Metacognition and Learning
https://doi.org/10.1007/s11409-019-09217-4

A comparison of non-verbal and verbal indicators of young children's metacognition

Claudia M. Roebers[1] · Sonja Kälin[1] · Eva A. Aeschlimann[1]

Received: 6 May 2019 / Accepted: 21 November 2019/Published online: 13 December 2019

Abstract
In the past decade, research increasingly uncovers emerging metacognitive skills in young children by using child-friendly, creative, and non-verbal measures of metacognition. In the present study, the now often used non-verbal "opt-out" paradigm and classical verbal metacognitive judgments (judgments-of-learning and confidence judgments) were used in the context of a paired associates learning task. Prospectively (before the memory test) and retrospectively (after recognition) $N = 138$ children ($N = 72$ 5-year-olds; $N = 66$ 6-year-olds) evaluated their performance. Results revealed evidence for existing metacognitive skills in their relative sense (discriminating correct from incorrect responses in the metacognitive measures) but poor metacognitive accuracy in absolute terms. The non-verbal opt-out measures were not found to consistently and substantially facilitate especially 5-year-olds` metacognitive skills. Correlating non-verbal and verbal measures revealed shared but also distinct, yet to be explained variances calling for a strongly differentiated concept of children's early metacognition.

Keywords Metacognitive development · Non-verbal metacognition · Judgments-of-learning · Confidence judgments · Choice latencies

There is no doubt that different species can experience uncertainty. Young children, apes, monkeys, and even dolphins can selectively choose to skip a trial when uncertain (e.g., Shields et al. 1997). However, the traditional developmental literature on metacognition suggests pronounced deficits in young children that are only slowly overcome (Schneider and Löffler 2016). As children's paradigms heavily rely on verbal measures of introspection (O'Leary and Sloutsky 2017), researchers adapted non-verbal paradigms stemming from comparative and animal studies to study young children's metacognition (Smith et al. 2003). This literature reports metacognitive skills early on in a relative

✉ Claudia M. Roebers
claudia.roebers@psy.unibe.ch

[1] University of Bern, Hochschulzentrum vonRoll, Fabrikstrasse 8, 3012 Bern, CH, Switzerland

sense (i.e., the ability to discriminate between correct and incorrect responses; e.g., Geurten and Bastin 2018; Goupil and Kouider 2016; Lyons and Ghetti 2013). It remains yet unknown, however, whether verbal and non-verbal indicators of metacognition tap the same underlying processes and produce similar results in terms of absolute metacognitive accuracy (i.e., number of correct metacognitive judgments when considering objective response accuracy; e.g., Washburn et al. 2016). The present study will systematically compare verbal and non-verbal indicators of metacognition to fill this gap in the literature. More detailed knowledge of explicit verbal and non-verbal measures of metacognition will help to integrate different, yet poorly related lines of the metacognition literature.

Typical *verbal measures* of metacognition are judgments-of-learning (JoLs, prospective monitoring), and confidence judgments (CJs, retrospective monitoring; Dunlosky and Metcalfe 2009). For these, individuals indicate – item by item - how sure they are that they will later recognize the correct information (JoL) or how sure they are that they recalled that piece of information correctly (CJs). This is typically done on pictorial Likert-scales accompanied by verbal labels ranging from, for example, "I am very sure" to "I am very unsure". Although 5-year-olds can give reliably lower judgments when incorrect than when correct (i.e., *relative* monitoring accuracy), the absolute accuracy of their judgments, especially when recall or recognition fails, is relatively low. *Absolute* accuracy of monitoring is thereby defined as the absolute level of confidence in relation to the absolute level of performance. In other words, young children tend to be a little less sure for incorrect recall than for correct recall (i.e., emerging relative monitoring accuracy), but their level of certainty is much higher than their performance (i.e., poor absolute monitoring accuracy). It is especially the accurate monitoring of errors together with the reporting of uncertainty that is challenging for kindergarten and young elementary school children and that impedes their absolute monitoring accuracy (e.g., Destan et al. 2014; Lyons and Ghetti 2013).

The most commonly used non-verbal method to assess metacognition in non-verbal species is the so-called "opt-out" paradigm. It is theoretically assumed that to deal with varying degrees of certainty and uncertainty, individuals have to rely on cues available to them. According to the trace accessibility model of metacognition (Koriat 1997), such cues are experienced based, subjective, and internally generated assessments of the competing memory representations (i.e. memory interference). The "opt-out" response option allows solving the conflict between competing memory representations. Thus, whenever rats, dolphins, monkeys, apes, and humans use the "opt-out" response, this is interpreted as showing that individuals have access to characteristics of their (competing) memory representations. Individuals then strategically guide their behavior based on these cues and use the "opt-out" option to avoid errors, or more generally, to optimize performance (Smith 2005).

On the one side, opt-out paradigms are used to investigate prospective metacognitive aspects of learning, allowing participants to skip the upcoming recall of items when unsure (e.g., Balcomb and Gerken 2008). On the other side, opt-out paradigms are also used after recall (retrospective), giving participants the option to withdraw uncertain responses (e.g., Geurten and Bastin 2018; Hembacher and Ghetti 2014). Even young children seem to choose the opt-out option metacognitively, that is, the recall of declined items have a higher likelihood to produce incorrect answers in a later test than accepted items (relative monitoring accuracy in terms of discrimination). However, a closer look at the reported results nevertheless reveals low absolute metacognitive accuracy. For example, in the Balcomb and Gerken's study (2008), when young children chose the opt-out response, they were still correct in as many as 61% of these declined trials, indicating a less-than-optimal opt-out use. In Hembacher and Ghetti's study (2014) children's opt-out use indicated emerging metacognitive skills, as 5-year-olds' recall of rejected items had a lower likelihood to be correct in a later test than the accepted items (however, children were still inaccurate in their opting-out in 30 to 40% of the trials). Overall, this suggests that they experienced uncertainty and tried to solve this conflict by opting-out. Yet, it remains unclear from what age on young children reliably use the opt-out option for avoiding memory failures.

There is inconsistency in the literature concerning the definition of implicit and explicit indicators of metacognition (Geurten and Bastin 2018; Goupil and Kouider 2016). In this contribution, we follow Goupil and Kouider's definition (2016) considering opt-out responses as explicit as individuals non-verbally communicate their uncertainty when opting-out. Implicit measures, in contrast, are defined as behavioral measures derived during the experimental task of which the individual is not (fully) aware. In theory-of-mind research, for example, the inclusion of implicit measures (e.g., anticipatory looking in false belief tasks, eye gaze in violation of expectations paradigms; see Kulke and Rakoczy 2018; Ruffman et al. 2001) has lead researchers to assume earlier-than-expected implicit skills compared to explicit skills. Implicit measures thus allow uncovering developmental continuity from implicit to explicit measures of these higher order cognitions (Kuhn 1999; Thoermer et al. 2012).

