (2003) showed with a sample of first and third grade children that both age groups differentiated between easy- and hard-to-learn picture pairs in their JOLs, indicating an ability to monitor their learning. Furthermore, the third graders but not the first graders were able to base their subsequent control behavior on their monitoring judgments, spending more time learning the item pairs judged as difficult and not well learned. For CJs and corresponding control behavior, individuals might be asked to report, withhold or correct answers in accordance with

their feeling of certainty. An adequate strategy would be to report answers with high confidence and withhold or correct errors of answers with low confidence (e.g., Koriat and Goldsmith 1996; Krebs and Roebers 2010; Roderer and Roebers 2009).

Overconfidence in children: costs or benefits?

According to Kruger and Dunning (1999), individuals who overestimate their performance might lack necessary metacognitive skills and therefore have difficulties in adjusting their learning behavior or making correct decisions, for example, whether to provide or withhold information in a test situation. For instance, overestimating one's performance (i.e., judging inaccurate answers as correct in the study phase) might lead to a premature termination of study time or to generally fewer items being chosen for self-paced study (Dunlosky and Rawson 2012; Metcalfe and Finn 2008; van Loon et al. 2013).

One popular explanation for the pronounced overestimation in children had been the assumption of a global metacognitive deficit (Flavell et al. 1970) such that children are generally not able to monitor their performance accurately and therefore cannot integrate information when making predictions, leading to overestimations. But, as shown in studies with kindergarteners and third graders, overestimation resulted due to wishful thinking rather than poor monitoring skills. In fact, children had difficulties differentiating between their wishes and expectations and overestimated the power of effort, speaking at least for a partial influence of motivational factors when making performance predictions (Schneider 1998; Visé and Schneider 2000).

Nevertheless, individuals with accurate predictions of their skills and performance should outperform their overconfident peers. Consistent with this assumption, student participants who were good at differentiating between what they know and don't know were found to score higher in exams than those who were not as precise when making performance predictions (e.g., Hacker et al. 2000; Sinkavich 1995). Regarding young children, a study with 3- to 4-year-olds demonstrated that the participants with the highest sensitivity to their lack of knowledge in a lexical discrimination task outperformed those who were unaware of their ignorance (Marazita and Merriman 2004). Similarly, Roderer and Roebers (2013) categorized fifth graders into three groups according to their test performance in mathematics and science (low, average and high achievers). The highest achieving children provided the most accurate performance predictions in comparison to average and low achievers. Together these results suggest that from early on, more accurate metacognitive skills are positively and substantially associated with performance.

On the other hand, the possibility of an adaptive function of overconfidence in children is discussed in the literature, suggesting that unrealistic optimism helps to approach novel and difficult tasks, which children wouldn't pursue if they had highly accurate metacognitive skills (Bjorklund et al. 1992; Shin et al. 2007). Shin and colleagues (2007) asked kindergarteners, first and third graders to learn different pictures across multiple trials and to predict how many items they would recall. Children were then separated into two groups based on their overconfidence (low and high overconfidence), followed by a performance test. Kindergarteners and first graders demonstrated a loss in recall performance in subsequent trials, compared to the first trial. Interestingly, the children with pronounced overconfidence had a lower loss in recall performance. Moreover, in the group of high overconfident third graders, recall performance actually increased across trials. Taken together, these findings speak for beneficial effects of pronounced overestimation. In view of that, Bjorklund and Bering (2002) argued that children's overconfidence helps them to persist in a task and keep their motivation high even when making errors and then, through their engagement, in fact improve performance.

Taken together, children as well as adults are overly optimistic when judging their own performance but at the same time, there are still evident individual differences in the accuracy

of these self-evaluations (i.e., over- and underestimation). Still, so far no consensus has been reached whether inaccurate performance predictions have mostly costs or benefits on children's performance. Nevertheless, one could imagine that accurate self-assessments may be especially critical for success in the context of learning and test taking. In such contexts, and on the level of single items, the ability to recognize what one knows and what one doesn't know and to be able to control how, when and what to study is very likely to be beneficial. In this regard, overestimating one's performance would be expected to be associated with costs rather than benefits. Unfortunately though, in most of the previous studies with either children or adults, only one overall measure of performance estimation was included, leaving the question open whether a more generalized metacognitive inability serves as underlying factor for an individual's over- or underestimation. Therefore, in the present study overall performance predictions will be contrasted with classical item-by-item measures of metacognitive monitoring and control. Besides, we will also compare overall predictions with overall postdictions to explore whether overestimation in predicting performance before the memory test is related to performance estimations after the test. Shedding light on possible individual differences in global as well as item-level metacognitive indices will be informative concerning the emergence of these skills.

Metacognition and executive functions

Individual differences in the ability to monitor and control one's thoughts and actions in learning and test situations has not only been investigated in the field of metacognition but also in the broader context of self-regulation, with the construct of executive functions playing a central role (Best and Miller 2010; Lyons and Zelazo 2011; Roebers et al. 2012). EF can be described as the ability to consciously control thoughts and actions (Lyons and Zelazo 2011) by storing and manipulating task related information, inhibiting inappropriate or automatized responses and by flexibly shifting between task rules or mental sets (Best and Miller 2010; Miyake and Friedman 2012; Miyake et al. 2000). Still, there is ongoing discussion on the underlying structure of executive functions (one factor vs. multiple factors). Regarding the kindergarten period, there is data from cross-sectional studies suggesting an undifferentiated, one-factor structure of EF (e.g., Hughes et al. 2010; Wiebe et al. 2008), with recent studies additionally proposing a rather slow differentiation in this structure across childhood and adolescence (e.g., Lee et al. 2013; Tucker-Drob 2009). For example, a two-factor structure of EF was confirmed in 5- to 13-year-olds, while only 15-year-olds were shown to have a well-separated three-factor structure (Lee et al. 2013). Although the issue considering developmental changes of EF and the underlying structure is not entirely clarified, most researchers agree that EF entail higher-order cognitive processes involving the dimensions of working memory (describing the ability to store and manipulate information for a short period of time), inhibition (entailing the ability to suppress dominant or impulsive responses), and cognitive flexibility (involving the flexible shift of mental sets or tasks rules to adapt to changes in the environment) (Best et al. 2009; Diamond 2013; Lyons and Zelazo 2011).

Even though metacognitive abilities and executive functions have rarely been investigated in relation to each other, there are results speaking for an interrelation between the two constructs (e.g., DeMarie et al. 2004; Fernandez-Duque et al. 2000; Roebers et al. 2012). Conceptually, both constructs involve self-reflection as part of self-regulatory processes, necessary to store and manipulate memory content, flexibly switch between stored knowledge and to be learned material, and to choose adequate strategies to regulate behavior (e.g., Lyons and Zelazo 2011). On a behavioral level, for example, the association between executive functions and metacognition was investigated in a longitudinal study with first graders, showing that executive functions were significantly related to metacognitive control, cross-

sectionally as well as longitudinally, with individual differences in both constructs being significantly related to academic performance in children's subsequent second grade (Roebbers et al. 2012). Similarly, Best et al. (2011) argued that metacognitive mechanisms of monitoring and control are related to the development of accuracy and reaction time in tasks measuring executive functions during the school years. This seems to be particularly the case because both constructs call for goal-oriented processing of information and for continuous adaptation of strategies used to master a given task. To exemplify, an age-related increase in accuracy and speed from middle childhood to young adulthood was observed: the awareness of failures in inhibiting an inappropriate response increased, followed by a slowing down in speed to improve accuracy in a switch-task (e.g., Best et al. 2011; Davidson et al. 2006).

The present study

The primary goal of this study was to advance our understanding of individual differences (i.e., over- and underestimation) in the accuracy of estimating one's overall performance in children. We aimed to explore whether differences in performance estimation accuracy are associated with costs in metacognitive skills and whether these differences are also related to executive functions. For analyses, children were separated into three groups called "under estimators", "realists" and "over estimators", based on their performance estimation accuracy. Additionally, information on sex, age, non-verbal IQ, vocabulary skills and SES were included.

In summary, the purpose of the current investigation was to examine the correspondence between metacognitive monitoring (JOLs and CJs) and control processes (study time allocation and control of answers) as a function of participants' performance estimation accuracy. Moreover, we aimed to explore whether performance estimation accuracy is systematically related to executive functions (cognitive flexibility, inhibition and working memory). We expected that (a) children with more accurate performance predictions would show better metacognitive skills on the global as well as item-level, allowing children to link more realistic monitoring of their performance with adequate control strategies. Accordingly, overestimating one's own performance should be associated with costs in metacognitive skills. We further hypothesized that (b) metacognitive monitoring and control skills would be substantially linked to executive functions, as the interplay of monitoring one's task performance and goal-directed behavior appear to be processes shared by both, metacognition and executive functions (Lyons and Zelazo 2011; Roebbers et al. 2012).

Method

Participants

The sample consisted of a total of 93 6-year-old children (47 girls and 46 boys) within an age range of 12 months (min. = 71 months, max. = 83 months) and a mean age of 75.9 months ($SD=2.9$ months). The initial pool included 100 participants. Data of seven participants was omitted due to missing data from one of the sessions (e.g., due to illness), comprehension difficulties of task instructions or because the set age criterion was not met.

Children were recruited in the German speaking part of Switzerland with participants being predominately of Caucasian origin and from middle-class families. Data collection started at the beginning of the kindergarten year and took place from August until December. All participants attended kindergarten. Of the total sample of $N=93$, 73 % of the children were native German speakers, while 27 % of the children were non-native German speaking

immigrants from different European countries. Children were not eligible for participation if they could read Asian ideograms to prevent that prior knowledge of the stimuli influenced performance in the paired associate learning task (see below). Besides children's nationality, information regarding the spoken languages in their homes was available. Children were only included in the study if they were sufficiently fluent in German due to the verbal instructions in the tasks used. Written informed parental consent was obtained prior to the study and children were asked to give their verbal assent.

