frontiers | Frontiers in Developmental Psychology

TYPE Original Research
PUBLISHED 06 November 2024
DOI 10.3389/fdpys.2024.1454717

OPEN ACCESS

EDITED BY
Loren Marulis,
Connecticut College, United States

REVIEWED BY
Shiyi Chen,
University of Idaho, United States
Jérôme Clerc,
Université Grenoble Alpes, France
Marion Leclercq,
University of Lille, France

*CORRESPONDENCE
Florian Jonas Buehler
✉ florianjbuehler@gmail.com

RECEIVED 25 June 2024
ACCEPTED 15 October 2024
PUBLISHED 06 November 2024

CITATION
Buehler FJ and Oeri N (2024) Sneaky Snake: assessing metacognitive behavior in 5 to 6 year-olds with an unsolvable task.
Front. Dev. Psychol. 2:1454717.
doi: 10.3389/fdpys.2024.1454717

Sneaky Snake: assessing metacognitive behavior in 5 to 6 year-olds with an unsolvable task

Florian Jonas Buehler* and Niamh Oeri

Department of Developmental Psychology, University of Bern, Bern, Switzerland

In the present study, we developed an unsolvable behavioral metacognitive task for kindergarten children. The task was designed to gain insight into how children's metacognitive processes, measured as monitoring (e.g., checking the plan) and control behavior (e.g., seeking a piece), operate in a problem-solving task that mimics real-life scenarios. Five to six-year-old kindergarten children ($N = 72$) were asked to build a wooden snake according to a plan. The middle piece of the snake (fourth out of seven pieces) was missing, making the task unsolvable. Other than expected, metacognitive behavior was not related to teacher ratings of metacognitive self-regulation. However, we found age differences. Children in kindergarten year two ($M = 5.85$ years old) showed more control behavior than children in kindergarten year one ($M = 5.05$ years old). Surprisingly, we did not find age differences in monitoring behavior. Lastly, we found that metacognitive behavior differed between the solvable part (before the missing piece is reached) and the unsolvable part (after the missing piece is reached). Children showed more monitoring and less control behavior in the solvable part than in the unsolvable part. The current study contributes to the metacognitive research methodology by capturing children's metacognitive processes in action using an ecological-valid, unsolvable behavioral task.

KEYWORDS

metacognition, monitoring, control, kindergarten, metacognitive behavior, unsolvable task

1 Introduction

Metacognitive processes are typically assessed verbally. Metacognitive processes describe the ability to monitor and control ongoing cognitive processes (Nelson and Narens, 1990). In metacognitive tasks that capture monitoring and control, children are often asked to evaluate their study progress (i.e., judgments of learning) to rate their confidence in a given answer (i.e., confidence judgments) or to decide which materials they would like to study again (i.e., restudy selections) (e.g., Baer et al., 2021; Bayard et al., 2021; Destan et al., 2014). For example, children are asked, "How sure are you that your answer is correct?" However, verbal metacognitive assessment requires language skills and conscious metacognitive awareness that may not be sufficiently developed, especially in young children. Thus, using verbal assessments to estimate metacognition may be misleading as children's metacognition in everyday situations differs from such verbal judgments. Behavioral metacognitive tasks have been proposed to address this shortcoming. In behavioral metacognitive tasks, children assemble train tracks (Bryce and Whitebread, 2012) or puzzles (Marulis and Nelson, 2021) according to a plan. Such problem-solving tasks allow observing spontaneously occurring metacognitive behavior, such as checking the plan or correcting a mistake. These tasks may be closer related to real-life scenarios because they are almost identical to typical children's games (e.g., assembling train tracks or puzzles). Although these behavioral tasks provide an opportunity to observe metacognitive behavior, our understanding of non-verbal metacognitive behaviors and developmental variation is currently limited due to different task constraints. The current study aims to address these shortcomings by introducing an unsolvable behavioral metacognition task. By providing a task with high ecological validity and systematically analyzing different monitoring and control behaviors, the study will contribute to a fine-grained understanding of metacognitive processes in 5–6-year-old kindergarten children.

Most research on metacognitive processes is based on the influential framework by Nelson and Narens (1990). Developed initially for metamemory, the framework distinguishes between metacognitive monitoring and control, which can be applied to other domains such as problem-solving (e.g., Bryce and Whitebread, 2012; Marulis and Nelson, 2021). Monitoring is a bottom-up process to accumulate task information (e.g., evaluating task difficulty). Control is a top-down process initiating actions at the task level. For example, evaluating a task as highly challenging is a monitoring process. Consequently, seeking help based on the evaluation is a typical control process. Both processes are closely related and crucial for children's self-regulated learning and academic achievement (Dunlosky and Metcalfe, 2009; Roebers, 2017).

Metacognitive monitoring and control develop from an early age. Research using perceptual tasks shows that from the age of 3, children can monitor their performance by reporting higher confidence in correct than incorrect trials (e.g., identifying degraded pictures; Coughlin et al., 2015; Gonzales et al., 2021; Lyons and Ghetti, 2011). From age 4, children seem to be able to monitor their performance in memory tasks (e.g., remembering picture pairs; Destan et al., 2014; Hembacher and Ghetti, 2014). From the age of 5, children show signs of metacognitive control as they are more likely to withdraw an incorrect answer than a correct answer (Bayard et al., 2021; Destan et al., 2014; Destan and Roebers, 2015; Kim et al., 2021). However, despite these impressive findings, it is important to acknowledge that these tasks require well-developed language skills and conscious metacognitive awareness. As mentioned above, these skills may not be fully developed in kindergarten children, yielding biased results for children with low language skills and/or lower metacognitive awareness. Non-verbal metacognitive tasks, therefore, provide an opportunity to analyze metacognitive processes independent of a child's language skills. For instance, behavioral tasks allow us to observe children's spontaneously occurring metacognitive processes without explicitly asking about them (e.g., "How sure are you that your answer is correct?"). In the following, we will refer to behavioral observations of metacognitive processes (e.g., monitoring and control) as metacognitive behavior. Thereby, it is important to note that previous studies (e.g., Bryce and Whitebread, 2012; Marulis and Nelson, 2021) have used the term metacognitive skills to describe metacognitive behavior. In the present study, we used the term metacognitive behavior to emphasize the behavioral and non-verbal aspects of behavioral assessments of metacognition. Studies focusing on metacognitive behavior (Bryce and Whitebread, 2012; Marulis and Nelson, 2021) are scarce but reveal similar developmental patterns: From the age of 3, children show monitoring (e.g., checking the construction) and control behavior (e.g., clearing space) when building three-dimensional puzzles according to a plan (Marulis and Nelson, 2021).

By simulating real-life scenarios in metacognitive tasks, we can gain insight into how metacognitive processes operate in everyday situations. Bryce and Whitebread (2012) introduced a problem-solving task in which children (5–7 years) were asked to assemble train tracks according to a model. The task allows one to observe metacognitive monitoring (e.g., checking the construction, checking the model) and control processes (e.g., clearing space, stating a plan). The results showed quantitative and qualitative differences in monitoring and control behaviors between 5- and 7-year-olds. Furthermore, metacognitive behavior was related to teacher ratings of children's metacognition [CHILD questionnaire by Whitebread et al. (2009)], suggesting convergent validity for the developed problem-solving task. Results confirmed the age-sensitivity of the task. Age differences indicated reliable age discrimination for both metacognitive processes, monitoring (e.g., checking the model) and control (e.g., sorting materials). Similarly, in the Wedgits© task (Marulis and Nelson, 2021), 3 to 5-year-olds had to assemble three-dimensional puzzles according to a plan. Metacognitive behavior was coded similarly to the train track task by Bryce and Whitebread (2012). However, in their analyses, the authors focused on aggregated scores of monitoring and control and did not distinguish between different types of monitoring or control behaviors. Results showed that metacognitive monitoring and control can be reliably observed at the age of 3. Overall, both studies suggest that metacognitive behavior in real-life play situations can be reliably observed at a very early age.