Against this background, choice latencies in the recognition test may be used as an implicit measure of monitoring. In the adult literature, choice latencies are used to quantify memory trace accessibility (Koriat 1997): to the extent that an individual has a strong memory trace for the information in question, she or he will be able to respond fast to a question. This experience of responding fast serves as a cue, that is as a mnemonic experience for monitoring, with monitoring judgments being considered inferential in nature (De Bruin and van Merriënboer 2017; Dunlosky and Thiede 2013). In other words, the more difficult an item is, the longer it will take to respond. The longer retrieval time will inform monitoring, consequently leading to low confidence in that particular piece of information. These assumptions have been confirmed as research shows that longer response choice latencies are associated with lower metacognitive judgments (Brewer and Day 2005; Koriat 2012; Robinson et al. 1997; Zakay and Tuvia 1998). Thus, choice latencies in a recognition test appear to be a suitable implicit research measure quantifying metacognitive experiences that are assumed to be detectable early in development (Efklides 2011; Thoermer et al. 2012).

Only three developmental studies have addressed choice latencies as measures of implicit metacognition (Coughlin et al. 2015; Koriat and Ackerman 2010; Roebers et al. 2019). These studies found that young elementary school children` choice latencies for correct answers were substantially shorter than those for incorrect answers. Even more importantly, choice latencies were found to be a significant predictor of classical confidence judgments suggesting that they may be a valuable implicit measure of early and emerging monitoring skills (Roebers et al. 2019). However, the participating children in these previous studies were already in school, leaving the question unanswered as to whether choice latencies mirror metacognitive processes in preschool samples, too. While the present study aims to make a first step into exploring implicit measures of metacognition in young children, we will not yet be able to present a theoretical account for the relation between explicit and implicit measures. The developmental relations between the two are likely to be very complex and developmentally dynamic. Nevertheless, we consider it important to gather first data and stimulate research into this direction.

The present study

The present study aimed to assess children's ability to metacognitively discriminate between correct and incorrect responses (in a *relative* sense) in verbal judgments and non-verbal opt-out responses (both measured prospectively and retrospectively). We expected to find hints of metacognitive skills in all measures of metacognition (Balcomb and Gerken 2008; Coughlin et al. 2015). Moreover, we intended to explore the *absolute* metacognitive accuracy and to investigate whether verbal and non-verbal measures produce similar results to bridge these two different research traditions.

We targeted 5- and 6-year-olds because in this age range, strong developmental improvements in metacognitive monitoring have been established (Ghetti et al. 2013; Roebers 2017; Schneider and Löffler 2016). Moreover, this age range allows using verbal measures as these children have repeatedly been shown to be able to use a pictorial confidence scale in which the different degrees of certainty and uncertainty are verbally labelled (e.g., Destan et al. 2014). The use of non-verbal measures should also be possible in these age groups. We expected to find age differences between the participating 5- and 6-year-olds for verbal and non-verbal indicators of metacognition (in a relative and absolute sense). As two previous studies suggest that verbal reports of uncertainty are less accurate than non-verbal measures in *absolute* terms mirroring earlier detectable skills on non-verbal measures than on verbal measures (Geurten and Bastin 2018; Paulus et al. 2013; Thoermer et al. 2012), and we expected a similar pattern.

Another goal of the present study was to explore common processes between verbal judgments and opting-out. By linking prospective and retrospective, verbal and non-verbal absolute measures of metacognitive accuracy to each other, we aimed to estimate the shared processes. We expected to find substantial associations, albeit no perfect links as the opting-out responses may also involve control or executive processes (Nelson and Narens 1994), more so than verbal judgments of certainty.

Finally, we aimed to explore choice latencies in the recognition test as an implicit measure and relate them to metacognitive processes. We expected that children select the correct alternative faster than the incorrect one, possibly mirroring ongoing metacognitive monitoring processes in terms of which alternative answer has the highest likelihood to be correct. If that is the case, then choice latencies should share substantial amounts of variance with the opting-out and the verbal metacognitive judgments, both retrospectively and prospectively, as is found in school-aged children and adults (Koriat and Ackerman 2010; Roebers et al. 2019).

Method

Open practice statement

The reported study was not preregistered but the data are available on a third-party permanent archive. All materials can be obtained from the first author.

Participants

Power analysis was conducted estimating the needed sample size with G*Power 3 (Faul et al. 2007). As we expected a large age effects (metacognitive skills develop rapidly between the age of 5 to 7 years), and aimed to obtain a statistical power of about 80%, 65 to 70 children per predefined age group were needed. Participants included $N = 138$ children from two age groups: 5-year-olds ($n = 72$, 43% girls ($N = 31$), $M_{age} = 63.6$ months, $SD = 3.8$ months, age range 57–69 months), and 6-year-olds ($n = 66$, 42% girls ($N = 28$), $M_{age} = 76.0$ months, $SD =$ 3.1 months, age range 72–83 months). Participants were recruited from 14 different public preschools and kindergartens in urban and rural regions near a Swiss university town. These regions covered both lower and upper middle class residential areas; no further information on participants' socio-economic background was available. Written consent was obtained from the main caregiver of the participating children (positive return rate was 90%); all children gave oral consent before testing. The study was approved by the ethics committee of the faculty and was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. Fifteen additional participants were tested but were not included in the analyses, because of technical problems ($n = 9$), or an accuracy of 0% in the recognition test phase (see below) of the paired associates task ($n = 6$). These children seemed to select the alternatives randomly but we had no prior indications that they did not understand the task; these children also were neither among the youngest nor were they non-native speakers.

Material and procedure

The experimenters were trained to administer the task on a tablet computer (Lenovo, Yoga Tab 3 Pro with the software OpenSesame; Mathôt et al. 2012). Testing was conducted in a quiet room in the child's school. The tablet registered the answers of the participants (touches on the screen) and response times. As learning material, we used 10 non-associative picture pairs (e.g., frog and slide). Additionally, 22 pictures were used as distractor pictures. All pictures depicted familiar objects, animals, fruits, and vegetables.

For the monitoring judgments, participants rated their confidence on a seven-point Likert scale presented on the bottom of the touchscreen (see Fig. 1). The scale depicted a thermometer with different colors, ranging from blue (representing "very unsure") to red (representing "very sure"; Koriat and Shitzer-Reichert 2002) and the metaphor with red representing hot and

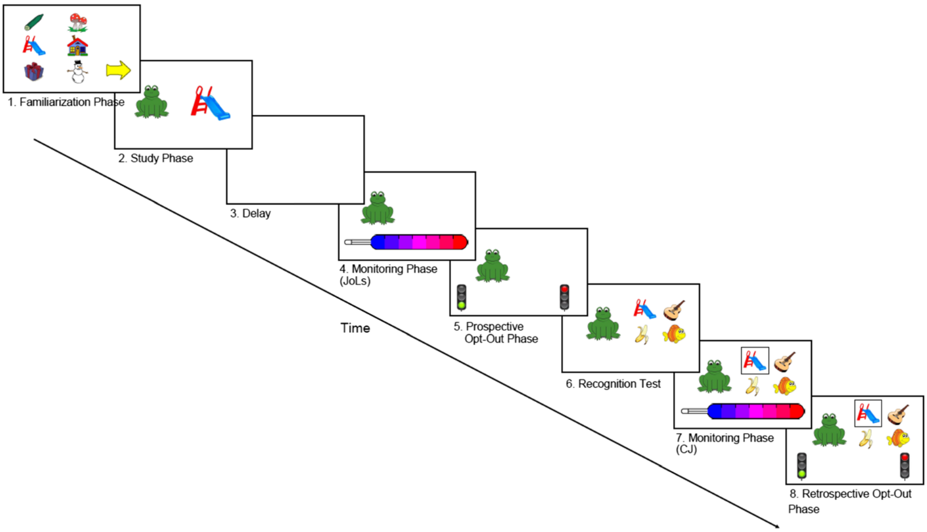

Fig. 1 Phases of the paired associates task

blue representing cold was explained. Example questions were asked for practicing the confidence ratings ensuring that children gave contrasting confidence judgments ("*What is the name of your teacher?*"- "*How sure are you about this?*"; "*When is my birthday?*" – "*How sure are you about this answer?*"). The categories in between very sure and very unsure were also explained; children easily understood the rationale of the scale, as self-ratings are also common kindergarten practice in Switzerland.