Procedure and materials

Three trained experimenters tested children individually in a quiet room during two visits (on average 14 days apart), lasting approximately 40 minutes each. The assessment included tasks on procedural metacognition (monitoring and control), executive functions (cognitive flexibility, inhibition and working memory), non-verbal IQ and vocabulary skills (expressive and receptive vocabulary). Moreover, information regarding the SES (mother's education, father's education and the family's monthly income) was obtained from the participant's parents. The order of the various tasks quantifying metacognitive monitoring and control, executive functions, non-verbal IQ and vocabulary skills was counterbalanced. Task orders differed such that children would start their first session either with the three tasks assessing executive functions or with the task quantifying metacognitive monitoring and control. In addition, the pairing of the EF tasks and the metacognition task with the non-verbal IQ measure and the vocabulary tests varied across the two sessions, resulting in four different task orders. Children were randomly assigned to one of these task orders.

Metacognitive monitoring and control task

Twenty Japanese characters called Kanji served as stimuli in the paired associate learning task, consisting of several phases involving metacognitive monitoring and control processes. Japanese characters have already been used as learning material in previous studies with 5- to 9-year-old children (e.g., Roderer and Roebers 2010, 2014). The paired associate learning task consisted of two sessions (A and B) with ten unique Kanji each, varying in difficulty and presented in randomized order. The order of the two item sets was counterbalanced across participants. The details of the experimental procedure are illustrated in Fig. 1.

Metacognitive judgments scale JOLs and CJs were provided using an external response pad (RB-730, Cedrus Cooperation), on which children could press one of five buttons, each representing one point of the Likert scale. The scale ranged from very unsure to very sure, with each point being depicted by a different smiley face (Allwood 2010; Krebs and Roebers 2010; Roebers et al. 2009). A pictorial image of the scale was provided on the response pad and on the laptop screen to ensure the participants' understanding of which button to press for the five scale points.

At the beginning, all children were introduced to the scale by receiving an explanation of the five scale points when asked to provide JOLs and CJs (e.g., the very sure button for JOLs: "*If you are really sure that you'll be able to remember what that sign means later on, and you think you will get it right, you should press this button*" [Experimenter points to the according smiley button]). To test the participants understanding of the scale, they were then asked to point to the requested scale points/buttons on the response pad (e.g., for JOLs: "*which button do you press when you're really sure you'll remember what the Kanji means if I ask you later on today?*"). The explanation of the scale points was repeated for CJs with an adaptation in wording (e.g., the very sure button for CJs: "*If you are really sure that you chose the right picture, and you think*

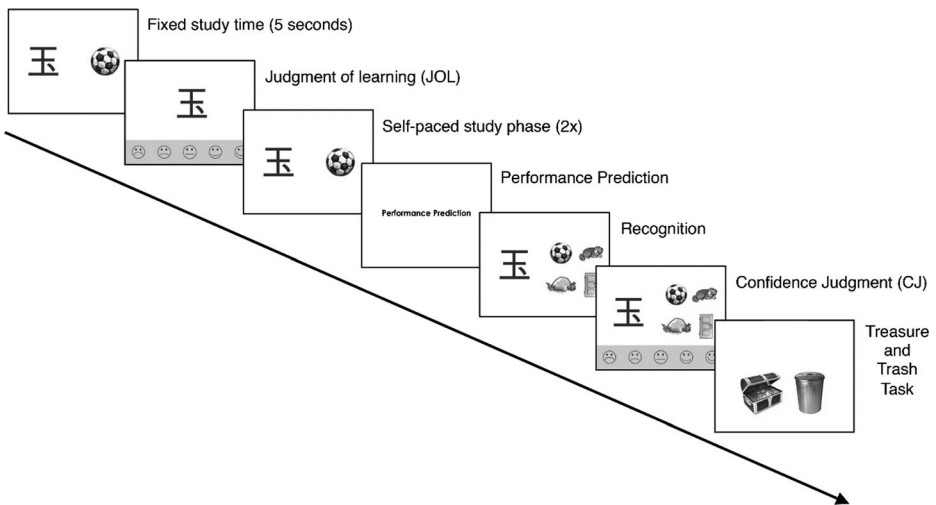


Fig. 1 Schematic of the paired associate learning task procedure

you got it right, you should press this button” [Experimenter points to the according smiley button]). Explanations were repeated if children pointed to the wrong smiley face.

Treasure chest and trash can A cardboard “treasure chest” and “trash can” were used as a task to quantify metacognitive control, in which children were told to win points for the accurate control of answers, i.e., correct items placed in the treasure chest and incorrect items placed in the trash can, which would allow them to choose a toy at the end. Similar control tasks involving choice embedded in a betting game have been previously used successfully to study metacognitive skills in young children (e.g., Destan et al. 2014; Hembacher and Ghetti 2013; Sussan and Son 2007). After task administration, all children received a toy of their choice, regardless of actual performance.

Task introduction Children were introduced to the paired associate learning task by being told a short story about how Japanese children learn to read and write. In this introductory part, children were familiarized with the various assignments (e.g., fixed and self-paced study phases, JOLs, recognition and CJs) and had the opportunity to practice the procedure with three different Kanji. After practicing, the actual test started with two identical sessions but two different sets of items (10 unique Kanji each), separated by a short break. In the following, we will describe the separate phases involved in each session.

Encoding phase The first step was a fixed-time study phase in which each of the 10 Kanji within the session was shown for 5 s together with a pictorial image representing the meaning of the Kanji. No verbal input was provided.

Judgments of learning (JOLs) Immediately after the fixed-study phase, children were shown one Kanji at a time and asked to provide a JOL for every Kanji they had learned by using the external response pad with the smiley scale. Specifically, they responded to the question “*how sure are you that you’ll remember what the Kanji means if I ask you later on?*” by pressing one of the five buttons. JOLs were scored from 0 (very

unsure) to 4 (very sure). JOLs “low” (response 0 and 1) and JOLs “high” (response 3 and 4) were used for analyses.

Allocation of study time Next, participants had the opportunity to re-learn the meaning of the 10 Kanji in two study phases within one session, allocating study time at their own pace. The child could press an arrow button provided on the external response pad next to the smiley scale to signal having finished studying one Kanji and to move on to the next item. It was not possible to go back to the previous Kanji once the arrow button was pressed. There was no upper limit for the self-paced study time and thus participants could spend as much time as needed on every Kanji. The second self-paced study phase followed immediately after the child finished studying the 10 Kanji for the first time. The Kanji were presented randomly in both self-paced study phases. Allocated study time as a function of study phase and averaged across session A and B was used for analyses.

Performance prediction An overall performance prediction followed the two self-paced study phases. Children were shown ten plain cardboard coins, with every coin symbolizing one of the 10 Kanji learned in session A and B, respectively. The child was asked to put as many coins in a cardboard box, as he or she thinks to have learned successfully in the session and would be able to remember by responding to the request: “*I want you to think about how many Kanji you think you still remember later on, and put that many coins in this box*”. The total number of Kanji estimated across both sessions (total of 20) was used for analyses.

Recognition test The performance prediction was followed by a recognition test in which children saw each of the 10 Kanji they had learned within one session presented on a screen with four different pictures. All pictures depicted the meanings of Kanji that had appeared in the session and were thus familiar to the child. One of the pictures was the correct meaning for the Kanji presented, while the remaining three were the correct meanings for different Kanji. The child was instructed to point to the picture that was the correct meaning for the Kanji presented. As the two sessions A and B did not differ significantly from each other in overall recognition accuracy, $t(92)=1.13$, $p=.26$, $d=.12$, the total number of correctly recognized Kanji across both sessions (20 Kanji) was used for analyses.

Confidence judgments (CJs) Immediately after the recognition test was completed, children were shown one Kanji at a time and asked to provide a CJ for the response they had given using the smiley scale. They were asked, “*how sure are you that you've just named the correct picture for the Kanji?*” and again responded by pressing one of the buttons on the response pad, in the same way that they had provided JOLs. CJs were scored from 0 (very unsure) to 4 (very sure). Mean CJs for correctly and incorrectly recognized Kanji, as well as the difference score in percentage (ranging from 0 to 100 %) between confidence judgments and corresponding recognition performance served as dependent measures for analyses.

Control of answers After children had provided CJs for every Kanji within one session, a final metacognitive control task was administered: participants were asked to control their answers into two different containers; either a treasure chest or a trash can, depending on whether or not children believed they had correctly identified the meaning of that Kanji during the recognition test. Children were told that only accurate recognitions should be placed in the treasure chest and inaccurate recognitions should be placed in the trash can to win points. Thus, children were required to exert control over their performance by placing their answers into the appropriate container. The total of accurate control decisions (i.e., correctly recognized and

placed in the treasure chest and incorrectly recognized and placed in the trash can) across both sessions (total of 20 Kanji) served as the dependent measures.

Executive functions tasks

Executive functions (EF) are measured using three different tasks quantifying cognitive flexibility, inhibition and working memory processes.

Cognitive flexibility Cognitive flexibility was measured using a computerized task in which children are asked to feed two categories of fish (either single-colored or multicolored), that are presented simultaneously on the left and the right side of the screen (Roebbers et al. 2012; Roebbers and Kauer 2009). The goal of the task is to feed the two categories of fish consecutively. The position of the fish is randomized (left or right) and therefore the child has to decide whose turn it is to be fed. The fish can be fed by pressing an external response button corresponding either to the left or the right side of the screen. There are two phases consisting of a total of 46 trials with inter-stimuli intervals varying from 300 to 700 ms. The overall accuracy (“the proportion of fishes correctly fed”) served as the dependent measure.