In addition to the benefits of observing metacognitive processes in real-life scenarios, behavioral metacognitive tasks have two further advantages. First, observing metacognition in behavioral tasks allows us to capture metacognitive behavior not only quantitatively but also qualitatively. Most standardized tasks provide quantitative, aggregated mean-based estimates for metacognitive processes. For example, typical memory tasks (Destan et al., 2014) and picture identification tasks (Lyons and Ghetti, 2011) yield aggregated (mean-) scores for metacognitive monitoring or control. While these tasks have provided insights into children's metacognitive development (see for an overview Roebers, 2017), the mean-based approach fails to capture different types of metacognitive monitoring and control processes involved in a task. Behavioral tasks, however, allow us to capture both quantifiable indexes and the opportunity to analyze the quality of the behavior. Thus, assessing metacognition through behavioral tasks not only provides insight into how often a behavioral strategy is displayed but also provides a more detailed understanding of the type of metacognitive behavior children display when faced with a challenge.

Second, behavioral tasks allow us to observe successful metacognitive performance as well as unsuccessful metacognitive performance, also known as metacognitive failure (e.g., Bryce and Whitebread, 2012; Marulis and Nelson, 2021). In an unsolvable behavioral task, two types of metacognitive failure can be observed: Failure of metacognitive monitoring and failure of metacognitive control. In our approach, monitoring failure occurs when participants mistakenly assemble the wrong piece without realizing the error. An incorrectly assembled piece suggests a failure in monitoring, such as failing to gather correct information about the piece. Furthermore, we conceptualize metacognitive control failure as any form of off-task behavior. Off-task is defined as any behavior that does not serve task completion constructively (Oeri and

Roebers, 2021). When showing off-task behavior, children fail to maintain goal-directed control behavior, such as seeking a piece. Especially when a child is asked to work independently on a task without any adult scaffolding, metacognitive control failure in terms of off-task behavior is likely to occur. Observing metacognitive failures, such as making mistakes and off-task behavior, provides insights into different aspects of the task that might be particularly challenging for children. Thus, observing successful metacognition (i.e., monitoring and control) and metacognitive failure (i.e., making mistakes and off-task) within the same task provides a comprehensive understanding of how metacognitive processes play out and which aspects might be especially challenging to exert metacognition successfully.

Despite these exciting advantages of behavioral tasks, methodological challenges currently limit our understanding of metacognitive behavior in more detail. A common challenge in any behavioral task is the intertwined effects of ability, age, and previous experience. More precisely, task difficulty can impact participants' performance, potentially leading to biased results if it varies significantly between participants. Thus, keeping task difficulty constant across participants is essential to capture the skills of interest reliably (e.g., Dunlosky et al., 2016). Bryce and Whitebread (2012) addressed the issue by introducing two different age-dependent train track tasks, an easy and a more difficult one. Even though performance between age groups was matched for task difficulty, such an approach does not control for ability differences within the age groups. Depending on previous experience with train tracks, task difficulties could still vary largely within the respective age groups. In the Wedgit task, difficulty was held constant by giving children increasingly complex puzzles until they could not complete them within 4 min (Marulis and Nelson, 2021). While such an approach ensures that metacognitive performance is assessed at the individual threshold of maximal performance, it may impact motivation and tiredness, as some children need to complete many more trials than others to achieve their maximum. Another less time-consuming approach is to make the task unsolvable. Despite the fact that previous experiences may influence motivation and potential strategies for approaching the task, the unsolvable nature of the task keeps task difficulty constant across the participants without requiring them to complete numerous trials below their performance threshold.

Second, to analyze different metacognitive strategies the monitoring and control strategies must be observed at a minimal frequency. The train track and the Wedgits task report an average of 8–11 monitoring and control behaviors per minute. However, for the train track task, the most frequent monitoring behavior (i.e., "checking own construction") was observed on average twice per minute, and the most frequent control behavior (i.e., "clearing space") was observed 0.5 times per minute. Furthermore, behaviors shown by less than 25% of the children were excluded from the micro-level analysis behaviors due to the limited range of scores. The low frequencies of target behaviors restrict the reliability of metacognitive behavior estimates, making it challenging to capture potential developmental shifts. A possibility to address this issue would be by introducing more diverse features of the target and distractor items. More specifically, using items that vary in color and symbols forces the participants to monitor and control their behavior more diligently, yielding more possibilities to observe monitoring and control behavior.

Lastly, previous behavioral tasks have focused solely on the metacognitive behavior's frequency. Although this provides important information on how often monitoring and control behavior can be observed, it does not give any information on how long participants engage in the respective behavior. Especially when trying to solve a problem, persisting with a behavior increases the chance for the behavior to be successful. For example, searching for the train track takes a minimal amount of time. If child A searches for a train track for 1 s and child B searches for 10 s, the likelihood of success is higher in child B, but the frequency score would be identical in both children. Although duration is by no means a guarantee for success, it does enhance the chance of being successful. Thus, including the duration of behavior in the coding and the behavior analysis may be a potential route to understanding behavioral patterns in more detail. Combining the frequency and duration may provide insight into different effective and non-effective patterns of monitoring and control when solving a problem.

Building on the foundation of Bryce and Whitebread's (2012) and Marulis and Nelson's (2021) behavioral tasks, we developed an unsolvable behavioral metacognition task (The Sneaky Snake) to observe monitoring and control behavior in kindergarten children. Similar to Bryce and Whitebread (2012), children (4–5 years) had to assemble a snake using wooden pieces according to a model. Different from Bryce and Whitebread (2012), the wooden pieces were colored (green, blue, yellow) and had different symbols on them (dots, triangles, squares). The fact that the snake pieces varied in color, size, shape, and symbols increased the need for a thorough inspection, verification, and reassessment. By increasing the complexity of the target and distractor features, we aimed to observe more metacognitive behavior per minute than in previous tasks and, potentially, more fine-grained developmental differences in the variety of observed metacognitive behaviors. The task was designed to be unsolvable. More specifically, the middle piece of the snake was missing. There were enough distractor pieces to ensure participants did not realize the piece was missing. Additionally, we verified that the children did not know that the piece was missing: We coded whether children had systematically tried all distractor pieces. If this was the case, we excluded the child ($n = 1$). Only by trying every distractor piece could one infer that the fourth piece was actually missing. Through the task's unsolvable nature, we aimed to keep task difficulty constant across participants. It prevents ceiling effects as no participant can fully complete the task (e.g., Dunlosky et al., 2016).

When developing a new task, it is also important to include established measures to examine the validity of the task. We evaluated the task's convergent validity by comparing the observed metacognitive behavior (monitoring and control) with the BRIEF-P Plan/Organize scale (German version: Daseking and Petermann, 2013). The BRIEF-P is a teacher questionnaire on children's self-regulation problems in the classroom. The scale includes items on metacognitive processes and executive functions. Recent studies suggest metacognition is closely related to self-regulation skills, such as executive functions (Bryce et al., 2015; Roebers, 2017; Marulis et al., 2020; Marulis and Nelson, 2021). For instance, Marulis and Nelson (2021) found a relationship between metacognitive behavior and executive functions. Therefore, the BRIEF-P is a suitable tool for assessing convergent validity with more general self-regulation skills in the classroom context. We also explored the relationship between the observed metacognitive behavior and a verbal assessment of children's metacognition in a ball-throwing task (Schneider, 1998). In the ball-throwing task, children had to estimate how many out of 10 balls they could successfully throw into a basket. This allows us to estimate convergent validity with a classical verbal assessment of metacognitive processes (i.e., What do you think how many of these 10 balls will you hit into the basket?).