Paired associates task

Metacognitive processes were measured in the context of a standard paired associates learning task consisting of eight phases (e.g., Lockl and Schneider 2004; Tsalas et al. 2017; see Fig. 1):

(1) In the familiarization phase, to make sure children knew the items, participants saw seven times a screen with six pictures (20 pictures were pictures of the to-be-learned picture pairs and 22 pictures were distractor pictures). Participants were instructed to ask the experimenter if they did not know what a picture depicted. In addition, on each screen the experimenter randomly selected three of the six pictures and asked the child to name them.
(2) During the study phase, the 10 picture pairs were presented one after the other for a fixed duration of three seconds. Children were instructed to remember which pictures belonged together and were told that they will be tested later on.
(3) During the delay, children did a filler task (mazes) on a sheet of paper (3 min).
(4) In the first monitoring phase (JoLs), the "cue pictures" (i.e., the left item of the item pair) were presented, one at a time, together with the thermometer scale. Children had to rate on the thermometer how sure they were that they would remember the corresponding picture later on.
(5) In the prospective opt-out phase, the cue pictures were presented again, one at a time. Now, children had to decide whether they wanted to be tested on this particular pair in the upcoming test or not. If they wanted to be tested later on (i.e., they accepted the item), they should touch the green traffic light. If they did not want to be tested later on (i.e., they rejected the item), they should touch the red traffic light. Children understood the logic of the traffic lights with ease.
(6) In the recognition test, children were first told that they would be tested only on the accepted items. Then, the computer simulated a problem, which was "solved" quickly. Next, the experimenter explained that, because of the computer problem, the child would now nevertheless be tested on all picture pairs. Hence, all cue pictures were presented one after the other together with four response alternatives (the matching picture, one picture belonging to another pair, and two distractors from the familiarization phase). There was no time limit for responding. Children were asked to guess if they did not know the answer.
(7) In the retrospective monitoring phase (CJ), the cues were presented again, one after the other, together with the four response alternatives. The previously chosen response alternative was circled, and children were asked to rate how sure they were that their response was correct.
(8) In the final retrospective opt-out phase, each cue picture was presented together with the four response alternatives, with the child's selected response again being circled. Now, children could gain points by deciding whether they wanted that the selected response

counted or not for "their final grading". If they wanted that their response counted (i.e., they accepted the item response), they should touch the green traffic light. If they did not want that their response counted (i.e., they rejected the item response), they should touch the red traffic light. Children were told that they would get one point if (a) they selected the green traffic light and their response was correct and if (b) they selected the red traffic light and their response turned out to be incorrect. Of course, all children were given positive feedback about the points they had earned (irrespective of their performance), and received a small gift for participating.

Altogether, task completion took between 15 and 20 min per child.

Data coding

As dependent variables for recognition different recognition accuracy variables were used [e.g., over all recognition accuracy: percent correct in the recognition test; recognition accuracy (as proportion) for accepted or rejected trials]. Choice latencies of the recognition test were computed as mean response times from the start of the stimuli presentation until the child touched the screen averaged over 9 trails (the response times of the first trial were excluded as for this trial the instructions were repeated). As the experimenter closely supervised children individually, no post hoc outlier analyses and no replacement of outliers were performed.

To enable direct comparisons between the verbal and non-verbal measures (monitoring judgments and rejected/accepted responses), we dichotomized JoLs and CJs into low- and high-judged items based on the observed frequencies of the judgments. We classified items with a judgment of the level 1 to 4 into low-judged items (thereby accounting for young children's overconfidence; e.g., Hembacher and Ghetti 2014), and items with a judgment of the level 5 to 7 into high-judged items to obtain an approximately equal number of observations for data analyses (see also Destan et al. 2014). This way, 51% of the monitoring judgments were classified as low and 49% were classified as high providing an ideal database for the analyses reported below.

To quantify metacognitive accuracy, we applied the following scoring for each item. No credit was given (a) if recognition was correct but the corresponding item judgment/selection was low/rejected, or (b) if recognition was false but the item judgment/selection was high/accepted. One credit was awarded (a) if recognition was correct and the item judgment/selection was high/accepted or (b) if recognition was false and the item judgment/selection was low/rejected. Dependent variable was the percentage of accurate metacognitive responses, that is, the number of credits awarded following the above-mentioned coding in percent.

Results

Statistical analyses were run with SPSS (IBM SPSS Statistics, Version 24.0). We used an alpha level of 5% (two-tailed) for significance tests. If multiple comparisons were computed, results were corrected by the Bonferroni-Holm familywise error rate correction and the corrected critical *p*-values (*p*-crit) are reported (Holm 1979). As estimates of effect size, the partial eta squared (η^2_p) or Pearson's correlation coefficient (*r*) are reported. A small, medium, and large effect is defined as $r = .10$, $r = .30$, and $r = .50$, or as $\eta^2_p = .01$, $\eta^2_p = .06$, and $\eta^2_p = .14$, respectively (Field 2013; Richardson 2011).

Descriptive statistics

Means and standard deviations are presented in Table 1 (recognition), in Table 2 (prospective metacognition), and in Table 3 (retrospective metacognition).

Recognition

First, we assessed whether accuracy of recognition was above chance (i.e., above 25%) and determined whether there was an age difference in recognition accuracy. Both age groups exhibited accuracy levels significantly greater than chance, 5-year-olds: $t(71) = 2.72$, $p = .008$, 95% confidence interval (CI) = [1.41, 9.14], $r = .31$; 6-year-olds: $t(65) = 7.62$, $p < .001$, 95% CI = [15.32, 26.20], $r = .69$. An independent *t*-test revealed that 6-year-olds outperformed 5-year-olds, $t(119.45) = -4.63$, $p < .001$, 95% CI = [−22.10, −8.86], $r = .39$. Thus, the recognition data reveals that with a recognition accuracy varying between 30 and 45%, there was a good database for assessing children's monitoring of both correct and incorrect recognition.

Metacognitive discrimination (an Indicator of *Relative* metacognitive skills)

Next, we determined whether children were more accurate in their recognition performance for items with a high judgment than for items with a low judgment, and for accepted items than for rejected items, respectively. This analysis allows estimating children's ability to metacognitively discriminate between correct and incorrect recognition in a *relative* sense. Moreover, we investigated age differences between the participating 5- and 6-year-olds for these verbal and non-verbal (relative) indicators of metacognition.