Inhibition The task used to assess inhibition is an adapted version of the commonly used stroop task by Archibald and Kerns (1999). On four different pages containing 25 stimuli each, children are asked to name the colors of fruits and vegetables as fast as possible. For black and white or wrong colors, children are asked to name the original color of the fruits and vegetables. The dependent measure of interference control (s), mirroring an individual’s ability to inhibit dominant or prepotent answers, is calculated using the formula from Archibald and Kerns (1999), with lower values indicating better inhibition/interference control. To simplify the comparison between the three measures of executive functions and the readability of the results in the current study, the values were converted, with now higher values indicating better inhibition/interference control.

Working memory Working memory was measured using a backward color recall task (Roebbers and Kauer 2009; Roethlisberger et al. 2010), a modified version of the backward digit span task from the WMTB-C (Pickering and Gathercole 2001). The goal of the task is to recall a sequence of colored discs in reversed order (one-syllable color labels in German), presented on a laptop. Every colored stimulus is presented for 1 s. Testing started with three trials of a two-stimulus sequence length. After completing at least two of the three trials correctly, the sequence length was increased by one item. The sum of correctly recalled sequences served as the dependent measure.

A principal component analysis (PCA) was conducted on the three EF tasks using varimax rotation. The Kaiser-Meyer Olkin measure confirmed the sampling adequacy, $KMO=.62$, and Bartlett’s Test of sphericity, $\chi^2(88)=17.63$, $p<.01$, indicated that the correlations between the three variables were large enough to conduct a PCA. One component with an Eigenvalue over Kaiser’s criterion of 1 resulted, explaining a total of 51.7 % of the variance. The resulting one-component structure of EF is in line with other cross-sectional findings, suggesting EF to be an unitary and more domain-general construct in children before school age (e.g., Hughes et al. 2010; Wiebe et al. 2008). Thus, for the current analyses a composite score of EF was built by adding up the z-scores of the three task measures.

Control measures

Non-verbal IQ To assess nonverbal intelligence, the TONI-III was administered (Brown et al. 1997). This test involves a set of 45 items with abstract contents and progressive difficulty. Each of the 45 items has six response options of which one is the correct answer. The dependent measure used for the current analyses was the raw score, consisting of the total number of correctly solved items.

Vocabulary skills Vocabulary skills were tested using the two subtests expressive and receptive vocabulary from the HAWIVA-III (Ricken et al. 2007), the German adaptation of the WPPSI-III (Wechsler 2002). The expressive vocabulary test consists of 26 items of which the child is asked to provide the correct label. To test receptive vocabulary, the participant is asked to point to the requested item (e.g., “show me the piano!”). Based on the significant correlation ($r_s = .21, p < .05$) between expressive and receptive vocabulary in our sample, the z-standardized total number of correctly solved items (raw score) from both tests is summed up and used as a “vocabulary skills” composite score for analyses.

SES To assess SES, both parents were asked to provide information about their level of education and the monthly income of the family. To indicate their educational level, parents could choose from four possible answers ranging from having no graduation to having a university degree. Similarly, four options were provided for the monthly income, with the numbers being based on representative incomes of Swiss families in 2012. For 75 children, information on SES was available. The SES composite score used for analyses consisted of the added z-scores of the three variables mother’s and father’s educational level and the family’s monthly income, which were all significantly correlated ($r_s \geq .30, p_s \leq .01$).

Results

The results are split into two sections: in the first section, multiple regression analyses for metacognitive monitoring and control skills as well as executive functions in relation to performance estimation accuracy are presented. In a second part, children are split into three groups (under estimators, realists, over estimators) based on their performance estimation accuracy, and then compared with each other in relation to their metacognitive skills and executive functions. The descriptive statistics of the dependent variables are presented in Table 1.

Multiple regression analyses

In a first step, multiple regression analyses were used to test if performance estimation accuracy in the current sample of 6-year-olds significantly predicted participants’ metacognitive monitoring and control skills as well as executive functions. Moreover, the control measures of age, sex, non-verbal IQ and vocabulary skills were included as predictors in the analyses.

Global metacognitive monitoring index

Calibration Performance estimation accuracy serves as a global monitoring index and is based on a judgment made prior to the actual recognition test. To see whether children’s performance prediction accuracy is related to their ability to monitor retrieval on a global level, i.e., to

Table 1 Descriptive statistics of the dependent variables

Measures (unit)	<i>M</i>	<i>SD</i>	Range
Metacognition			
Global-level monitoring: postdiction	26.85	22.28	-21.25–75.00
Item-level monitoring: relative accuracy (G correlation)	0.26	0.55	-1.00–1.00
Item-level monitoring: discrimination in CJ between correct and incorrect recognition (mean CJ)	0.26	0.61	-0.95–2.00
Accurate control of correct recognition (%)	71.28	19.28	30.00–100.00
Accurate control of incorrect recognition (%)	46.23	29.12	0.00–100.00
Difference in study time between JOL low and JOL high (s)	-0.07	0.73	-1.75–1.52
Study time JOL low (s)	2.03	1.33	0.36–5.78
Study time JOL high (s)	1.86	0.90	0.61–4.08
Difference in CJs in the control of correct/incorrect recognition (mean CJ)	0.34	0.80	-1.75–2.50
CJ for accurate control of correct recognition (mean CJ)	3.45	0.62	1.67–4.00
CJs for accurate control of incorrect recognition (mean CJ)	3.03	0.87	1.00–4.00
Executive functions			
Cognitive flexibility (%)	62.24	16.50	28.00–95.00
Inhibition (s)	32.67	9.88	15.62–59.72
Working memory (trials)	5.36	1.94	1.00–10.00

“postdict” their performance, calibration (postdiction accuracy) was assessed for every participant by calculating the mean difference score in percentage (ranging from 0 to 100 %) between confidence judgments and corresponding recognition performance for the total of 20 items. To do so, confidence judgments provided on the 5-point Likert scale were transformed into percentages (0=0 %, 1=25 %, 2=50 %, 3=75 %, 4=100 %). The same was done for the recognition accuracy (incorrect=0 %, correct=100 %) to be able to compute the difference between confidence rating and recognition accuracy for every item. The lower the resulting difference score (bias) in percentage, the higher the level of correspondence between confidence judgment and actual performance.

A forced entry regression analysis was used to test whether performance estimation accuracy, the control measures age, sex, non-verbal IQ and vocabulary skills significantly predicted children’s ability to monitor retrieval on a global level (Table 2). The results indicated that the predictors explained 41.2 % of the variance, $F(5, 81)=11.34$, $p<.001$, with performance estimation accuracy significantly predicting the ability to postdict performance, $\beta=.64$, $t(81)=6.86$, $p<.001$. None of the other predictors turned out to be significant, $p_s \geq .75$.

Item-level metacognitive monitoring indices

Metacognitive discrimination in CJs Because children had the opportunity to study the items at their own pace after having given the JOLs and before the recognition test, item-by-item monitoring accuracy is addressed via confidence judgments only.

In a first step, a measure of item-level metacognitive discrimination, i.e., the difference in CJs for correctly recognized items compared to CJs for incorrectly recognized items was assessed by calculating a discrimination score (mean CJs correct recognition – mean CJs for incorrect recognition) for each participant. Using a forced entry regression analysis, it was

evaluated whether performance estimation accuracy and the control measures (age, sex, non-verbal IQ and vocabulary skills) significantly predicted the ability to monitor retrieval on the item level (Table 2). It was found that the predictors explained 17.5 % of the variance, $F(5, 78)=3.32$, $p<.01$, with performance prediction accuracy, $\beta=-.34$, $t(78)=-3.02$, $p<.01$, being the only significant predictor of item-level metacognitive discrimination.

Resolution (Gammas) To determine whether the 6-year-old children in our sample were able to discriminate in the memorability of the different to-be-remembered items after the actual memory test, non-parametric gamma correlations between confidence judgments and recognition performance (i.e., resolution) for each participant were calculated. A total of $n=13$ children dropped out of the analysis because of missing variability in their judgments. For the remaining 80 children, a gamma correlation of $M=.26$ ($SD=.55$) resulted. This correlation was significantly greater than 0, $t(79)=4.25$, $p<.001$, indicating above chance accuracy of the confidence judgments in discriminating children's performance with respect to one Kanji relative to another.

In a next step, the ability to discriminate in the memorability of the Kanji was examined using forced entry regression analysis with performance estimation accuracy, age, sex, non-verbal IQ and vocabulary skills as predictors (Table 2). Results showed a total of 17.1 % of variance being explained, which did turned out to be significant, $F(5, 69)=2.85$, $p<.05$. Performance estimation accuracy, $\beta=-.26$, $t(69)=-2.11$, $p<.05$, and vocabulary skills, $p_s \geq .11$ $\beta=-.25$, $t(69)=-2.07$, $p<.05$, significantly predicted resolution accuracy.

Metacognitive control

Control of answers A descriptive analysis of the percentage of items correctly placed in the treasure chest (i.e., items placed in the treasure chest if correctly recognized) divided by the total percentage of correctly recognized items, and the percentage of items correctly placed in the trash can (i.e., items placed in the trash can if incorrectly recognized) divided by the total percentage of items incorrectly recognized, revealed that on average children controlled 71 % ($SD=19$ %) of their answers accurately in the treasure chest and 46 % ($SD=29$ %) of their incorrect answers accurately in the trash can.