The hypotheses were the following: (1) We expect the observed monitoring and control behaviors in 5–6-year-old kindergarten children to be negatively related to teacher ratings of self-regulation problems (BRIEF-P Plan/organize scale). More precisely, we expect more frequent and longer monitoring and control behavior to be related to fewer problems reported in the Plan and Organize scale. (2) We expect more frequent monitoring and control behavior in older than younger children. (3) We expect that older children spend more time (longer durations) with monitoring and control behaviors than younger children. (4) Without any *a priori* expectation of the change in metacognitive behavior, we exploratory compared metacognitive behavior before reaching the missing piece (solvable interval) with metacognitive behavior after the missing piece (unsolvable interval).

2 Method

2.1 Participants

In the preregistered study,[1] we relied on a random subsample from a larger study (https://doi.org/10.17605/OSF.IO/JYCV7) on children's self-regulated learning ($N = 193$). The target sample size for the present study was $N = 66$ children and is based on previous studies (Bryce and Whitebread, 2012; Marulis and Nelson, 2021). We included six additional children to account for potential dropouts, resulting in $N = 72$ children (47% female). Children were recruited from different public kindergartens. Seventy-four percent of the children in the sample had at least one parent with a university degree, indicating a high socioeconomic background. Moreover, 67% of the sample were native speakers, which reflects the number of native children in Swiss schools (Federal Statistical Office, 2024). A majority of children had Swiss parents (59%), 13% had parents from other European countries, 4% had African parents, 4% had Asian parents, and for 19%, we do not have any information on the ethnic background. Children in the first kindergarten year ($n = 37$) were $M = 5.05$ ($SD = 0.33$) years old, and children in the second kindergarten ($n = 35$) were $M = 5.85$ ($SD = 0.46$) years old.

2.2 Procedure

Before testing, parents gave informed written consent, and children gave verbal assent. Ethical approval for the study was obtained from the faculty ethics committee (Approval No. 2023-07-01). We collected data from September 2023 until December 2023. Six trained experimenters individually tested participants. The child's parents were not present during testing. Testing took place in a quiet room at the kindergarten. Among all administered tasks, task order was counterbalanced.

2.3 Sneaky Snake

The task measures metacognitive behavior in an unsolvable task. The task was adapted from Bryce and Whitebread (2012). The participants had to assemble a colored wooden snake according to a model. The model (picture) and a box with the target and distractor pieces were placed on a mat (Figure 1). The snake (test trial version) consisted of seven target pieces. Overall, 38 additional distractor pieces were placed in a box. To increase task difficulty and elicit different task behaviors, several distractors were used: The snake pieces differed in colors (green, blue, yellow), shape (short and long bent pieces, four different sizes of straight pieces), and symbols (dots, triangles, squares). First, the participants completed a practice trial (a snake with three pieces) to ensure they understood the task. During the practice trial, participants received feedback from the experimenter. After a successful practice trial, children were asked to assemble another snake (test trial). Children were instructed to start building the snake from the head. They were also instructed to build the snake on the mat. The test trial was unsolvable. The picture model of the test trial consisted of seven pieces, but the fourth piece of the snake was missing. For the test trial, the experimenter left the room and returned after 5 min or whenever the child ended the task. Children's behavior was videorecorded. We piloted the task several times to determine the optimal number and qualities (colors, shapes, sizes, and symbols) of target and distractor pieces. Intensive piloting was necessary to ensure that the task was suitable for 5–6-year-olds yet challenging enough to observe metacognitive behavior. We excluded 10 children from the analyses because they were not able to successfully complete the practice trial, decided to interrupt the task (i.e., going to the bathroom), were interrupted by a third person, or understood that the task was unsolvable (children who had systematically tried all the distractor pieces, $n = 1$). Excluding children who understood the task is unsolvable is relevant to keeping the task demands constant across participants.

2.4 Coding scheme for metacognitive behavior

The foundation of the developed coding scheme is based on the coding scheme for metacognitive behavior by Bryce and Whitebread (2012). The coding scheme developed by Bryce and Whitebread has four distinct features that suggest it is a valuable tool for observing metacognitive behavior: (1) It is based on Nelson and Narens' (1990) theoretical framework. (2) It is age sensitive. (3) It shows convergent validity with teacher ratings of children's metacognition. (4) It has been successfully applied to a slightly different problem-solving task by Marulis and Nelson (2021). Based on the metacognitive behaviors described by Bryce and Whitebread (2012) and Nelson and Narens' theoretical framework (1990), we adapted the coding scheme for the Sneaky Snake task. We coded monitoring, control, making mistakes, and off-task behaviors. To differentiate between monitoring and control behaviors, we relied on Nelson and Narens' framework (1990), describing monitoring as a bottom-up process accumulating task

1 https://doi.org/10.17605/OSF.IO/ZX86H

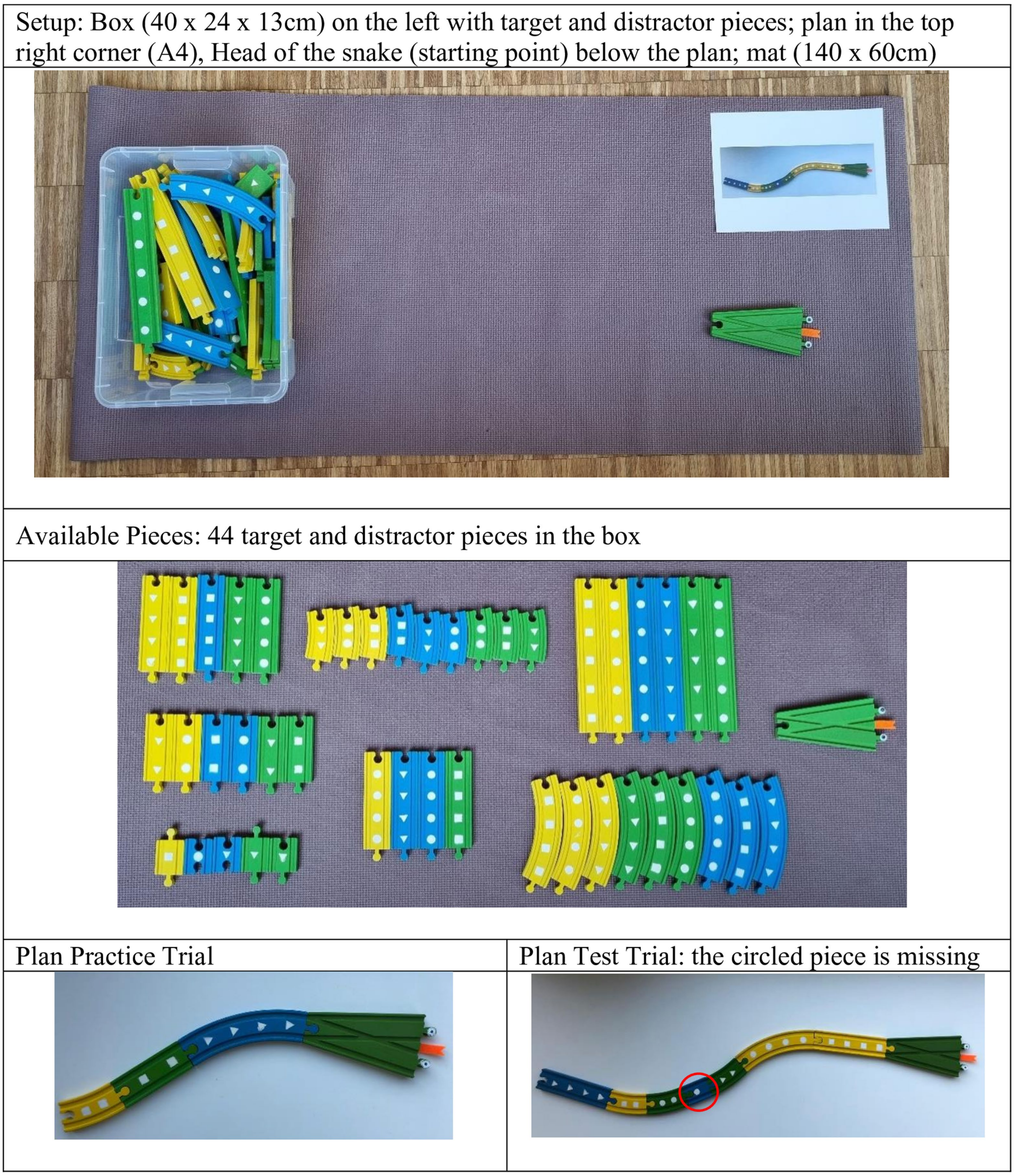

FIGURE 1
The Sneaky Snake task.