For each measure, we performed a 2 × 2 ANOVA with age as between-subject factor (5 vs. 6 year olds) and item judgment/selection as within-subject factor (for verbal measures we included the factor judgment: low vs. high; for non-verbal measures we included the factor selection: rejected vs. accepted). Figure 2 illustrates recognition accuracy as a function of metacognition (low-judged/rejected items and high-judged/accepted items), age group, for the different measures separately.

Results revealed main effects of age and item judgment/selection for each measure. As expected and as can be seen in Fig. 2, children in both age groups showed the ability to discriminate between correct and incorrect recognition. That is, for each of these relative metacognitive discrimination indicators recognition accuracy was overall significantly higher for high-judged/accepted items compared to low-judged/rejected items (prospective opt-out: $F(1, 120) = 13.96$, $p < .001$, $\eta^2_p = .10$; JoL: $F(1, 108) = 25.86$, $p < .001$, $\eta^2_p = .19$; retrospective opt-out: $F(1, 109) = 39.69$, $p < .001$, $\eta^2_p = .27$; CJ: $F(1, 112) = 38.62$, $p < .001$, $\eta^2_p = .26$).

Table 1 Descriptive statistics of the recognition variables

Variables	5-Year-Olds				6-Year-Olds			
	n	*M*	*SD*	Range	*n*	*M*	*SD*	Range
ACC of Recognition (% correct)	72	30.28	16.44	10–90	66	45.76	22.12	10–100
Choice Latencies for Correct Items (s)	70	7.61	4.24	2.11–25.03	64	6.43	3.21	2.41–19.95
Choice Latencies for Incorrect Items (s)	72	8.13	4.09	2.68–21.92	65	8.08	3.71	2.62–17.95

ACC, Accuracy; there were 10 to-be-remembered item pairs

Table 2 Descriptive statistics of the prospective metacognition variables

	5-Year-Olds				6-Year-Olds			
Variables	*n*	*M*	*SD*	Range	*n*	*M*	*SD*	Range
ACC of Recognition (% correct)								
If JoL is Low	65	27.18	18.85	0–83.33	61	42.13	25.85	0–100
If JoL is High	63	40.50	32.98	0–100	59	63.40	33.02	0–100
If Item is Rejected[a]	66	30.54	26.94	0–100	57	40.32	25.33	0–100
If Item is Accepted[a]	71	34.67	27.25	0–100	66	58.17	32.45	0–100
Metacognitive ACC								
Of JoL	72	59.17	19.84	10–90	66	59.09	20.06	10–90
Of Prosp. Opt-Out	72	48.19	19.88	0–90	66	57.27	21.31	10–90
Percentage of Item Judgment/Selection								
JoL Low	72	64.17	30.06	0–100	66	57.58	28.67	0–100
1JoL High	72	35.83	30.06	0–100	66	42.42	28.67	0–100
Rejected Items[a]	72	44.03	26.14	0–100	66	46.36	25.82	0–90
Accepted Items[a]	72	55.97	26.14	0–100	66	53.64	25.82	10–100

ACC, Accuracy; JoL, judgment of learning

[a] Accpeted/rejected in the prospective (prosp.) opt-out measure

Beside the main effects of age and item judgment/selection (see above), the interactions between age and item judgment/selection were significant for both non-verbal measures, prospective opt-out: $F(1, 120) = 5.84$, $p = .017$, $\eta^2_p = .05$; retrospective opt-out: $F(1, 109) = 5.07$, $p = .026$, $\eta^2_p = .04$. As expected and as can be seen in Fig. 2, the ability to non-verbally discriminate between correct and incorrect recognition was more pronounced in the 6- compared to the 5-year-olds, although the 5-year-olds also showed substantial relative monitoring accuracy. Specifically, when 6-year-olds accepted a trial, they had a substantially higher likelihood that their recognition of that item would be correct compared to when they had rejected an item. And that was true for prospective and retrospective opting out (Prospective

Table 3 Descriptive statistics of the retrospective metacognition variables

	5-Year-Olds				6-Year-Olds			
Variable	*n*	*M*	*SD*	Range	*n*	*M*	*SD*	Range
ACC of Recognition (% correct)								
If CJ is Low	64	22.81	21.24	0–100	54	34.29	31.32	0–100
If CJ is High	69	37.21	28.88	0–100	65	56.82	28.24	0–100
If Item is Rejected[a]	60	25.32	24.37	0–100	58	33.12	30.29	0–100
If Item is Accepted[a]	67	37.91	25.24	0–100	64	60.99	31.70	0–100
Metacognitive ACC								
Of CJ	72	52.92	19.82	10–100	66	57.88	20.94	10–100
Of Retrosp. Opt-Out	72	53.47	20.50	10–90	66	61.82	20.82	10–100
Percentage of Item Judgment/Selection								
CJ Low	72	44.86	30.26	0–100	66	36.67	27.81	0–100
CJ High	72	55.14	30.26	0–100	66	63.33	27.81	0–100
Rejected Items[a]	72	44.03	29.68	0–100	66	40.00	27.85	0–100
Accepted Items[a]	72	55.97	29.68	0–100	66	60.00	27.85	0–100

ACC, Accuracy; CJ, confidence judgment

[a] Accepted/rejected in the retrospective (retrosp.) opt-out measure

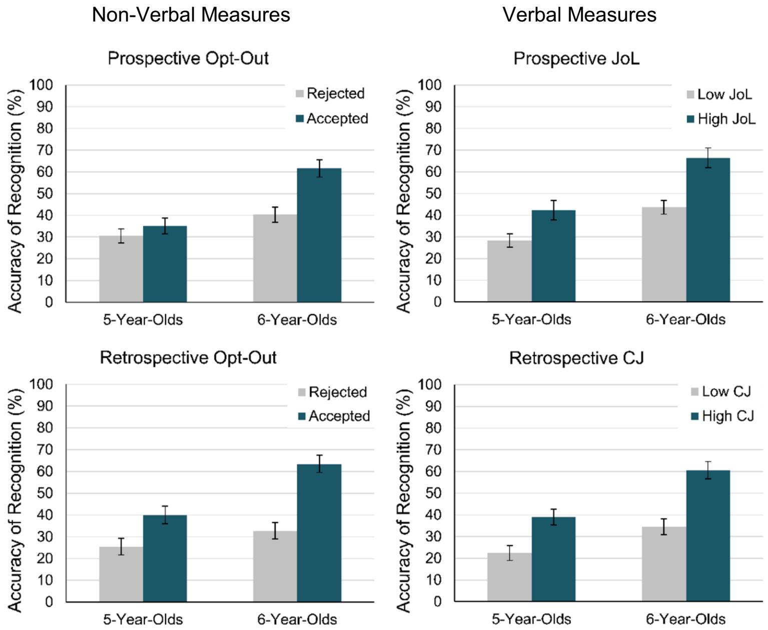

Fig. 2 Estimated marginal means (± SE) for recognition accuracy for low-judged/rejected items and high-judged/accepted items for the two age groups and for each measure. JoL = judgment of learning; CJ = confidence judgment

opt-out: rejected items: $M = 40.32$, $SE = 3.36$, accepted items: $M = 61.56$, $SE = 4.21$, $t(56) = -4.24$, $p < .001$, p-crit = .050, 95% CI = [−31.27, −11.21], $r = .49$; retrospective opt-out: rejected items: $M = 32.70$, $SE = 4.11$, accepted items: $M = 63.45$, $SE = 4.31$, $t(55) = -5.61$, $p < .001$, p-crit = .025, 95% CI = [−41.74, −19.77], $r = .60$).