In a next step, two separate multiple regression analyses using the forced entry method were run to test whether performance estimation accuracy and the control variables significantly predicted participants accurate control of answers in the treasure chest and the trash can

Table 2 Summary of the forced entry multiple regressions for the metacognitive monitoring variables

Variable	Calibration (Postdiction accuracy)			Metacognitive discrimination in CJs			Resolution (Gammas)		
	B	SE B	β	B	SE B	β	B	SE B	β
Performance estimation accuracy	2.03	0.30	.64***	-0.03	0.01	-.34**	-0.02	0.01	-.26*
Age	-0.16	0.72	-.02	-0.02	0.23	-.10	-0.04	0.02	-.18
Sex	-1.26	4.01	-.03	0.12	0.13	.10	0.16	0.13	.02
Non-verbal IQ	0.08	0.53	.02	0.03	0.02	.17	0.03	0.02	.23
Vocabulary skills	-.17	1.38	-.01	-0.01	0.04	-.03	-0.09	0.04	-.25*
R ²	.41			.18			.17		
n	87			84			75		

* $p<.05$. ** $p<.01$. *** $p<.001$

(Table 3). The percentage of items correctly recognized and placed in the treasure chest, and the percentage of items incorrectly recognized and placed in the trash can, both constituting measures of accurate control, were included in the analyses. The results of the regression for the accurate control of answers in the treasure chest, indicated that the five predictors explained 14.5 % of the variance, $F(5, 81)=2.57, p<.05$, with performance estimation accuracy ($\beta=-.33, t(81)=-2.93, p<.01$), as well as vocabulary skills ($\beta=-.31, t(81)=-2.73, p<.01$) significantly predicting the accurate control of answers in the treasure chest.

In the second regression including the accurate control of answers in the trash can, 16.2 % of variance was explained by the entered predictors, $F(5, 80)=3.09, p<.05$. The results indicated that both, performance estimation accuracy ($\beta=-.25, t(80)=-2.19, p<.05$) and non-verbal IQ ($\beta=.24, t(80)=2.18, p<.05$), significantly predicted the accurate control of answers in the trash can.

Metacognitive monitoring and control interplay

To examine the correspondence between monitoring judgments (JOLs and CJs, respectively) and their related control behavior (study time allocation and control decisions, respectively) in relation to the five predictors (performance estimation accuracy, age, sex, non-verbal IQ and vocabulary skills), two separate multiple regression analyses were run (Table 4).

JOLs and study time allocation For the first monitoring and control interplay measure, JOLs were analyzed in relation to study time allocation. Children in our sample spent on average 1.93 s ($SD=0.88$) studying the items in both self-paced study phases and most of the responses provided by the participants were either low JOLs (a judgment of 0 and 1 on the 5-point Likert scale) or high JOLs (a judgment of 3 and 4 on the 5-point Likert scale), in 19.1 % and 64.2 % of the cases, respectively. Therefore, we decided to include the mean difference in study time between items with low JOLs (a judgment of 0 or 1) and items with high JOLs (a judgment of 3 or 4) in seconds as a dependent measure. Using forced entry regression analysis, it was tested whether performance estimation accuracy and the control variables significantly predicted the study time difference between items with low and high JOLs. The five predictors were shown to explain 16.8 % of the variance. Only performance estimation accuracy significantly predicted study time differences between items with low and high JOLs,

Table 3 Summary of the forced entry multiple regressions for the metacognitive control variables

Variable	Control of answers correct recognition (treasure chest)			Control of answers incorrect recognition (trash can)		
	B	SE B	β	B	SE B	β
Performance estimation accuracy	-0.90	0.31	-.33**	-1.04	0.48	-.25*
Age	0.57	0.75	.08	-0.32	1.13	-.03
Sex	5.14	4.14	.13	6.86	6.27	.12
Non-verbal IQ	0.06	0.55	.01	1.82	0.83	.24*
Vocabulary skills	-3.90	1.43	-.31**	0.30	2.16	.02
R ²	.15			.16		
n	87			86		

* $p<.05$. ** $p<.01$. *** $p<.001$

Table 4 Summary of the forced entry multiple regressions for the metacognitive monitoring and control interplay

Variable	Study time as a function of JOLs			Control of answers as a function of CJs		
	B	SE B	β	B	SE B	β
Performance estimation accuracy	-43.16	14.88	-.46**	-0.03	0.01	-.26*
Age	2.26	39.14	.01	-0.32	0.27	-.13
Sex	8.99	208.68	.01	0.02	0.15	.02
Non-verbal IQ	-9.98	27.82	-.05	0.07	0.02	.37**
Vocabulary skills	-142.24	71.49	-.33	0.10	0.05	.20
R ²	.17			.31		
n	52			72		

* $p < .05$. ** $p < .01$. *** $p < .001$

$\beta = -.46$, $t(46) = -2.90$, $p < .01$, even though the regression analysis did not reach significance, $F(5, 46) = 1.90$, $p = .11$.

CJs and control of answers Regarding the second measure of the monitoring and control interplay, that is, the question of whether items correctly placed in the treasure chest and the trash can differed in their confidence ratings, the mean CJs for items correctly placed in the treasure chest (i.e., correctly recognized and placed in the treasure chest) and for items correctly placed in the trash can (i.e., incorrectly recognized and placed in the trash can) was considered. A forced entry regression analysis was run to evaluate if performance estimation accuracy and the control measures of age, sex, non-verbal IQ and vocabulary skills significantly predicted participants mean discrimination in CJs between items correctly placed in the treasure chest and items correctly placed in the trash can. The results showed that the predictors explained 30.5 % of the variance, $F(5, 66) = 5.79$, $p < .001$. Performance estimation accuracy ($\beta = -.26$, $t(66) = -2.36$, $p < .05$) significantly predicted the criterion variable, as did non-verbal IQ ($\beta = .37$, $t(66) = 3.35$, $p < .01$).

Executive functions

In a final multiple regression analysis, we examined whether the metacognitive ability of making accurate performance estimations might also be a significant predictor of children's executive functions. Using the forced entry method, it was tested whether performance estimation accuracy, age, sex, non-verbal IQ and vocabulary skills significantly predicted executive functions. The predictors were found to explain 2.0 % of variance, $F(5, 82) = .31$, $p = .91$, with none of the predictors reaching significance, $p_s \geq .26$.

For the subsample of 75 children for whom SES information was available, the same multiple regression analyses were run including SES as another potential predictor. SES turned out to be a significant predictor only in the analysis including EF as a dependent measure, $\beta = -.29$, $t(66) = -2.31$, $p < .05$, although the regression analysis itself did not reach significance, $F(6, 66) = 1.03$, $p = .41$ (explained variance = 8.0 %). As SES did not significantly predict any of the other dependent measures ($p_s \geq .07$), it was excluded from the analyses, except from the ANCOVA with group as factor and EF as the dependent measure (reported below), to prevent a loss of statistical power as a result of the smaller sample size.

Analyses of covariance (ANCOVAs)

The second part of the results section is focusing on group differences as a function of children's performance estimation accuracy for other metacognitive monitoring as well as control skills and executive functions. For these analyses, children were divided into three groups (under estimators, realists, over estimators) based on their performance estimation accuracy. All analyses were run with the same dependent measures used in the multiple regression analyses reported in the previous results sections.

Group formation The groups were formed based on children's overall performance predictions provided after the two self-paced study phases and before taking the recognition test (see Fig. 1 for a schematic of the task procedure). On average, children in the current sample predicted to remember the meaning of $M=12.30$ ($SD=5.10$) Kanji and correctly recognized $M=10.72$ ($SD=4.28$) out of the total of 20 Kanji. To see whether the participants differed in their performance estimation accuracy in the paired associate learning task in the first place, the number of correctly recognized items in the recognition test collapsed across both sessions (total of 20 Kanji) was subtracted from the total number of items the child predicted to remember, with the resulting absolute value of the difference between the two variables representing a "calibration" score for every participant (e.g., Dunlosky and Thiede 2013; Roderer and Roebbers 2013; van Loon et al. 2013). The resulting performance estimation accuracy ($M=1.58$, $SD=7.12$), met the criteria of a normal distribution, $W(93)=.98$, $p=.30$. Children were assigned to three groups (under estimators, realists, over estimators) by choosing half of the standard deviation of the sample's performance estimation accuracy (± 3.56 Kanji) as a cut-off point. Consequently, the first cut-point was set at -3.56 Kanji and the other cut-off point at $+3.56$ Kanji.

The group of under estimators ($n=26$) consisted of children whose performance prediction was more than 3.56 Kanji below their actual number of correctly recognized Kanji out of the total of 20 Kanji. The realists group ($n=29$) consisted of children whose performance prediction was between 3.56 Kanji below and 3.56 Kanji above their actual number of correctly recognized Kanji. Finally, the group called over estimators ($n=38$) consisted of children whose performance prediction was more than 3.56 Kanji above their actual number of correctly recognized Kanji.

Global metacognitive monitoring index The three groups of under estimators, realists and over estimators were formed based on the accuracy of making performance predictions before the actual recognition test, serving as a global monitoring index. To examine to what extent overall performance prediction accuracy generalizes to overall performance "postdictions", we examined whether the three groups differed in their ability to monitor retrieval. Therefore, an ANCOVA with group as factor, age, sex, non-verbal IQ and vocabulary skills as covariates and calibration as the dependent measure was run. The analysis revealed a significant group effect on calibration, $F(2, 80)=19.96$, $p<.001$, $\eta^2_p=.33$, after controlling for the covariates. Pairwise comparisons showed that under estimators had significantly lower difference scores ($M=10.07$, $SE=3.96$) in comparison to over estimators ($M=42.52$, $SE=3.30$), $p<.001$. The difference in calibration between realists ($M=20.63$, $SE=3.58$) and over estimators also reached significance, $p<.001$, with realists showing better calibration. In terms of costs or benefits, over estimators - when looking back on their performance - were too confident and gave higher CJs than the other two groups, resulting in a higher bias between CJs and recognition accuracy.