information (e.g., studying the plan) and control as a top-down process initiating actions at the task level (e.g., seeking a piece in the box). Making mistakes and off-task behavior describe two types of metacognitive failure. Making mistakes describes a monitoring failure (e.g., building in an incorrect piece) and occurs when a person may not have accumulated enough task information to make an accurate decision. Off-task behavior describes a control failure (e.g., walking away from the task) and occurs when a person fails to maintain goal-directed behavior at the task level. To examine the validity of the coding scheme, i.e., if these behaviors were observable, we randomly selected and coded 10 videos. After these 10 initial codings, we had to revise the coding scheme because some behaviors did not occur as we theoretically assumed. For example, because of the low occurrences of verbalizations in the Sneaky Snake we reduced the number of verbalization categories to monitoring and control. The final coding scheme is displayed in Table 1.

After the first full coding round, we had to drop four behaviors ("Comparing a single piece with the plan," "Comparing pieces,"

TABLE 1 Sneaky Snake coding scheme.

Behavior	Example	Occurrences per minute *M* (*SD*)	Seconds per minute *M* (*SD*)
Monitoring behavior			
Checking plan	Child glances back to the plan while seeking in the box for a piece.	4.2 (1.47)	11.25 (3.51)
Inspecting a piece	Child takes a closer look at a snake piece by counting the number of symbols on the piece.	2.69 (1.25)	4.76 (2.59)
Monitoring verbalization	"This is a difficult task" or "1, 2, 3, 4 [child counts squares on a piece]"	0.18 (0.54)	0.65 (2.16)
Comparing a single piece with the plan	Child puts a snake piece next to the plan and glances back and forth between the piece and the plan.	–	–
Comparing pieces	Child compares a blue curve with dots with a blue curve with squares.	–	–
Checking own construction	Child checks their construction by overviewing the built snake.	–	–
Sum score monitoring	–	7.08 (2.6)	16.67 (6.17)
Control behavior			
Seeking	Child seeks in the box for a piece.	4.44 (1.21)	24.78 (7.51)
Adjustments	Child replaces a piece in the snake to correct an error.	0.13 (0.17)	0.66 (1.21)
Control verbalization	"A yellow curve with four squares [child repeats what they are seeking]"	0.05 (0.32)	0.22 (1.37)
Grouping pieces	The child groups yellow pieces in one place.	–	–
Empty the box	Child empties the box.	–	–
Sum score control	–	4.63 (1.29)	25.66 (7.25)
Monitoring failure			
Mistakes	Child builds in an incorrect piece.	0.58 (0.73)	5.15 (5.6)
Control failure			
On-task-off-task	Child builds a snake that is unrelated to the task and the plan.	0.75 (0.75)	3.8 (4.12)
Off-task	Child walks away and does not interact with the task anymore.	0.3 (0.26)	4.26 (4.65)
Sum score off-task	–	1.06 (0.8)	8.05 (6.63)

All behaviors were coded in the first round of coding. In the second coding round, behaviors in bold were maintained, whereas all other behaviors were dropped because of low frequencies and reliabilities.

"Checking own construction," and "Grouping pieces") because they were not uniquely identifiable and reliably distinguishable from other behaviors. It was challenging to distinguish between "comparing a piece with the plan," "checking the original plan," and "inspecting an object." For example, a child looks back and forth between a piece and the plan while holding the piece in their hand. This behavior could be coded as a single instance of "comparing the piece with the plan" or as two instances of "inspecting the piece" and "checking the plan." To decrease ambiguity in the coding process, we decided to focus on fewer behaviors but clearly identifiable and reliably observable behaviors. Finally, we excluded "emptying the box" because it was not observed. Table 1 shows an overview of the coded behaviors, including their occurrences and durations. A more detailed version of the coding scheme is available here: https://osf.io/3dtnv/.

Because the total test time varied between subjects, we divided the duration and occurrences per behavior through the total minutes spent on the task. If children were briefly disturbed by a third party (e.g., another child running in the test room), the disturbance time was deducted from the total time. (No behaviors were rated when the child was disturbed; $n = 8$; disturbance time: $M = 3.39$ s; $range = 1–6.13$ s). We also computed aggregated scores of monitoring, control, and off-task behavior. We summed all individual behaviors contributing to monitoring, control, or off-task behavior. Monitoring failures consisted of a single measurement based on mistakes. See Table 1 for the mean scores.

We double-coded 28 (39%) of the videos. Interrater reliability for monitoring (*ICC* occurrence/min = 0.85; *ICC* duration/min = 0.59), control (*ICC* occurrence/min = 0.94; *ICC* duration/min = 0.95), making mistakes (*ICC* occurrence/min = 0.93; *ICC* duration/min = 0.96), and off-task behavior (*ICC* occurrence/min = 0.93; *ICC* duration/min = 0.97) was excellent. We transcribed all verbalizations during the coding process. Two independent raters categorized the transcriptions as monitoring or control behaviors and solved disagreements by discussion.

2.5 Self-regulation skills

To assess children's self-regulation skills in the classroom, teachers filled out two subscales, Plan/Organize and Emotional Control of the Behavior rating inventory of executive function-preschool version (Brief-P German version; Daseking and Petermann, 2013). We computed normed T-scores separated by age and gender for the Plan/Organize and Emotional Control scales. Normed mean scores are presented in Table 2. A lower T-score indicates fewer problems reported in the Plan/Organize and Emotional Control scales.

2.6 Ball-throwing

The task based on Schneider (1998) measures overconfidence. Participants were asked to throw 10 balls into a basket from a 120 cm distance. Participants started with a practice trial (10 balls). After the practice trial, they were asked to predict how many balls they would successfully throw into the box in the test trial. In the test trial, children threw 10 balls again.

We calculated metacognitive accuracy based on children's prediction [0–10] and test trial scores [0–10]. The accuracy score indicates the absolute difference between predicted and scored balls. A score closer to 0 indicates higher accuracy. Mean scores can be found in Table 2.

2.7 Statistical analysis

Shapiro Wilk tests revealed non-normal distributions of metacognitive behavior. Therefore, we relied on Spearman correlations and Mann–Whitney U Tests for group comparisons and Cohen's d as effect sizes. For data analysis, we used R (R Core Team, 2021). We computed Spearman correlations with the purrr package [version 1.0.2], MANOVAs with the manova() function of base R, and Mann–Whitney U Tests with the wilcox.test() function of base R. The R code for data analysis was developed with the support of OpenAI's GPT-4 model (OpenAI, 2024 version). The dataset and R script are available here: https://osf.io/3dtnv/.

3 Results

3.1 Descriptives

The final sample consisted of $n = 62$ children. Children spent $M = 247.84$ s ($SD = 67.48$) on the Sneaky Snake task. Regarding the solvable and the unsolvable parts, most children ($n = 55/62$) reached the unsolvable part of the task: They assembled the first three snake pieces correctly and started looking for the fourth missing piece. On average, participants worked on the task for $M = 94.03$ s ($SD = 47.45$) during the solvable part and for $M = 156.21$ s ($SD = 68.78$) during the unsolvable part. Table 1 reports the mean occurrences and duration of all observed behaviors in the Sneaky Snake task. Single categories dominated monitoring and control behavior. The most prevalent monitoring behavior was checking the plan (occurrences/min $M = 4.2$; duration/min $M = 11.25$), followed by inspecting a piece (occurrences/min $M = 2.69$; duration/min $M = 4.76$). The most prevalent control behavior was seeking (occurrences/min $M = 4.44$; duration/min $M = 24.78$), followed by adjustments (occurrences/min $M = 0.13$; duration/min $M = 0.66$). Therefore, we relied on sum scores of monitoring, control, and off-task behavior for the analyses. Making mistakes consisted of a single score. Interestingly, inspections of occurrences and duration of behaviors revealed slightly different patterns. While the most frequent behavior (occurrences/min) was monitoring, the longest (duration/min) behavior was control.