In contrast to our expectations, 5-year-olds` recognition accuracy was found *not* to differ between rejected and accepted items in the non-verbal prospective measure. (Rejected items: $M = 30.55$, $SE = 3.37$, accepted items: $M = 35.10$, $SE = 3.52$, $t(64) = -0.96$, $p = .341$, p-crit = .050, 95% CI = [−14.03, 4.92], $r = .12$). This suggests that 5-year-olds had difficulties in metacognitively predicting correct and incorrect recognition. When metacognitively evaluating their recognition retrospectively, in contrast, 5-year-olds metacognitively discriminated substantially: their recognition accuracy differed significantly between rejected and accepted items (rejected items: $M = 25.44$, $SE = 3.4$, accepted items: $M = 40$, $SE = 3.57$, $t(54) = -3.14$, $p = .003$, p-crit = .025, 95% CI = [−23.87, −5.26], $r = .39$). These results indicate that 5-year-olds were in principle able to use the opt-out option strategically but had problems in doing so prospectively.

At the same time, also in contrast to our expectations, the interactions between age and item judgment/selection were non-significant in the verbal measures, JoL: $F(1, 108) = 1.47$, $p = .228$, $\eta^2_p = .01$; CJ: $F(1, 112) = 1.94$, $p = .166$, $\eta^2_p = .02$. That is, independent of age and the timing of the metacognitive judgment, lower judgments were associated with a lower likelihood that recognition of that item was correct and higher judgments were associated with a higher likelihood that recognition of that item was correct. Thus, it was not the case that the younger age group had disproportionally larger problems with the verbal compared to the non-verbal monitoring.

Metacognitive Accuracy (an Indicator of *Absolute* Metacognitive Skills)

In addition, we investigated whether verbal and non-verbal measures produced similar results in terms of metacognitive accuracy in an absolute sense, that is, we assessed how many metacognitive responses matched with recognition accuracy. Moreover, we considered age differences between the 5- and 6-year-olds for verbal and non-verbal metacognitive accuracy.

In a first step, we tested whether metacognitive accuracy was above chance level. This proved to be the case for all four measures in the 6-year-olds, JoL: $t(65) = 3.68$, $p < .001$, p-crit = .017, 95% CI = [4.16, 14.02], $r = .42$; prospective opt-out: $t(65) = 2.77$, $p = .007$, p-crit = .050, 95% CI = [2.03, 12.51], $r = .33$; CJ: $t(65) = 3.06$, $p = .003$, p-crit = .025, 95% CI = [2.73, 13.03], $r = .35$; retrospective opt-out: $t(65) = 4.61$, $p < .001$, p-crit = .013, 95% CI = [6.70, 16.94], $r = .50$. In 5-year-olds, in contrast, only the verbal prospective judgments (JoLs) were more accurate than chance level, JoL: $t(71) = 3.92$, $p < .001$, p-crit = .013, 95% CI = [4.50, 13.83], $r = .42$; prospective opt-out: $t(71) = -0.77$, $p = .444$, p-crit = .050, 95% CI = [−6.48, 2.87], $r = .09$; CJ: $t(71) = 1.25$, $p = .216$, p-crit = .025, 95% CI = [−1.74, 7.57], $r = .15$; retrospective opt-out: $t(71) = 1.44$, $p = .155$, p-crit = .017, 95% CI = [−1.34, 8.29], $r = .17$.

Second, we calculated ANOVAs separately for the prospective and the retrospective measures, with age as between-subject factor and modality of metacognitive judgment (prospective measures: JoLs vs. prospective opt-out; retrospective measures: CJs vs. retrospective opt-out) as within-subject factor. Figure 3 depicts the metacognitive accuracy (in percent) for the verbal and non-verbal measures and for both age groups separately.

Results comparing the two *prospective* metacognitive measures (verbal JoLs vs. non-verbal opt-out) revealed a significant main effect of metacognitive modality, $F(1, 136) = 9.14$, $p = .003$, $\eta^2_p = .06$. Contrary to our expectation, no main effect of age was found, $F(1, 136) = 2.72$, $p = .101$, $\eta^2_p = .02$; but the interaction between age and metacognitive modality was significant, $F(1, 136) = 4.68$, $p = .032$, $\eta^2_p = .03$. In contrast to our expectations, the main effect of metacognitive modality was due to higher metacognitive accuracy in the verbal (i.e., JoLs) compared to the non-verbal measure (i.e., opt-out). This was true in both age groups but the effect was unexpectedly even stronger in the 5-year-olds compared to the 6-year-olds (5-year-olds: verbal JoL: $M = 59.17$, $SE = 2.34$; non-verbal prospective opt-out: $M = 48.19$, $SE = 2.34$; $t(71) = 3.49$, $p = .001$, p-crit = .025, 95% CI = [4.70, 17.25], $r = .38$; 6-year-olds: verbal JoL: $M = 59.09$, $SE = 2.47$; non-verbal prospective opt-out: $M = 57.27$, $SE = 2.62$; $t(65) = 0.65$, $p = .516$, p-crit = .050, 95% CI = [−3.74, 7.37], $r = .08$).

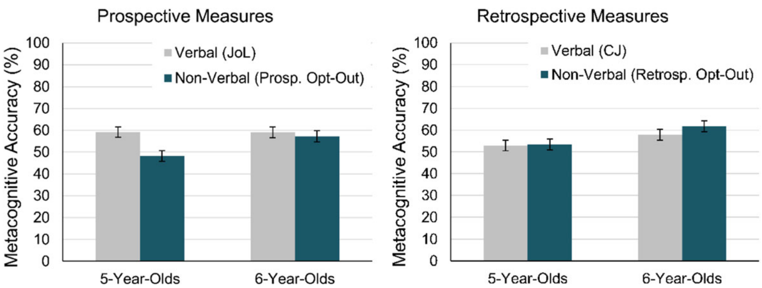

Fig. 3 Estimated marginal means (± SE) for metacognitive accuracy for the verbal and non-verbal measures as a function of age group. JoL = judgment of learning; CJ = confidence judgment; Prosp. = prospective; Retrosp. = retrospective

Results comparing the *retrospective* metacognitive measures (verbal CJ vs. non-verbal opt-out) yielded a significant main effect of age, $F(1, 136) = 4.65$, $p = .033$, $\eta^2_p = .03$. As presumed, the metacognitive accuracy was significantly higher in 6-year-olds, compared to 5-year-olds. The main effect of metacognitive modality, however, as well as the interaction were non-significant ($F(1, 136) = 1.88$, $p = .173$, $\eta^2_p = .01$; interaction age by modality: $F(1, 136) = 1.06$, $p = .304$, $\eta^2_p = .01$). Surprisingly thus, for retrospective metacognition, the modality in which children gave their metacognitive judgment did not affect accuracy of these processes.