Item-level metacognitive monitoring indices

Metacognitive discrimination in CJs To examine possible group differences in the ability to monitor retrieval on an item level, an ANCOVA with group as factor, the variables age, sex, non-verbal IQ and vocabulary skills as covariates and metacognitive discrimination in CJs between correct and incorrect recognition as the dependent measure was run. The analysis revealed a significant effect of group on metacognitive discrimination, $F(2, 77)=5.92, p<.01, \eta^2_p=.13$, after controlling for the covariates. Pairwise comparisons indicated that under estimators showed significantly stronger metacognitive discrimination in their CJs between correctly and incorrectly recognized items ($M=.52, SE=.13$) in comparison to over estimators ($M=-.02, SE=.10$), $p<.01$. The difference in metacognitive discrimination between over estimators and realists ($M=.35, SE=.11$) also reached significance, $p<.05$. Thus, overestimating performance was related to less sophisticated abilities in monitoring retrieval on the item level.

Resolution accuracy Next, possible group differences between under estimators, realists and over estimators in the ability to discriminate in the memorability of the Kanji were examined running an ANCOVA with group as factor, the control variables as covariates, and resolution (Gammas) as dependent variable. No significant effect of group resulted, $F(2, 68)=2.30, p=.11, \eta^2_p=.06$. Although under estimators ($M=.45, SE=.12$) seemingly had better resolution on the descriptive level compared to realists ($M=.23, SE=.11$) and over estimators ($M=.10, SE=.11$), the three groups did not differ significantly from each other, $p_s \geq .11$.

Metacognitive control: control of answers

Two separate ANCOVAs were run to examine participants' ability to control their answers accurately in the treasure chest (if correctly recognized) and the trash can (if incorrectly recognized). Children's ability to accurately control their answers in the treasure chest and the trash can as a function of group is displayed in Fig. 2a and b, respectively.

First, an ANCOVA with group as factor, the covariates age, sex, non-verbal IQ and vocabulary skills, and the percentage of accurate control of correctly recognized items (treasure chest) as dependent measure was run. No significant effect of group on the accurate control of correct recognition resulted after controlling for the covariates, $F(2, 80)=2.55, p=.08, \eta^2_p=.06$. Hence, under estimators ($M=79.06, SE=4.06$), realists ($M=69.42, SE=3.67$) and over estimators ($M=67.21, SE=3.39$) did not differ significantly from each other in their control of correctly recognized Kanji, $p_s \geq .10$.

Next, to explore whether the three groups differed in their accurate control of incorrect recognition, an ANCOVA with group as factor, the same covariates, and the percentage of accurate control of incorrect recognition (trash can) as the dependent measure was run. The effect of group turned out to be significant, $F(2, 79)=5.49, p<.01, \eta^2_p=.12$, after controlling for the covariates. Pairwise comparisons revealed that under estimators ($M=63.03, SE=5.90$) had a significantly higher percentage of accurate control of incorrect recognition compared to over estimators ($M=37.12, SE=4.83$), $p<.01$, while no difference between under estimators and realists ($M=45.61, SE=5.24$) as well as realists and over estimators resulted, $p_s \geq .08$.

Taken together, the two analyses demonstrated that overestimation does not generally lead to lower accuracy in metacognitive control of answers. Still, over estimators, even though showing similar accuracy in the accurate control of correct recognition, were outperformed by the under estimators when it came to the control of incorrectly recognized items.

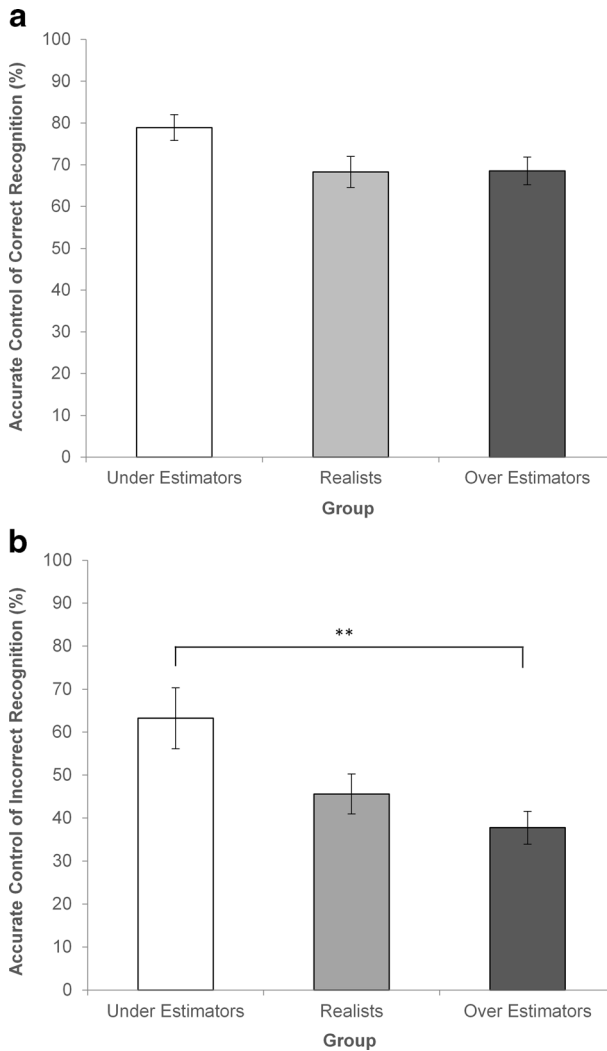


Fig. 2 Mean percentage of accurate control of correctly recognized Kanji in under estimators, realists and over estimators (a) and mean percentage of accurate control of incorrectly recognized Kanji in under estimators, realists and over estimators (b). Standard errors of the mean are represented by the error bars attached to each column. ** $p < .01$

Metacognitive monitoring and control interplay

JOLs and study time allocation In a first step, we were interested to see whether children in the three groups differed in their amount of time spent learning the Kanji in the self-paced study phases. Therefore, an ANOVA with group (under estimators vs. realists vs. over estimators) and overall allocated study time across session A and B for both self-paced study phases in seconds as dependent measure was run. No significant group effect resulted, $F(2, 85)=2.91$, $p=.06$, $\eta^2_p=.07$. Thus, over estimators ($M=1.69$, $SD=.73$) did not differ significantly from under estimators ($M=2.23$, $SD=1.07$) and realists ($M=2.02$, $SD=.85$) in their allocated study time.

To illustrate children's study time for low and high JOLs as a function of group, the descriptive values for the overall study time allocation are displayed in Fig. 3a. An ANCOVA was run to determine whether there were significant group differences in the overall study time allocated to items with low JOLs vs. high JOLs. The analysis with group as factor, age, sex, non-verbal IQ and vocabulary skills as covariates and the mean difference in overall study time for low vs. high JOLs in seconds (mean study time for items with low JOLs – mean study time for items with high JOLs) as dependent measure, revealed a significant group effect after controlling for the covariates, $F(2, 45)=4.12, p<.05, \eta^2_p=.15$.

Following up on the significant group effect, pairwise comparisons showed that under estimators ($M=.43, SE=0.19$) differed significantly from over estimators ($M=-.33, SE=0.17$) in their study time difference between items with low vs. high JOLs, $p<.05$, such that under estimators spent more time studying items which had previously received low JOLs compared to items with high JOLs, while the opposite was true for the group of over estimators. The difference in study time between low and high JOLs did not reach significance for the comparisons between realists ($M=-.10, SE=0.19$) and over estimators as well as realists and under estimators, $p_s \geq .15$.

CJs and control of answers For the second monitoring and control interplay measure, the mean CJs for items correctly placed in the treasure chest and for items correctly placed in the trash can for the three groups are illustrated in Fig. 3b.

To follow up on the descriptive group differences in the control of answers in relation to CJs presented in Fig. 3b, an ANCOVA with group as factor, the covariates age, sex, non-verbal IQ and vocabulary skills and the mean difference in CJs for items correctly placed in the treasure chest (i.e., correctly recognized and placed in the treasure chest) compared to items correctly placed in the trash can (i.e., incorrectly recognized and placed in the trash can) as dependent measure was run. A significant effect of group resulted after controlling for the covariates, $F(2, 67)=4.69, p<.05, \eta^2_p=.14$. Pairwise comparisons revealed that under estimators ($M=.52, SE=.13$) as well as realists ($M=.49, SE=.12$) differed significantly from over estimators ($M=.06, SE=.11$) in their discrimination in CJs of items correctly placed in the treasure chest compared to CJs of items correctly placed in the trash can, with items in the treasure chest having received higher CJs, $p_s < .05$. Realists and under estimators did not differ significantly from each other, $p=.10$. In other words, being overconfident in terms of one's overall performance prediction was not only linked to inferior metacognitive discrimination between correct and incorrect recognition but was also associated with a less systematic correspondence between monitoring (CJs) and the accurate control of answers.