As indicated in Table 2, normed t-scores (normed for age and gender) on the BRIEF-P Plan/Organize scale and the Emotional Control scale were normally distributed, indicated by mean scores close to 50 and standard deviations close to 10. Moreover, in the ball-throwing task, children overestimated their performance. They predicted to score more balls than they did, which is typical for this age group (Schneider, 1998; Xia et al., 2023).

3.2 Validating the metacognitive behavior codings

To evaluate convergent validity we correlated metacognitive behavior in the Sneaky Snake task with a validated teacher questionnaire of self-regulated behavior (BRIEF-P). We expected negative correlations between the observed monitoring and control behaviors (occurrences and duration of the behaviors) and the BRIEF-P Plan/Organize scale. However, while controlling for age, we found no relation between monitoring or control behavior and the Plan/Organize scale for either of the variables, occurrence or duration. Also, off-task behavior and making mistakes were unrelated to the Plan/Organize scale.

Next, we explored the relationship between the observed monitoring and control behaviors and teacher ratings of self-regulated behavior in the BRIEF-P Emotional Control scale. While controlling for age, the results showed a negative correlation between monitoring occurrences and the Emotional Control scale ($\rho = -0.27$, $p = 0.042$), indicating that more monitoring is associated with fewer emotional control issues reported in the Emotional Control scale. Moreover, results showed a trend toward a negative correlation between control duration and the Emotional Control

TABLE 2 Mean scores BRIEF-P and Ball-throwing.

Scale	*M* (*SD*)	Range
BRIEF-P		
Plan/organize [t]	50.2 (10.31)	39–74
Emotional control [t]	47.45 (9.72)	40–76
Ball-throwing		
Prediction	7.1 (2.59)	2–10
Performance	5.03 (2.24)	0–10
Metacognitive accuracy	2.77 (2.43)	0–9

t = age and gender normed t-scores; Metacognitive accuracy = Performance - Prediction.

TABLE 3 Sneaky Snake mean scores for first and second kindergarten year.

Score	5-year-olds (Kindergarten 1) $n = 31$		6-year-olds (Kindergarten 2) $n = 31$	
	M (*SD*)	Range	*M* (*SD*)	Range
Monitoring				
Occurrences [occ/min]	6.71 (1.52)	3.8–10.4	7.44 (3.34)	1.6–17.11
Duration [s/min]	16.39 (4.58)	6.05–28.63	16.94 (7.5)	3.13–47.12
Control				
Occurrences [occ/min]	4.59 (1.29)	2.4–7.2	4.67 (1.32)	3–8.82
Duration [s/min]	23.05 (5.47)	14.98–37.58	28.27 (7.92)	9.9–43.66
Mistakes				
Occurrences [occ/min]	0.78 (0.92)	0–4.2	0.38 (0.38)	0–1.34
Duration [s/min]	7.00 (6.71)	0–24.14	3.3 (3.4)	0–10.95
Off-task				
Occurrences [occ/min]	1.24 (0.93)	0–3.6	0.87 (0.61)	0–2.75
Duration [s/min]	9.29 (6.48)	0–28.86	6.81 (6.66)	0–31.54

Significant differences between kindergarten 1 and 2 are bold ($p < 0.05$).

scale ($\rho = -0.25$, $p = 0.054$), indicating that longer control behavior is associated with fewer problems in regulating and controlling emotions. All other behaviors (monitoring duration, control occurrences, making mistake occurrences and duration, and off-task occurrences and duration) were unrelated to the Emotional Control scale.

Lastly, we evaluated the convergent validity between metacognitive behavior and metacognitive verbal performance prediction. Correlations controlled for age revealed no relation between monitoring or control behavior and metacognitive accuracy in the ball-throwing task for either of the variables, occurrence or duration. Also, off-task behavior and making mistakes were unrelated to the performance prediction in the ball-throwing task.

3.3 Age differences in metacognitive behavior

We compared 5 and 6-year-olds on occurrences and duration of monitoring, control, making mistakes, and off-task behavior. A MANOVA revealed a trend toward a significant age difference (*Pillai's trace* = 0.23, F (8, 53) = 1.99, $p = 0.066$). Following up with Mann–Whitney U Tests revealed that 5-year-olds showed less control behavior (duration: $U = 265$, $p = 0.002$; $d = -0.77$), a trend to more off-task behavior (duration: $U = 619.5$, $p = 0.051$; $d = 0.38$), and more mistakes (duration: $U = 630.5$, $p = 0.034$; $d = 0.69$) than 6-year-olds. Mann–Whitney U Tests for all other age comparisons (monitoring occurrences and duration, control occurrences, making mistake occurrences, off-task occurrences) were not significant. See Table 3 for means and Figure 2 for boxplots. In summary, we partially confirmed our hypothesis regarding age: Older children show more control and less off-task behavior and made fewer mistakes. However, we did not find an age difference in monitoring behavior.

3.4 Metacognitive behavior in the solvable and unsolvable task

To explore how the task's unsolvable nature affected children's behavior, we compared the behavior during the solvable part of the task (i.e., before reaching the missing piece) to the behavior during the unsolvable part of the task (i.e., after reaching the missing piece). Therefore, we compared the 60 s before the missing piece (solvable interval) with the 60 s after the missing piece (unsolvable interval). Wilcoxon signed-rank tests revealed that compared to the unsolvable interval in the solvable interval children exhibited more monitoring behavior (occurrences: $W = 983.5$, $p = 0.003$; $d = 0.43$; duration: $W = 1'049$, $p = 0.02$; $d = 0.28$), less control behavior (duration: $W = 372$, $p = 0.001$; $d = -0.48$), made more mistakes (duration: $W = 560$, $p < 0.001$; $d = 0.53$), and less off-task behavior (duration: $W = 150$, $p = 0.02$; $d = -0.34$). Wilcoxon signed-rank tests for all other task solvability comparisons (control occurrences, mistake occurrences, and off-task occurrences) were not significant. See Table 4 for mean scores and Figure 3 for boxplots. In summary, children show more monitoring, less control, and less off-task behavior in the solvable than unsolvable interval.

4 Discussion

Simulations of real-life scenarios in behavioral metacognitive tasks can provide insights into how young children's emerging metacognitive processes operate in everyday situations. In the present study, we developed an unsolvable behavioral task with high ecological validity to improve current behavioral methods to capture metacognitive monitoring and control behaviors in 5–6-year-olds. Building on existing behavioral metacognitive tasks (Bryce and Whitebread, 2012; Marulis and Nelson, 2021), three features were modified: First, to hold task difficulty constant across all participants, the task was designed to be unsolvable. Second, three distractors, shape, color, and size, were used to increase the frequency of different metacognitive strategies.