Shared processes in verbal and non-verbal metacognitive measures

We explored to what extent verbal and non-verbal metacognition tap similar processes, without having strong assumptions about developmental causes and outcomes. For the purpose of estimating shared processes, we calculated partial correlations among the different measures of metacognitive accuracy, controlling for age. Results are depicted in Table 4 and revealed small to large significant correlations between all measures, except between the metacognitive accuracy of the CJ and the prospective opt-out measure.

Choice latencies in recognition: A window into young Children's implicit metacognition?

First, we explored whether choice latencies in recognition are longer for incorrectly compared to correctly recognized items (see Fig. 4).

Results of the conducted 2 × 2 ANOVA with age as between-subject factor and recognition correctness as within-subject factor revealed a main effect of recognition correctness. As expected, the main effect of correctness was due to substantially longer choice latencies for incorrectly compared to correctly recognized items, $F(1, 131) = 6.48$, $p = .012$, $\eta^2_p = .05$. The main effect age and the interaction were non-significant (age: $F(1, 131) = 1.37$, $p = .245$, $\eta^2_p = .01$; interaction: $F(1, 131) = 1.92$, $p = .168$, $\eta^2_p = .01$). Thus, independent of age, when children selected the correct response alternative, they were significantly faster than when selecting an incorrect alternative.

Secondly, we investigated the relations between choice latencies and the indicators of metacognition, for the two age groups separately. For this, we calculated within-subject correlations on the item level between choice latencies in recognition and the responses in the four metacognitive measures (i.e., JoL and CJ: low- and high-judged items; prospective and retrospective opt-out: rejected and accepted items). Using one-sample *t*-tests, we tested

Table 4 Partial Correlations Between the Metacognitive Accuracy (ACC) measures while controlling for age

Metacognitive ACC (%)	JoL	CJ	Prosp. Opt-Out
Verbal Measures			
JoL	–		
CJ	.32***	–	
Non-Verbal Measures			
Prosp. Opt-Out	.24**	.09	–
Retrosp. Opt-Out	.17*	.55***	.31***

JoL = judgment of learning; CJ = confidence judgment; Prosp, prospective; Retrosp, retrospective

* $p < 0.5$; ** $p < 0.01$; *** $p < 0.001$

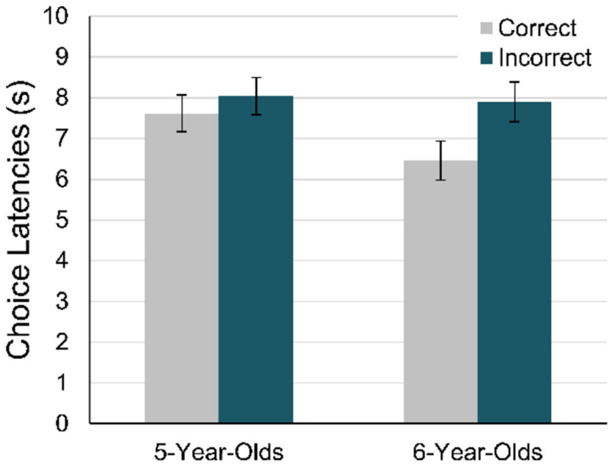

Fig. 4 Estimated marginal means (± SE) for choice latencies in recognition for correct and incorrect recognized items, as a function of age group

whether the mean correlations of both age groups were significantly higher than zero. In addition, with independent *t*-tests we analyzed whether mean correlations differed significantly between the two age groups. Means and standard deviations of the correlations are presented in Table 5, separately for each age group.

Results revealed that in the 6-year-olds, choice latencies in the recognition test correlated significantly and negatively with the responses in all four metacognitive measures. With the 5-year-olds, in contrast, choice latencies correlated only significantly and negatively with the responses of the prospective opt-out (see Table 5 with the results of the one-sample *t*-tests). Readers are reminded that negative correlations indicate – as expected - that *increases* in choice latencies were related to *lower* monitoring judgments and to a *higher* likelihood of rejecting an item.

When comparing the magnitudes of the correlations across age groups, results showed that mean correlations between choice latencies and the two prospective measures were significantly stronger in the 6-year-olds compared to the 5-year-olds, JoL ($t(107) = 3.66$, $p < .001$, p-crit = .013, 95% CI = [0.11, 0.37], $r = .33$; prospective opt-out: $t(119) = 3.01$, $p = .003$, p-crit = .017, 95% CI = [0.07, 0.33], $r = .27$). However, mean correlations between choice latencies and the two retrospective measures did not differ between the two age groups, CJ: $t(108) = 1.65$, $p = .101$, p-crit = .025, 95% CI = [−0.03, 0.28], $r = .16$; retrospective opt-out: $t(105) = -0.36$, $p = .717$, p-crit = .050, 95% CI = [−0.18, 0.12], $r = .04$.

Discussion

Results of verbal and non-verbal indicators of metacognition revealed that – as expected – 5- and 6-year-olds can – in *relative* terms - discriminate between correct and incorrect recognition, relatively independent of modality. That is, both for the verbal judgments and for the opt-out responses, when children chose a low judgment or opted-out, their subsequent responses had a higher likelihood to be incorrect than correct. At the same time, we observed age-related improvements underlining that these skills are rapidly developing in this age range.

Table 5 Descriptive statistics of the correlations on the item level between the choice latencies in the recognition test and the different indicators of metacognition

	5-Year-Olds								6-Year-Olds							
Variables (Correlations)	*n*	*M*	*SD*	*t (df)*	*p*	*p* –crit	95% CI	*r*	*n*	*M*	*SD*	*t (df)*	*p*	*p* –crit	95% CI	*r*
Choice Latencies and JoL	55	−0.07	0.36	−1.41 (54)	.164	.025	[−0.17, 0.03]	.19	54	−0.31	0.32	−7.06 (53)	**.000**	.013	[−0.139, −0.22]	.70
Choice Latencies and Prosp. Opt-Out	64	0.03	0.35	0.64 (63)	.523	.050	[−0.06, 0.12]	.08	57	−0.17	0.37	−3.441 (56)	**001**	.025	[−0.27, −0.07]	.42
Choice Latencies and CJ	59	−0.09	0.41	−1.63 (58)	.108	.017	[−0.20, 0.02]	.21	51	−0.22	0.40	−3.84 (50)	**.000**	.017	[−0.33, −0.10]	.48
Choice Latencies and Retrosp. Opt-Out	55	−0.16	0.39	−3.10 (54)	**.003**	.013	[−0.27, −0.06]	.39	52	−0.14	0.40	−2.43(51)	**.019**	.050	[−0.25, −0.02]	.32

Bold p values = $p <$ Holm critical p -value (p -crit) and thus significant >0

CI, confidence interval; JoL, judgment of learning; Prosp, prospective; CJ, confidence judgment; Retrosp, retrospective

In contrast to our expectations, 5-year-olds performed more poorly in the prospective opting-out than in the corresponding classical monitoring judgment suggesting that the opt-out paradigm is not a "cure-all" method to young children's fragile metacognitive skills. Admittedly, the task was difficult (recognition accuracy was relatively low) and thus, the memory load was high, absorbing younger children's cognitive resources for the learning task and leaving fewer resources available for metacognition (Dunlosky and Metcalfe 2009). The disparate and more positive picture emerging from previous studies using perceptual discrimination tasks (e.g., Lyons and Ghetti 2013) points to the strong impact of the task: the need to build a memory representation and judge or act metacognitively on the basis of that representation is likely to affect relative metacognitive skills. One might argue that our memory task was not a very common task for preschool children and thus that we are underestimating children's "true" metacognitive skills. However, many kindergarten activities involve memory-related games and thus, our paradigm was – although demanding – manageable. It may be considered as a "testing-the-limits" paradigm, certainly also absorbing children's executive functions, but this way providing a realistic estimation of their emerging metacognitive abilities.