Executive functions

Against the background of the above noted differences between under estimators, realists, and over estimators with regard to their (a) metacognitive discrimination and (b) their ability to act adequately on monitoring when it comes to metacognitive control (study time allocation as well as the accurate control of answers), we further explored the three groups' executive functioning skills. Therefore, we examined whether the metacognitive ability of making accurate performance estimations might be associated with superior executive functions. Accordingly, an ANCOVA with group as factor, the covariates age, sex, non-verbal IQ, vocabulary skills and SES, and a composite score of executive functions as dependent measure was performed. As we expected, based on the multiple regression analyses with the entire sample (see above), the effect of group did not reach significance after controlling for the covariates, $F(2, 65)=.16, p<.86, \eta^2_p=.01$. Thus, over- or underestimating one's performance was not significantly linked to differences in executive functions.

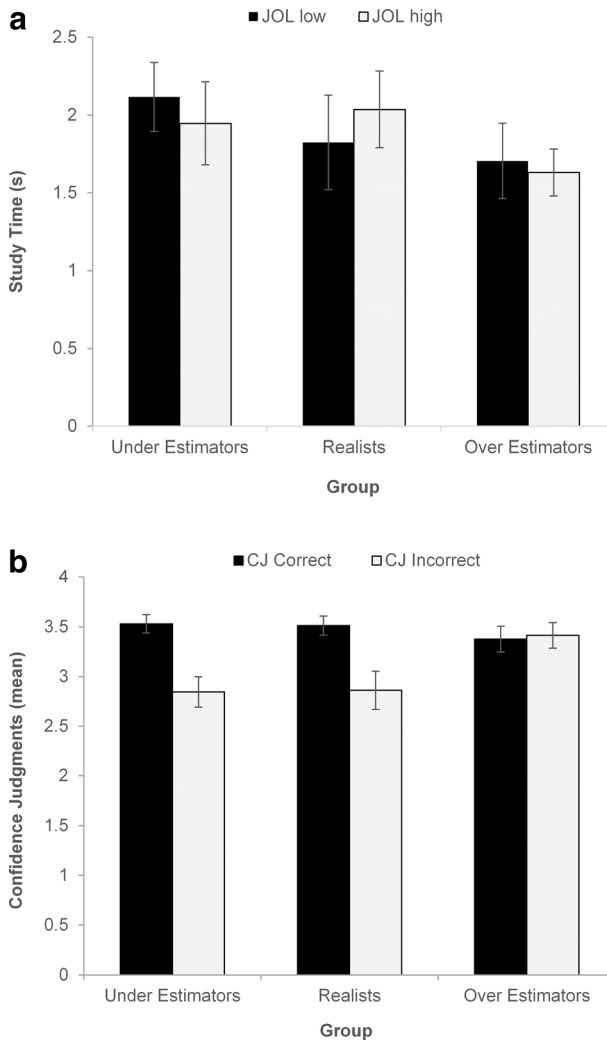


Fig. 3 Mean study time (s) for low (0 and 1) and high (3 and 4) judgments of learning (JOLs) in under estimators, realists and over estimators (a) and mean confidence judgments (CJs) for the accurate control of correct and incorrect recognition in under estimators, realists and over estimators (b). Standard errors of the mean are represented by the *error bars* attached to each column. * $p < .05$. ** $p < .01$

Discussion

The literature focusing on children's performance estimation accuracy has discussed potential costs as well as benefits of overestimating one's own performance for learning outcomes (e.g., Bjorklund and Bering 2002; Marazita and Merriman 2004; Roderer and Roebbers 2013; Shin et al. 2007). In this sense, the present study provides insights into the potential *costs* of the ability to accurately estimate one's own performance with respect to global as well as item-level monitoring skills, metacognitive control abilities, and executive functions as another, related aspect of self-regulation. It was demonstrated that there were large differences in children as young as six years in their ability to estimate their performance. Overestimating

performance was not a uniform phenomenon in the current study: rather, performance predictions were normally distributed, ranging from under- to pronounced overestimation. Finding such variation and a normal distribution in 6-year-olds' performance estimation accuracy was surprising as many previous studies suggested a general and uniform tendency of overconfidence in young children in various social, physical and cognitive tasks (e.g., Lipko et al. 2012; Plumert and Schwebel 1997; Schneider 1998; Shin et al. 2007).

Although there was a normal distribution of prospective under- and overestimation of performance, it is important to notice that when it came to the monitoring of single items, children were still uniformly overconfident. In fact, when making JOLs, judgments of "very unsure" and "unsure" were only used in 19 % of the cases, while in 64 % of the cases children chose being "very sure" and "sure" (for CJs: 10 % and 76 %, respectively), and even underestimators gave a mean CJ of "being rather sure" after incorrect responses, suggesting that even the underestimating children were still relatively confident regarding incorrectly recognized items. In previous studies addressing the age-related progression in the ability to monitor correct and incorrect responses, this pattern of results has also been documented and it was argued that the ability to monitor certainty might be acquired earlier in childhood than the ability to monitor uncertainty (e.g., Roebers et al. 2007).

Overestimation generalizes to postdicting performance

In the present investigation, the importance of being able to estimate one's performance accurately even in young children was highlighted by the significant regression analysis, in which solely the predictor of performance estimation accuracy explained 41.2 % of variance in postdicting performance. It was demonstrated that 6-year-olds' overestimation in predicting performance prospectively also led to a greater bias in "postdicting" overall performance retrospectively. Presumably, there are different factors involved in children's metacognitive judgments which might explain the overestimation or "bias" in judging their own performance. One such explanation for the overestimation in overall performance predictions of pre-schoolers and kindergarteners involves wishful thinking processes (e.g., Schneider 1998; Stipek et al. 1984), such that young children have difficulties in differentiating between what they expect to achieve and what they wish to achieve. It has also been argued that young children typically overestimate their performance because of an inability to differentiate effort from actual ability, in the sense that trying hard should automatically lead to superior performance (Nicholls 1978). Finally, young children's monitoring skills have recently been shown to be closely related to their (very positive) self-concept of abilities (Roebers et al. 2012) and this general and very positive perspective on their achievement might bias their metacognitive judgments on the global as well as on the item-level. Nevertheless, the exact costs or benefits of a generalized positively biased view of one's future and past performance were not addressed in the present paper. It therefore remains to be investigated whether overestimation in performance pre- and postdiction is beneficial in the long run, in terms of self-efficacy, motivation and effort investment.

Costs of overestimation in terms of metacognitive discrimination

To our knowledge, this study is among the first to show that global overestimation led to lower levels of monitoring accuracy on the item-level, namely the discrimination of confidence judgments between correct and incorrect recognition. Based on theoretical assumptions, an individual can only execute adequate control processes when relying on accurate metacognitive judgments (i.e., CJs discriminating reliably between correct and incorrect

responses), ideally benefiting overall performance (e.g., Metcalfe 2009; for similar results with third and fifth graders see; Roebers et al. 2014). In our study, overestimation came at significant costs of a 6-year-old child's ability to experience and report such accurate and fine-tuned metacognitive monitoring processes on the item level. Consequently, efficient metacognitive control behaviors might be impeded (e.g., differentially allocating study time based on the metacognitive judgments, revising longer or re-studying uncertain items, correcting errors – see below).

Costs of overestimation in metacognitive judgments: missing realism

Bearing in mind the positive effect of accurate metacognitive judgments, underestimating children in the current sample had in fact turned into “realists” when looking back at their performance, while over estimators still missed a sufficient degree of realism in their self-perceptions. One possible interpretation for under estimator's changed self-perception from prediction to postdiction is that these children were able to take their performance into account when retrospectively estimating their overall performance. Or, some individuals may under estimate their performance ahead of time to protect themselves (and others) against disappointment later on, and to activate sufficient effort (e.g., prolonged study time) to master the task to the degree they wish to achieve. In fact, a study with college students demonstrated that those participants who held the strongest concerns about other people's opinion spent more time studying for tests (Lin-Agler et al. 2004). In the present study, however, only partial confirmation for this interpretation was found: even though performance estimation accuracy was a meaningful predictor of item-level monitoring skills, only approximately 18 % of variance was explained, leaving room for other potential predictors in explaining children's accuracy in metacognitive judgments.

Personality factors might be candidate factors in explaining individual differences in monitoring accuracy (i.e., over- and underestimation). To our knowledge, the potential influence of personality factors on metacognitive skills has only been investigated in adults (e.g., Buratti et al. 2013; Pallier et al. 2002; Schaefer et al. 2004). In the study by Buratti et al. (2013), an association between overconfidence and the personality traits of “extraversion” and “openness” was reported, such that individuals scoring higher on these traits were more likely to be overconfident. Against this background, it may seem that our underestimating 6-year-olds were not generally pessimistic; instead, they appeared to be relatively cautious towards their own task mastery. Future studies should try to directly link metacognition to personality traits such as self-concept and self-efficacy, possibly illuminating potential sources of individual differences in metacognitive abilities.

Costs of overestimation in terms of control and the interplay of monitoring and control

First of all, costs of overestimating future performance were not only found with respect to monitoring, but also with respect to control processes. Even though children in all three groups were equally good at placing correctly recognized Kanji into the treasure chest, a meaningful difference in the accurate control of incorrect answers in the trash can resulted: over estimators had a significantly lower percentage of Kanji accurately placed into the trash can (i.e., incorrectly recognized items), compared to the under estimators. Possibly, the overestimating children made their decisions based on very optimistic and rather “biased” perceptions, resulting in more incorrect items being erroneously placed into the treasure chest instead of the trash can. In fact, over estimators put 60 % of the incorrectly recognized items into the trash can while under estimators showed this tendency of inadequate metacognitive control in only

17 % of the incorrectly recognized items. The fact that over estimators remained overconfident in their confidence ratings after taking the test (e.g., showing a greater bias in their postdictions) might support this explanation.