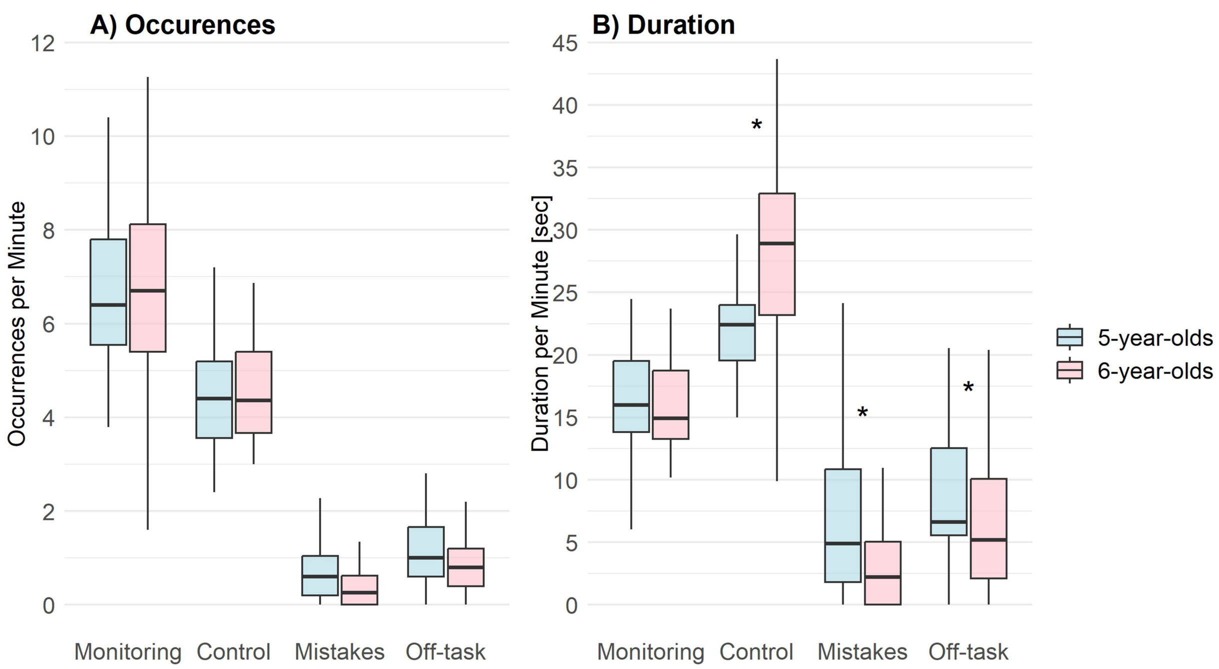

FIGURE 2
Boxplots for (A) occurrences and (B) duration of metacognitive behavior for 5 and 6-year-olds. *Significant differences between 5 (kindergarten 1) and 6-year-olds (kindergarten 2) at $p < 0.05$.

TABLE 4 Mean scores for behavior in the solvable and unsolvable task part.

Score	Solvable *n* = 55		Unsolvable *n* = 55	
	M (*SD*)	Range	*M* (*SD*)	Range
Monitoring				
Occurrences [occ/min]	9.29 (4.21)	2–18	7.19 (3.92)	0–20
Duration [s/min]	19.27 (8.64)	5.32–42.18	15.96 (9.53)	0–60
Control				
Occurrences [occ/min]	5.25 (1.75)	2–9	4.84 (2.14)	0–10
Duration [s/min]	22.3 (8.92)	8.81–48.53	29.9 (13.05)	0–50.31
Mistakes				
Occurrences [occ/min]	1.06 (1.03)	0–4	0.87 (1.2)	0–6
Duration [s/min]	8.85 (10.04)	0–39.88	3.12 (6.38)	0–28.15
Off-task				
Occurrences [occ/min]	0.51 (1.12)	0–7	0.75 (1.14)	0–5
Duration [s/min]	1.53 (4.8)	0–31.61	4.37 (7.48)	0–40.92

Scores for the solvable part are based on the 60 s before children reached the unsolvable piece, and scores for the unsolvable part are based on the 60 s after children reached the unsolvable piece. Significant differences between the solvable and unsolvable task part are bold ($p < 0.05$).

Third, to understand metacognitive processes more comprehensively, in addition to observing the frequency of the observable behaviors, the duration of the behaviors was coded, too.

The results for the Sneaky Snake task can be summarized as follows: Overall, the analysis showed that the two most frequently observed metacognitive behaviors were "seeking" (i.e., metacognitive control behavior) and "checking the plan" (i.e., metacognitive monitoring behavior). "Checking the plan" was shown twice as often than the next frequent behavior, "inspecting a piece." The difference for metacognitive control was even more pronounced: "Seeking" was shown four times more than the next frequent behavior, "adjusting." These differences are also reflected in the duration of how long the behaviors were shown.

Other than expected, the correlation analysis showed that the observed metacognitive behaviors were not related to the teacher's estimations of children's metacognitive regulation skills measured with the BRIEF-P Plan/Organize scale. Moreover, metacognitive

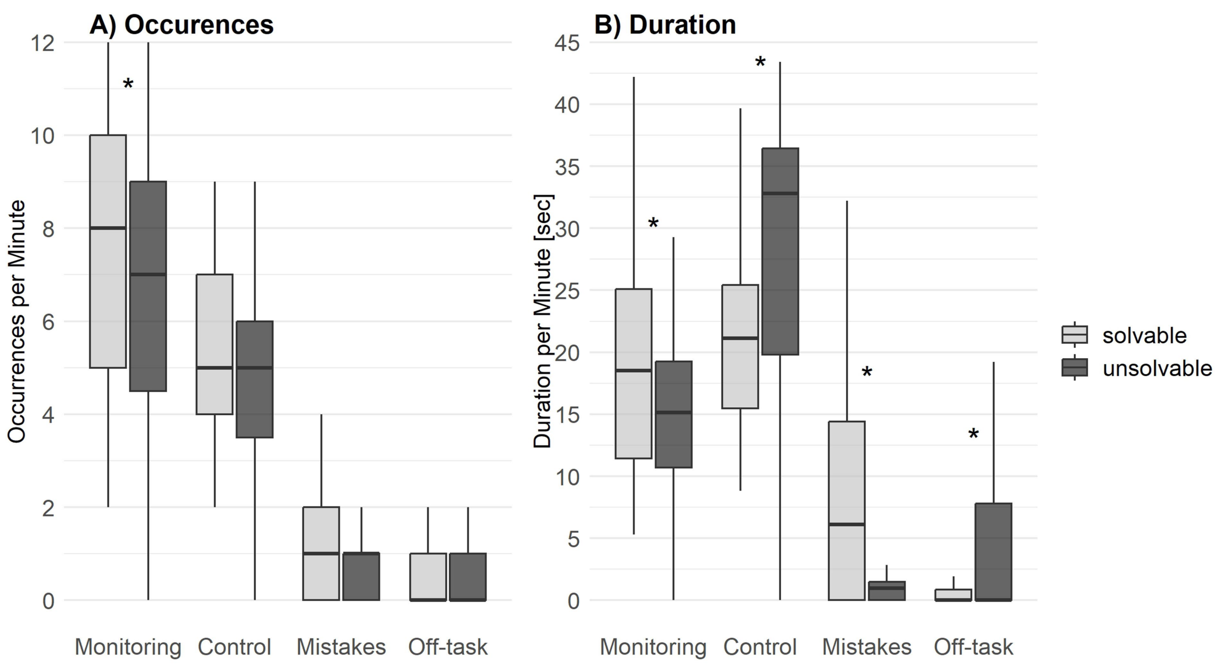

FIGURE 3
Boxplots for (A) occurrences and (B) duration of metacognitive behavior in the solvable and unsolvable part. *Significant differences in metacognitive behavior between the solvable and unsolvable part at $p < 0.05$.

accuracy in a ball-throwing task (i.e., a classical verbal assessment of metacognition) was not related to metacognitive behavior. Expected age differences were found for control behavior (i.e., seeking behavior) and off-task behavior but not for monitoring behavior or making mistakes. Comparing the solvable part of the task to the unsolvable part showed that while the task was solvable, children showed more monitoring but less control behavior. Once the task was unsolvable, children tended to show more off-task behavior.