The reported results revealed low *absolute* metacognitive accuracy, confirming the literature. In Balcomb and Gerken's study (2008), when 3.5-year-olds chose the opt-out response, they were nevertheless correct in 61% of these declined trials. Similarly, when having the option to retrospectively withdraw answers, Hembacher and Ghetti's (2014) participating 5-year-olds withdrew answers of which about 50% would have been correct. In our study, children's metacognitive selections (opt-out or not) and judgments (high or low certainty) were often erroneous (in 40% to 50% of the cases), and in 5-year-olds not reliably above chance level. Thus, although children as young as 5 years are able to verbally and non-verbally report on uncertainty (in a *relative* sense), the absolute accuracy of these skills remains far from perfect (Schneider and Löffler 2016).

This is the first study to include classical judgments and opt-out measures of metacognition. Direct comparisons of the two modalities revealed unexpected findings. While there were no systematic differences in absolute metacognitive accuracy for the retrospective indicators, the pattern of results for the prospective measures was just in the opposite direction of what we had expected. The reporting of relative metacognitive skills thus seems to overestimate children's skills and the present findings argue for the reporting of both, relative and absolute metacognitive skills in any study.

Although we share many researchers concerns about classical verbally reported metacognitive judgments (Geurten and Bastin 2018; Hembacher and Ghetti 2014; O'Leary and Sloutsky 2017), the present results revealed that the modality in which metacognitive processes are communicated is not the only and most important influencing factor. In fact, the present study revealed that the timing of the metacognitive measure in the task procedure may be more crucial. Across measures, the prospective in direct comparison to retrospective measures of metacognition yielded poorer performance. And, this difference was more pronounced in the 5- compared to the 6-year-olds. It thus seems that foreseeing and predicting performance is more difficult for young children than retrospectively evaluating their performance. The present findings underscore the necessity to precisely conceptualize children's emerging metacognitive skills and to draw a differentiated picture concerning developmental improvements thereof.

This is also the first study to link verbal and non-verbal measures of metacognition in children to each other. Results revealed that – at least in part – these are sharing similar processes. Both before and after the memory test, opting-out and classical verbal judgments were significantly related, indicating that monitoring is a common process. However, the associations explained a maximum of 30% of the observed variance. This may support the theoretical distinction between monitoring and control processes that is dominant in the cognitive and developmental, but not in the comparative literature (Schneider and Löffler 2016; Washburn et al. 2016). While the opting-out can be considered as mirroring preliminary control processes, the judgments belong to the monitoring aspect of metacognition, and this difference is likely to affect the correlations. Motivational (earning credits, avoiding errors), temperamental (conscientiousness), and cognitive (e.g., inhibitory control) factors might additionally explain individual differences in non-verbal and verbal measures differently. Thus, also other, yet-to-be-explored processes seem to be involved when children are given and using the opt-out option.

In a first attempt to uncover implicit metacognitive processes while young children undergo a recognition test revealed that both opt-out responses and verbal judgments are not strongly driven by fast and implicit monitoring processes (but see Hoffmann-Biencourt et al. 2010; Koriat 2012). There were small (6-year-olds) and only negligible intra-individual correlations (5-year-olds) between choice latencies in recognition and measures of metacognition (see Table 5). This might suggest that children were engaged in memory retrieval processes rather than in comparing different degrees of certainty for the alternatives and disregarding experiences of effortful remembering when opting-out and giving metacognitive judgments (Goldsmith and Koriat 2007). The finding that children took more time when selecting the incorrect compared to the correct alternative nevertheless seems to indicate that they were engaging in on-task processing, although probably not in monitoring (see O'Leary and Sloutsky 2017).

The study had not been set up to provide an elaborated theoretical account for the relation between implicit and explicit measures of metacognition but rather to explore the possibility that choice latencies may be an early detectable indicator of on-going metacognitive processes. The relation between these two classes of measures is likely to be complex and developmentally dynamic and may be best described with a dual-processes account (similar to the distinction of implicit and explicit theory of mind skills; see Grosse Wiesmann et al. 2017). The present data shows that preschool children are making metacognitive experiences (Efklides 2011): they may be hesitating before selecting an answer because they are uncertain, or they may jump on an answer right away because they are sure about it. These differences in response latencies can be considered as metacognitive experiences and may be building blocks and important prerequisites for children to develop accurate metacognitive skills (Roebers et al. 2019). To use this for informing monitoring and for improving monitoring accuracy has yet to be discovered or learned.

No study is perfect and this one is no exception: For one, the within-subject design implicated that the different metacognitive indicator were assessed in a sequence. This bears the disadvantage that an earlier assessment may influence a subsequent. We assumed that some monitoring processes need to take place before children can decide whether to opt-out (reject) or accept a trial, and thus decided that the classical judgments should rather proceed the opting-out phases instead of vice versa. Unfortunately, we cannot rule out the possibility that the results would appear different if the order had been reversed. Yet, as we did not find consistent and strong advantages of the opting-out in comparison with the classical judgments (in the sense that monitoring was triggered and thereby substantially improved opting-out), it seems safe to assume that the influence of the fixed order was only marginal. For another, the limited number of items prevented the calculation of within-person correlations (Gammas) as often-used measures of metacognitive accuracy. The task proved to be difficult (recognition accuracy was low to moderate) and thus, adding more items would not have been beneficial. This limitation is not unique to the present study but rather the rule in developmental research. Fortunately, other measures, such as discrimination or resolution and the distinctions used in the present approach have been proven to be similarly informative. Finally, and strictly spoken, the dichotomization of the different indicators into verbal and non-verbal may not be aptly named, considering that the opt-out tasks were also instructed verbally. The labels were chosen to refer to the literature these paradigms stem from.

Taken together, the present study contributes to an ongoing debate about how to best capture young children's emerging metacognitive skills by showing that non-verbal measures tap similar, although not the identical processes as classical verbal metacognitive judgments. Time-based implicit measures, in contrast, seemed to be rather distinct from metacognitive indicators. More research is clearly needed to better understand the processes involved when children undergo memory tests and evaluate their performance and to explore the dynamic nature of the relationships between the different indicators. Understanding the early roots of these important metacognitive skills should be a central objective for cognitive, developmental, and comparative psychology alike.