Most importantly, costs of overestimation were then also found in the interplay of monitoring and control processes. Even though over estimators did not necessarily terminate their self-paced study time prematurely in comparison to under estimators, it was nevertheless found that overestimation led to less sensible study time decisions in relation to monitoring judgments (low vs. high JOLs). Hence, under estimators but not over estimators spent more time studying items which had previously received low JOLs. Similarly, being overconfident when predicting performance resulted in less sophisticated control of answers in relation to certainty: the control of answers of under estimators (placing items in the treasure chest and the trash can) was based on more differentiated monitoring (confidence judgments).

Despite the cost of overestimating performance for study time allocation in relation to low and high JOLs, it must be noted that these study time differences were small. Generally, children only spent about 2 s studying each item. When further taking into account the mean recognition performance of only 11 out of 20 Kanji, all 6-year-olds of the current sample might have benefited from a more prolonged allocation of study time, regardless of group membership. On the one side, these results demonstrate that there are costs of overestimating one's performance concerning the interplay of monitoring and control in children as young as 6 years old, supported by the result of performance estimation accuracy being a significant predictor in both regression analyses on the monitoring-control interplay. On the other side, our findings confirm that children's metacognitive abilities to monitor and control lower levels of certainty (i.e., especially uncertainty) are still developing and are becoming increasingly sophisticated throughout middle childhood (e.g., Hoffmann-Biencourt et al. 2010; Krebs and Roebers 2012; Lockl and Schneider 2003).

Circumscribed costs of overestimation

Even though we did find differences between over and under estimators when considering metacognitive skills, and single significant associations between metacognitive indicators and executive functions in the correlational analyses for the whole sample (see Appendix A) resulted, these links were not strong enough to yield substantial group differences in executive functions. A link between metacognitive control and executive functions has been reported in previous studies with adults and children (e.g., Bryce et al. *in press*; DeMarie et al. 2004; Fernandez-Duque et al. 2000; Roebers et al. 2012). These studies suggest circumscribed rather than global common grounds of metacognition and executive functions, possibly due to the fact that metacognitive processes are strongly dependent on the task at hand while executive functioning skills are generally thought to be involved in many everyday life activities, including learning and memory. Thus, the costs of metacognitive overestimation are important to be considered, yet fortunately, they appear to be circumscribed and domain-specific to learning and memory processes.

Furthermore, although we measured EF with commonly used tasks, there is disagreement on the tasks which best exploit cognitive flexibility, inhibition and working memory, and there are ongoing discussions about the underlying structure of EF (i.e., one-factorial vs. multi-factorial structure) in childhood and adolescence (see Best and Miller 2010 for a review). In the current study we focused on the relation between a composite score of EF (entailing cognitive flexibility, inhibition and working memory) and performance estimation accuracy as one measure of metacognitive monitoring. The regression analysis conducted with performance estimation as a predictor only explained 8 % of variance in the composite score of EF

(taking into account SES). As suggested by (Bryce et al. *in press*), specific executive functions might be more closely related to monitoring and control processes. In their recently published study on the association between EF and metacognitive skills, inhibitory control was related to monitoring in 5- to 7-year-old children. In this sense, more research focusing on how different aspects and tasks of EF are related to the various aspects of both, monitoring and control throughout childhood seems warranted.

Methodological challenges when quantifying monitoring skills in young samples

Admittedly, the Gamma correlations, a classical and widely used measure of relative monitoring accuracy, did not yield substantial differences between under estimators, realists, and over estimators. While Gammas are based on pairwise comparisons between monitoring judgments (and their respective recognition performance), metacognitive discrimination is based on the whole set of monitoring judgments (as a function of their respective recognition performance). Consequently, these measures may differ in their sensitivity, with Gammas possibly not being able to detect the subtle differences observed in this young age group in which metacognitive skills seem to fluctuate more strongly than in older samples. Moreover, studies with adults have shown that it is more difficult to find individual differences in within-subject measures of metacognition compared to global performance predictions and postdictions (e.g., Dunlosky and Thiede 2013; Maki and McGuire 2002). Our findings with 6-year-olds likely support this methodological interpretation, as no association between postdictions and the ability to discriminate among the differences in the memorability of the Kanji was found (see correlations in Appendix A) and Gamma coefficients were generally rather low in our sample.

Conclusions

Taken together, the present findings contribute to a better understanding of 6-year-old children's emerging ability to estimate their performance in relation to other metacognitive monitoring and control skills, on the global as well as on the item-level. Our findings primarily speak for costs of young children's overestimation: children who overestimated their own performance had less accurate monitoring as well as control skills compared to realistic, and most of all, to children who underestimated their own performance. These results are further supported by the regression analyses: performance estimation accuracy on the whole seems to be a good indicator for the different metacognitive skills reported. In addition, the findings have implications for learning in educational contexts: when learning new material, overestimating one's own performance could lead to a premature termination of study time or less sophisticated error correction, resulting in lower performance. It might therefore be of particular importance for educators to help children even at this young age to identify difficult and potentially incorrect items, supporting beneficial control strategies in the process of self-regulated learning. It may actually be valuable to model useful control strategies in classrooms and to emphasize the role of monitoring for self-regulation recurrently and in various school domains.

Acknowledgments This project was partially financed by a grant of the Swiss National Science Foundation (SNSF-Grant No.100014_126559 provided to Claudia M. Roebbers). We would like to thank the participating children and their families, the children's teachers and the student research assistants who helped with data collection.

Appendix A

Table 5 Intercorrelations (Spearman-rho) of the study variables

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Age in Months	–												
2 Sex	-.02	–											
3 Non-verbal IQ	.25*	.03	–										
4 Vocabulary skills	-.08	.16	.03	–									
5 EF composite	.05	.01	.07	-.03	–								
6 Performance estimation accuracy	.01	.08	-.20	-.32**	-.17	–							
7 Global-level monitoring: Postdiction	-.04	.02	-.13	-.27*	-.28**	.66**	–						
8 Item-level monitoring: Discrimination CJs	.00	-.01	.25*	.08	.11	-.40**	-.32**	–					
9 Item-level monitoring: Relative accuracy	-.07	-.07	.25*	-.16	.02	-.31**	-.17	.81**	–				
10 Accurate control of correct recognition	.11	.05	.05	-.17	.22*	-.17	-.09	-.05	.07	–			
11 Accurate control of incorrect recognition	.04	.11	.30**	.10	-.04	-.33**	-.25*	.28**	.29*	-.31**	–		
12 ST Difference JOL high/low	-.01	-.05	-.01	-.13	.12	-.23	-.10	.04	.16	.11	.15	–	
13 CJ Difference in the control of Correct/Incorrect recognition	.00	-.06	.41**	.27*	.09	-.43**	-.46**	.78**	.57**	.02	.35**	.03	–

EF=Executive Functions. ST=Study Time. JOL=Judgment of Learning. CJ=Confidence Judgment. * $p < .05$. ** $p < .01$

References

- Allwood, C. M. (2010). The realism in children's metacognitive judgments of their episodic memory performance. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (pp. 149–169). New York: Springer.
- Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology*, 5, 115–129. doi:10.1076/chin.5.2.115.3167.
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, 81, 1641–1660. doi:10.1111/j.1467-8624.2010.01499.x 21077853.
- Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review*, 29, 180–200. doi:10.1016/j.dr.2009.05.002 20161467.
- Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, 21, 327–336. doi:10.1016/j.lindif.2011.01.007.
- Bjorklund, D. F., & Bering, J. M. (2002). The evolved child applying evolutionary developmental psychology to modern schooling. *Learning and Individual Differences*, 12, 1–27. doi:10.1016/S1041-6080(02)00047-X.
- Bjorklund, D. F., Coyle, T. R., & Gaultney, J. F. (1992). Developmental differences in the acquisition and maintenance of an organizational strategy - Evidence for the utilization deficiency hypothesis. *Journal of Experimental Child Psychology*, 54, 434–448. doi:10.1016/0022-0965(92)90029-6.
- Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *Test of nonverbal intelligence (TONI-3)* (3rd ed.). Austin: Pro-Ed.
- Bryce, D., Whitebread, D., & Szűcs, D. (in press). The relationships among executive functions, metacognitive skills and educational achievement in 5 and 7 year-old children. *Metacognition and Learning*, doi:10.1007/s11409-014-9120-4.
- Buratti, S., Allwood, C. M., & Kleitman, S. (2013). First- and second-order metacognitive judgments of semantic memory reports: The influence of personality traits and cognitive styles. *Metacognition and Learning*, 8, 79–102. doi:10.1007/s11409-013-9096-5.
- Clayson, D. E. (2005). Performance overconfidence: Metacognitive effects or misplaced student expectations? *Journal of Marketing Education*, 27, 122–129. doi:10.1177/0273475304273525.
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44, 2037–2078. doi:10.1016/j.neuropsychologia.2006.02.006.
- DeMarie, D., Miller, P. H., Ferron, J., & Cunningham, W. R. (2004). Path analysis tests of theoretical models of children's memory performance. *Journal of Cognition and Development*, 5, 461–492. doi:10.1207/s15327647jcd0504_4.
- Destan, N., Hembacher, E., Ghetti, S., & Roebers, C. M. (2014). Early metacognitive abilities: The interplay of monitoring and control processes in 5- to 7-year-old children. *Journal of Experimental Child Psychology*, 126, 213–228.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168. doi:10.1146/annurev-psych-113011-143750.
- Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition*, 25, 691–700. doi:10.3758/BF03211311.
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks: Sage.
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, 22, 271–280. doi:10.1016/j.learninstruc.2011.08.003.
- Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction*, 24, 58–61. doi:10.1016/j.learninstruc.2012.05.002.
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12, 83–87. doi:10.1111/1467-8721.01235.
- Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. *Consciousness and Cognition*, 9, 288–307. doi:10.1006/ccog.2000.0447.
- Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of learning. *Learning and Instruction*, 32, 1–9. doi:10.1016/j.learninstruc.2014.01.001.
- Flavell, J. H. (1979). Meta-cognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906–911. doi:10.1037/0003-066x.34.10.906.
- Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology*, 1, 324–340. doi:10.1016/0010-0285(70)90019-8.

- Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*, 160–170. doi:10.1037//0022-0663.92.1.160.
- Hembacher, E., & Ghetti, S. (2013). How to bet on a memory: Developmental linkages between subjective recollection and decision making. *Journal of Experimental Child Psychology, 115*, 436–452. doi:10.1016/j.jecp.2013.03.010.
- Hoffmann-Biencourt, A., Lockl, K., Schneider, W., Ackerman, R., & Koriat, A. (2010). Self-paced study time as a cue for recall predictions across school age. *British Journal of Developmental Psychology, 28*, 767–784. doi:10.1348/026151009x479042.
- Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology, 35*, 20–36. doi:10.1080/87565640903325691.20390590.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517. doi:10.1037/0033-295x.103.3.490.
- 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. doi:10.1348/000709910x485719.
- Krebs, S. S., & Roebers, C. M. (2012). 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. doi:10.1007/s11409-011-9079-3.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134. doi:10.1037//0022-3514.77.6.1121.
- Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development, 84*, 1933–1953. doi:10.1111/Cdev.12096.
- Lin-Agler, L. M., Moore, D., & Zabrocky, K. M. (2004). Effects of personality on metacognitive self-assessments. *College Student Journal, 38*(3), 453–461.
- Lipko, A. R., Dunlosky, J., Hartwig, M. K., Rawson, K. A., Swan, K., & Cook, D. (2009). Using standards to improve middle school students' accuracy at evaluating the quality of their recall. *Journal of Experimental Psychology: Applied, 15*, 307–318. doi:10.1037/a0017599.20025417.
- Lipko, A. R., Dunlosky, J., Lipowski, S. L., & Merriman, W. E. (2012). Young children are not underconfident with practice: The benefit of ignoring a fallible memory heuristic. *Journal of Cognition and Development, 13*, 174–188. doi:10.1080/15248372.2011.577760.
- Lockl, K., & Schneider, W. (2003). Metacognitive monitoring and self-control processes for children's allocation of study time. *Zeitschrift für Paedagogische Psychologie, 17*, 173–183. doi:10.1024//1010-0652.17.3.173.
- Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development, 82*, 1178–1187. doi:10.1111/j.1467-8624.2011.01649.x.21954871.
- Lyons, K. E., & Zelazo, P. D. (2011). Monitoring, metacognition, and executive function: Elucidating the role of self-reflection in the development of self-regulation. In B. J. Benson (Ed.), *Advances in child development and behavior* (pp. 379–412). Oxford: Elsevier.
- Maki, R. H., & McGuire, M. J. (2002). Metacognition for text: Implications for education. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied Metacognition* (pp. 39–67). Cambridge: Cambridge University Press.
- Marazita, J. M., & Merriman, W. E. (2004). Young children's judgment of whether they know names for objects: The metalinguistic ability it reflects and the processes it involves. *Journal of Memory and Language, 51*, 458–472. doi:10.1016/j.jml.2004.06.008.
- Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*, 159–163.
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15*, 174–179. doi:10.3758/PBR.15.1.174.
- Metcalfe, J., & Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and Learning, 8*, 19–46. doi:10.1007/s11409-013-9094-7.
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*, 8–14. doi:10.1177/0963721411429458.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100. doi:10.1006/cogp.1999.0734.
- Nelson, T. O., Dunlosky, K. J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5*, 207–213. doi:10.1111/j.1467-9280.1994.tb00502.x.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). New York: Academic.

- Nicholls, J. G. (1978). Development of concepts of effort and ability, perception of academic attainment, and understanding that difficult tasks require more ability. *Child Development, 49*, 800–814. doi:10.1111/j.1467-8624.1978.tb02383.x.
- Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., et al. (2002). The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology, 129*, 257–299. doi:10.1080/00221300209602099.
- Pickering, S. J., & Gathercole, S. E. (2001). *Working memory test battery for children (WMTB-C)*. London: Psychological Cooperation.
- Plumert, J. M., & Schwebel, D. C. (1997). Social and temperamental influences on children's overestimation of their physical abilities: Links to accidental injuries. *Journal of Experimental Child Psychology, 67*, 317–337. doi:10.1006/jecp.1997.2411.
- Ricken, G., Fritz, A., Schuck, K. D., & Preuß, U. (2007). *HAWIVA-III, Hannover-Wechsler-Intelligenztest fuer das Vorschulalter-III. Manual zur Durchfuehrung und Auswertung* (3rd ed.). Goettingen, Germany: Huber.
- Roderer, T., & Roebbers, C. M. (2009). *Children's strategic regulation of memory accuracy* (pp. 253–274). Hauppauge: Nova.
- Roderer, T., & Roebbers, C. M. (2010). Explicit and implicit confidence judgments and developmental differences in metamemory: An eye-tracking approach. *Metacognition and Learning, 5*, 229–250. doi:10.1007/s11409-010-9059-z.
- Roderer, T., & Roebbers, C. M. (2013). Children's performance estimation in mathematics and science tests over a school year: A pilot study. *Electronic Journal of Research in Educational Psychology, 11*(1), 5–24.
- Roderer, T., & Roebbers, C. M. (2014). 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. *Metacognition and Learning, 9*, 1–23. doi:10.1007/s11409-013-9109-4.
- Roebbers, C. M., Cimeli, P., Roethlisberger, M., & Neuenschwander, R. (2012). Executive functioning, metacognition, and self-perceived competence in elementary school children: An explorative study on their interrelations and their role for school achievement. *Metacognition and Learning, 7*, 151–173. doi:10.1007/s11409-012-9089-9.
- Roebbers, C. M., & Kauer, M. (2009). Motor and cognitive control in a normative sample of 7-year-olds. *Developmental Science, 2*, 175–181. doi:10.1111/j.1467-7687.2008.00755.x 19120425.
- Roebbers, C. M., Krebs, S. S., & Roderer, T. (2014). Metacognitive monitoring and control in elementary school children: Their interrelations and their role for test performance. *Learning and Individual Differences, 29*, 141–149. doi:10.1016/j.lindif.2012.12.003.
- Roebbers, 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. doi:10.1348/026151006X104392.
- Roebbers, C. M., von der Linden, N., & Howie, P. (2007). Favourable and unfavourable conditions for children's confidence judgments. *British Journal of Developmental Psychology, 25*, 109–134. doi:10.1348/026151006X104392.
- Roethlisberger, M., Neuenschwander, R., Michel, E., & Roebbers, C. M. (2010). Executive functions: Underlying cognitive processes and their correlates in late preschool children. *Zeitschrift fuer Entwicklungspsychologie und Paedagogische Psychologie, 42*, 99–110. doi:10.1026/0049-8637/A000010.
- Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. K. (2004). Overconfidence and the big five. *Journal of Research in Personality, 38*, 473–480. doi:10.1016/j.jrp.2003.09.010.
- Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science, 1*, 291–297. doi:10.1111/1467-7687.00044.
- Schneider, W., Vise, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring - Evidence from a judgment-of-learning task. *Cognitive Development, 15*, 115–134. doi:10.1016/S0885-2014(00)00024-1.
- Shin, H., Bjorklund, D. F., & Beck, E. F. (2007). The adaptive nature of children's overestimation in a strategic memory task. *Cognitive Development, 22*, 197–212. doi:10.1016/j.cogdev.2006.10.001.
- Sinkavich, F. J. (1995). Performance and metamemory: Do students know what they don't know? *Journal of Instructional Psychology, 22*(1), 77–87.
- Son, L. K. (2005). Metacognitive control: Children's short-term versus long-term study strategies. *Journal of General Psychology, 132*, 347–363. doi:10.3200/Genp.132.4.347-364.
- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning Memory and Cognition, 26*, 204–221. doi:10.1037/0278-7393.26.1.204.
- Stipek, D. J., Roberts, T. A., & Sanborn, M. E. (1984). Preschool-age children's performance expectations for themselves and another child as a function of the incentive value of success and the salience of past performance. *Child Development, 55*, 1983–1989. doi:10.1111/j.1467-8624.1984.tb03896.x.

- Sussan, D., & Son, L. K. (2007). The training of metacognitive monitoring in children. *Columbia Undergraduate Science Journal*, 2.
- Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition*, 3, 161–164. doi:10.1016/j.jarmac.2014.02.001.
- Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology*, 45, 1097–1118. doi:10.1037/a0015864.19586182.
- van Loon, M. H., de Bruin, A. B. H., van Gog, T., & van Merriënboer, J. J. G. (2013). Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration. *Learning and Instruction*, 24, 15–25. doi:10.1016/j.learninstruc.2012.08.005.
- Visé, M., & Schneider, W. (2000). Determinanten der Leistungsvorhersage bei Kindergarten- und Grundschulkindern: Zur Bedeutung metakognitiver und motivationaler Einflußfaktoren. *Zeitschrift fuer Entwicklungspsychologie und Paedagogische Psychologie*, 32, 51–58. doi:10.1026//0049-8637.32.2.51.
- Wechsler, D. (2002). *The Wechsler Preschool and primary scale of intelligence (WPPSI-III)* (3rd ed.). San Antonio: The Psychological Cooperation.
- Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, 44, 575–587. doi:10.1037/0012-1649.44.2.575.18331145.