4.1 Evaluating the unsolvable metacognitive task

The Sneaky Snake task was developed to address existing constraints in capturing behavioral metacognitive processes in young children in a real-life, familiar play context. The task consists of wooden train track pieces. These wooden train tracks are a common toy in kindergartens and children's homes in Switzerland. However, to ensure that all children were familiar with the task, we included an extensive practice trial with feedback. The practice trial was without any time limit, so every child could take as much time as they needed to get familiar with the task.

We included the BRIEF-P questionnaire (Daseking and Petermann, 2013) to validate the behavioral task. Other than expected, we found no relationship between metacognitive behavior in the Sneaky Snake task and teacher ratings of metacognitive regulation for either of the two variables, occurrences or duration. These findings indicate that the behavior children show when asked to work independently on a problem-solving task does not match the teacher's estimation of how well a child is able to plan and organize their behaviors to pursue a goal in the classroom context. While these findings are different from what we expected, it might be that the regulation demands during the Sneaky Snake tasks differ substantially from the metacognitive regulation demands in the classroom context. Thus, it might be that the two measurements focus on different aspects of metacognition. More specifically, the Sneaky Snake captures metacognitive behavior in a single play session, whereas in a typical classroom setting, children must monitor and control their behavior in the presence of many other children or when working in a group setting with peers. Given that metacognition has been suggested to be domain-specific (e.g., Baer et al., 2021; van Loon et al., 2024), and these two measurements capture different metacognitive aspects, this may explain the lack of correlations between the BRIEF-P and the Sneaky Snake.

Moreover, the plan/organize scale includes only one item that captures the child's ability to work on a difficult task; the remaining nine questions capture more classroom situations demanding to follow classroom rules. In addition to the plan/organize scale, the BRIEF-P also allows the computation of an emergent metacognition index as described in the BRIEF-P manual. This index combines the plan/organize and working memory scale. The working memory scale captures additional metacognitive components (e.g., difficulty staying on task or following multi-step instructions). However, the working memory scale was not assessed in the present study, and consequently, the metacognitive index could not be computed. It might be that the metacognitive index would reveal relationships with metacognitive behaviors that were not apparent with the plan/organize scale. Using a different questionnaire, Bryce and Whitebread (2012) found positive relations between metacognitive behavior and teacher rating with the CHILD questionnaire (Whitebread et al., 2009). The different findings may be explained by a slightly different focus of the two questionnaires: Whereas the CHILD questionnaire assesses adaptive metacognitive skills in the classroom (e.g., uses previously taught strategies), the BRIEF-P is a clinical scale focusing on metacognitive regulation problems (e.g., does not complete tasks, even after receiving hints). It may be that the CHILD questionnaire captures metacognitive skills more closely aligned with the Sneaky Snake task than the BRIEF-P.

Future research could also consider a questionnaire more closely related to metacognitive behavior reflected in the Sneaky Snake, such as the CHILD questionnaire. Moreover, a questionnaire on children's metacognitive knowledge about the Sneaky Snake could be valuable for gaining further insights into children's metacognitive awareness. For instance, Marulis and Nelson (2021) interviewed children after the Wedgits© Puzzle with the Metacognitive Knowledge Interview (e.g., Do you think you did a good or not so good job on the puzzles?; see Marulis et al., 2016). By relating the child's answers during the interview to their behavior, the Metacognitive Knowledge Interview may provide insight into the child's level of consciousness during the Sneaky Snake task.

Interestingly, metacognitive monitoring and control were related to teacher-rated emotional control skills. We found that children who more frequently monitored their behavior and spent more time with control actions (mainly seeking the snake pieces) were also better able to control their emotions. Thus, it seems that the ability to control one's own emotions may be crucial to maintaining metacognitive monitoring and control when facing an unsolvable task.

Lastly, we also explored the relationship between metacognitive accuracy in the ball-throwing task and metacognitive behavior in the Sneaky Snake. The results showed that metacognitive accuracy in the ball-throwing task was not related to metacognitive monitoring and control behavior in the Sneaky Snake task. As mentioned previously, such performance predictions as an indicator of metacognitive accuracy are an explicit and verbal assessment of children's metacognition (e.g., Xia et al., 2023). The present finding suggests that verbal assessment of metacognition and metacognitive behavioral processes reflect distinct metacognitive processes that, especially in early development, may operate more independently. The distinct measurement domains might also explain the zero correlations. Furthermore, the Sneaky Snake task is a problem-solving task, whereas the ball-throwing task is a motor task. As suggested above, metacognition in young children may be domain-specific (Baer et al., 2021; van Loon et al., 2024), explaining why metacognition in the Sneaky Snake task and metacognition in the ball-throwing task were unrelated. Future research could assess behavioral and verbal metacognition in the same task to further investigate whether behavioral and verbal metacognitive processes develop independently or develop differently between domains.

Comparing the Sneaky Snake task to existing behavioral tasks such as the Train track (Bryce and Whitebread, 2012) and the Wedgits© task (Marulis and Nelson, 2021) shows that increasing the number of distractors does not necessarily increase the frequency of the metacognitive monitoring and control behaviors. Similar to the train track study in the present study, not all behaviors were observed at a minimum frequency to be analyzed. In fact, three monitoring behaviors and two control behaviors were so rarely shown that we had to exclude them from the analysis. To address the limited frequency issue, developing more complex behavioral tasks involving multiple subsequent steps may be a way to elicit more diverse metacognitive behaviors in the participants.

4.2 Metacognitive processes captured through behavioral observation

As expected, we found age differences in several metacognitive indices (i.e., control, off-task, mistakes). Older children showed more control behavior than younger children. More specifically, when older children showed metacognitive control, they tended to show the behavior for longer periods but not necessarily more frequently. Contrary to our expectations and different than Bryce and Whitebread's (2012) findings, we did not find age differences in monitoring behavior. The age range between the investigated groups might explain the different findings. Bryce and Whitebread (2012) compared 5-year-olds with 7-year-olds, whereas we compared 5-year-olds with 5 years and 9-months-olds. Age differences in monitoring are likely more pronounced when comparing groups of children with a more significant age difference. Finally, we also found shorter periods of off-task behavior and mistakes in older children than in younger ones. These findings align with the literature suggesting that metacognitive failure decreases with age (Bryce and Whitebread, 2012). Overall, the present findings suggest that in kindergarten, differences in metacognitive behavior occur primarily in the duration, not the frequency of metacognitive behavior. More specifically, older children may spend more time with goal-directed control behavior, which may be related to making fewer mistakes and showing less off-task behavior. The differing pattern of results for occurrences and duration emphasizes the importance of including both measurements in future studies.

Furthermore, the present findings are also somewhat contradictory to longitudinal studies with verbal metacognitive assessments. Whereas Bayard et al. (2021) and Gonzales et al. (2021) found more pronounced developmental improvements in metacognitive monitoring than control, we found age differences in metacognitive control but not in metacognitive monitoring. This opposing result pattern between the present results and verbal metacognition assessments emphasizes the need to further understand how language skills and metacognitive awareness may influence these verbal judgments. One way to address this knowledge gap is through longitudinal research, which includes both measurement approaches: behavioral metacognitive task and verbal metacognitive assessments. Through such an approach, we would be able to disentangle how language and metacognitive awareness might be driving developmental trajectories of monitoring and control.