Acknowledgments We wish to thank Nike Tsalas for her help designing the paired associates task and setting up the study. We further gratefully acknowledge the participation of our preschool and kindergarten children, as well as the cooperation of their teachers and parents. The help of Stefanie Mahrer, Florence Ibrahim, and Massimo Gut during data collection was also greatly appreciated. Our gratitude includes Stefan Kodzhabashev from the Faculty's technology platform for programming the task.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Research involving human participants The authors assure that the study has been realized in accordance with the ethical principle of the APA. Ethical approval from the Faculty's Ethical Review Board had been obtained prior to the study.

Informed consent Children's parents gave informed consent; children themselves gave oral consent prior to the testing.

References

Balcomb, F. K., & Gerken, L. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Developmental Science, 11*, 750–760. https://doi.org/10.1111/j.1467-7687.2008.00725.x.

Brewer, N., & Day, K. (2005). The confidence-accuracy and decision latency-accuracy relationships in children's eyewitness identification. *Psychiatry, Psychology and Law, 12*, 119–128. https://doi.org/10.1375/pplt.2005.12.1.119.

Coughlin, C., Hembacher, E., Lyons, K. E., & Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Developmental Science, 8*, 957–971. https://doi.org/10.1111/desc.12271.

De Bruin, A. B. H., & van Merriënboer, J. J. G. (2017). Bridging cognitive load and self-regulated learning research: A complementary approach to contemporary issues in educational research. *Learning and Instruction, 51*, 1–9. https://doi.org/10.1016/j.learninstruc.2017.06.001.

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. https://doi.org/10.1016/j.jecp.2014.04.001.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks: Sage.
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. https://doi.org/10.1016/j.learninstruc.2012.05.002.
Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46*, 6–25. https://doi.org/10.1080/00461520.2011.538645.
Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146.
Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (3rd ed.). London: Sage Publications.
Geurten, M., & Bastin, C. (2018). Behaviors speak louder than explicit reports: Implicit metacognition in 2.5-year-old children. *Developmental Science, 22*, 1–6. https://doi.org/10.1111/desc.12742.
Ghetti, S., Hembacher, E., & Coughlin, C. A. (2013). Feeling uncertain and acting on it during the preschool years: A metacognitive approach. *Child Development Perspectives, 7*, 160–165. https://doi.org/10.1111/Cdep.12035.
Goldsmith, M., & Koriat, A. (2007). The strategic regulation of memory accuracy and informativeness. In A. S. Benjamin & B. H. Ross (Eds.), *Skill and strategy in memory use* (pp. 1–60). San Diego: Elsevier Academic Press.
Goupil, L., & Kouider, S. (2016). Behavioral and neural indices of metacognitive sensitivity in preverbal infants. *Current Biology, 26*, 3038–3045. https://doi.org/10.1016/j.cub.2016.09.004.
Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeiss, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental Science, 20*, e12445.
Hembacher, E., & Ghetti, S. (2014). Don't look at my answer: Subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. *Psychological Science, 25*, 1768–1776. https://doi.org/10.1177/0956797614542273.
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. https://doi.org/10.1348/026151009x479042.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65–70 Retrieved from https://www.jstor.org/stable/4615733.
Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370. https://doi.org/10.1037/0096-3445.126.4.349.
Koriat, A. (2012). The relationships between monitoring, regulation and performance. *Learning and Instruction, 22*, 296–298. https://doi.org/10.1016/j.learninstruc.2012.01.002.
Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science, 13*, 441–453. https://doi.org/10.1111/j.1467-7687.2009.00907.x.
Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy. Insights from the processes underlying judgments of learning in children. In P. Chambres, M. Izaute, & P.-J. Marescaux (Eds.), *Metacognition, process, function and use* (pp. 1–18). Norwell: Kluwer Academic Publishers.
Kuhn, D. (1999). Metacognitive development. In: L. Balter & C. T. Lamonda: *Child psychology: A handbook of contemporary issues (p.* 259–286). Philadelphia, PA, Psychology Press.
Kulke, L., & Rakoczy, H. (2018). Implicit theory of mind- an overview of current replications and non-replications. *Data in Brief, 16*, 101–104.
Lockl, K., & Schneider, W. (2004). The effects of incentives and instructions on children's allocation of study time. *European Journal of Developmental Psychology, 1*, 153–169. https://doi.org/10.1080/17405620444000085.
Lyons, K. E., & Ghetti, S. (2013). I don't want to pick! Introspection on uncertainty supports early strategic behavior. *Child Development, 84*, 726–736. https://doi.org/10.1111/cdev.12004.
Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*, 314–324. https://doi.org/10.3758/s13428-011-0168-7.
Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge: The MIT Press.
O'Leary, A. P., & Sloutsky, V. M. (2017). Carving metacognition at its joints: Protracted development of component processes. *Child Development, 88*, 1015–1032. https://doi.org/10.1111/cdev.12644.
Paulus, M., Proust, J., & Sodian, B. (2013). Examining implicit metacognition in 3.5-year-old children: An eye-tracking and pupillometric study. *Frontiers in Psychology, 4*, 145. https://doi.org/10.3389/fpsyg.2013.00145.
Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6*, 135–147. https://doi.org/10.1016/j.edurev.2010.12.001.

Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology, 82*, 416–425.

Roebers C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. Developmental Review 45:31-51

Roebers, C. M., Mayer, B., Steiner, M., Bayard, N., & Van Loon, M. (2019). The role of children's metacognitive experiences for cue utilization and monitoring accuracy: A longitudinal study. *Developmental Psychology.* Advance online publication. https://doi.org/10.1037/dev0000776.

Ruffman, T., Garnham, W., Import, A., & Conolly, C. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology, 80*, 201–224.

Schneider, W., & Löffler, E. (2016). The development of metacognitive knowledge in children and adolescents. In J. Dunlosky & S. U. K. Tauber (Eds.), *The Oxford handbook of Metamemory* (pp. 491–518). New York: Oxford University Press.

Shields, W. E., Smith, J. D., & Washburn, D. A. (1997). Uncertain responses by humans and rhesus monkeys (Macaca mulatta) in a psychophysical same-different task. *Journal of Experimental Psychology: General, 126*, 147–164. https://doi.org/10.1037/0096-3445.126.2.147.

Smith, D. J. (2005). Studies of uncertainty monitoring and metacognition in animals and humans. In H. S. Terrace & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (pp. 242–271). New York: Oxford University Press.

Smith, J. D., Shields, W. E., & Washburn, D. A. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and Brain Sciences, 26*, 317–339. https://doi.org/10.1017/S0140525X03000086.

Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from implicit to an explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology, 30*, 172–187.

Tsalas, N., Sodian, B., & Paulus, M. (2017). Correlates of metacognitive control in 10-year old children and adults. *Metacognition and Learning, 12*, 297–314. https://doi.org/10.1007/s11409-016-9168-4.

Washburn, D. A., Beran, M. J., & Smith, D. J. (2016). Metamemory in comparative context. In J. Dunlosky & S. U. K. Tauber (Eds.), *The Oxford handbook of Metamemory* (pp. 269–288). New York: Oxford University Press.

Zakay, D., & Tuvia, R. (1998). Choice latency times as determinants of post-decisional confidence. *Acta Psychologica, 98*, 103–115.