The unique feature of the Sneaky Snake task is that it is unsolvable. The main aim of designing an unsolvable task was to hold task difficulty constant for all participants. Most children (89%) reached the unsolvable part of the task, indicating that task difficulty was indeed comparable between children. However, the fact that the task consists of a solvable part and then becomes unsolvable allows us to examine an increase in metacognitive regulation demands. When the participant reaches the unsolvable part of the task, the regulation demands increase significantly as no moment of success facilitates metacognitive regulation and motivation to complete the task. The change in regulation demands was mirrored nicely in all four observed behaviors: Comparing the behaviors shown in the solvable part to the unsolvable part showed that while monitoring behavior and making mistakes decreased from the solvable to the unsolvable task part, control and off-task behavior increased. These behavioral changes may reflect the changing task demands from the solvable to the unsolvable part. While monitoring one's progress when building the snake is crucial, the same amount of progress cannot be made in the unsolvable part when searching for the missing piece. Searching for the missing piece requires a high maintenance of goal-directed behavior when facing difficulty. The higher metacognitive demands in the unsolvable part may also explain the increase in off-task behavior. Finally, fewer mistakes in the unsolvable part may result from the seeking behavior; while children were searching for the next piece, they did not place any pieces, consequently lowering the risk of making mistakes. Overall, in terms of ecological validity, the shift in the task from solvable to unsolvable mimics real-life situations quite accurately. In class and more generally when learning something new, most children are faced with the situation that initially, when starting the task, they can complete the first part but then are confronted with difficulty. Maintaining this edge is where learning eventually happens. It is also precisely at this threshold and beyond where metacognitive skills are most needed to accomplish a goal successfully. The current version of the task has yet to be improved. However, examining metacognitive processes at the threshold from solvable to unsolvable, as well as when the task is unsolvable, may be interesting for future research to gain a more detailed understanding of metacognitive processes in action.

4.3 Limitations

Even though our aim was to develop a task to address constraints in existing behavioral tasks, with our adaptions, we could not reliably observe all behaviors as planned in the first version of the coding scheme (see Table 1). Especially "glancing behavior" was difficult to distinguish. For instance, when a child puts a piece next to the plan, it was difficult to distinguish whether the child solely glanced at the plan ("checking the plan") or actively glanced back and forth between the piece and the plan ("comparing a single piece with the plan"). Therefore, we combined some of the categories in the second version of the coding scheme. For instance, we coded "checking the plan" and "comparing a single piece with the plan" as the same behavior ("checking the plan"). To address this issue in future studies, the task size should be increased. For example, instead of a picture of the snake, a same-size snake model could be used. The snake model should be placed further away from the building mat to allow for a more precise distinction between checking the model or comparing the piece with the model.

Motivation has an essential impact on any human behavior including metacognition (e.g., Efklides, 2011; Marulis and Nelson, 2021; Zimmerman and Moylan, 2009). Therefore, it's possible that motivation affects the four observed behaviors (i.e., monitoring, control, mistakes, and off-task) to different degrees and that the demands on motivation even increased during the unsolvable part of the task. Unfortunately, in the current design, it is not possible to disentangle motivation from the observed behaviors. Further research using different incentives could investigate how motivation is related to monitoring, control, mistakes, and off-task behavior.

Furthermore, the focus of the BRIEF-P questionnaire made it difficult to validate the task. The zero relation between the metacognitive behavior and the teacher ratings limit our understanding of the extent to which the task effectively captures metacognitive behavior in 5–6-year-old kindergarten children. However, the BRIEF-P working memory scale may be interesting to include in future studies. Combining the working memory scale with the plan/organize scale would allow computing an emergent metacognition index, which might be more closely related to metacognitive behavior. Finally, the present cross-sectional study limits our understanding of developmental differences in metacognitive behavior. Longitudinal designs are required to understand developmental differences in more detail.

4.4 Conclusion

We investigated young children's metacognitive behavior in an unsolvable task. The task was designed to gain insight into how children's metacognitive processes operate in a problem-solving task that mimics real-life scenarios (e.g., Bryce and Whitebread, 2012; Marulis and Nelson, 2021). Similar to previous studies (e.g., Bryce and Whitebread, 2012; see for reviews Roebers, 2017; Xia et al., 2023), we found age differences in metacognition. Older children showed longer control behavior than younger children. Furthermore, results suggest differing metacognitive behaviors depending on whether a task is solvable or unsolvable. We observed more monitoring and less control behaviors in the solvable than unsolvable part of the task. Although the task still needs further improvement, the unsolvable nature of the task assesses metacognitive processes at a crucial threshold: Most learning happens at the edge between solvable and unsolvable, similar to what Vygotsky described as the zone of proximal development (Vygotsky, 1978). The nature of the Sneaky Snake task allows us to capture metacognitive processes precisely at this edge, potentially providing insight into metacognitive processes during a crucial moment in the learning process. The current study contributes to the research methodology to capture metacognitive processes in action by introducing an unsolvable behavioral metacognitive task.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/3dtnv/.

Ethics statement

The study involving humans was approved by the Ethikkommission der Phil.-hum. Fakultät University of Bern. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

FB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. NO: Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The study was supported by the following grant: Swiss National Science Foundation, Grant/Award Number: 1001C_197336.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Baer, C., Ghetti, S., and Odic, D. (2021). Perceptual and memory metacognition in children. Proceedings of the annual meeting of the cognitive science perceptual and memory metacognition in children 43, 3061–3067.

Bayard, N. S., van Loon, M. H., Steiner, M., and Roebers, C. M. (2021). Developmental improvements and persisting difficulties in children's metacognitive monitoring and control skills: cross-sectional and longitudinal perspectives. *Child Dev.* 92, 1118–1136. doi: 10.1111/cdev.13486

Bryce, D., and Whitebread, D. (2012). The development of metacognitive skills: evidence from observational analysis of young children's behavior during problem-solving. *Metacogn. Learn.* 7, 197–217. doi: 10.1007/s11409-012-9091-2

Bryce, D., Whitebread, D., and Szűcs, D. (2015). The relationships among executive functions, metacognitive skills and educational achievement in 5 and 7 year-old children. *Metacogn. Learn.* 10, 181–198. doi: 10.1007/s11409-014-9120-4

Coughlin, C., Hembacher, E., Lyons, K. E., and Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Dev. Sci.* 18, 957–971. doi: 10.1111/desc.12271

Daseking, M., and Petermann, F. (2013). BRIEF-P-Verhaltensinventar zur Beurteilung exekutiver Funktionen für das Kindergartenalter [German adaptation of the behavior rating inventory of executive function® - preschool version (BRIEF®-P)]. Bern: HUBER.

Destan, N., Hembacher, E., Ghetti, S., and Roebers, C. M. (2014). Early metacognitive abilities: the interplay of monitoring and control processes in 5-to 7-year-old children. *J. Exp. Child Psychol.* 126, 213–228. doi: 10.1016/j.jecp.2014.04.001

Destan, N., and Roebers, C. M. (2015). What are the metacognitive costs of young children's overconfidence? *Metacogn. Learn.* 10, 347–374. doi: 10.1007/s11409-014-9133-z

Dunlosky, J., and Metcalfe, J. (2009). Metacognition: a textbook for cognitive, educational, life span & applied psychology. Thousand Oaks, CA: Sage Publications.

Dunlosky, J., Mueller, M. L., and Thiede, K. W. (2016). "Methodology for investigating human metamemory: problems and pitfalls" in The Oxford handbook of metamemory. eds. J. Dunlosky and S. K. Tauber (New York, NY: Oxford University Press).

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: the MASRL model. *Educ. Psychol.* 46, 6–25. doi: 10.1080/00461520.2011.538645

Federal Statistical Office. (2024). *Compulsory education*. Available at: https://www.bfs.admin.ch/bfs/en/home/statistics/education-science/pupils-students/compulsory.html.

Gonzales, C. R., Merculief, A., McClelland, M. M., and Ghetti, S. (2021). The development of uncertainty monitoring during kindergarten: change and longitudinal relations with executive function and vocabulary in children from low-income backgrounds. *Child Dev.* 93, 524–539. doi: 10.1111/cdev.13714

Hembacher, E., and Ghetti, S. (2014). Don't look at my answer: subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. *Psychol. Sci.* 25, 1768–1776. doi: 10.1177/0956797614542273

Kim, S., Senju, A., Sodian, B., Paulus, M., Itakura, S., Okuno, A., et al. (2021). Memory monitoring and control in Japanese and German preschoolers. *Mem. Cogn.* 51, 708–717. doi: 10.3758/s13421-021-01263-1

Lyons, K. E., and Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Dev.* 82, 1778–1787. doi: 10.1111/j.1467-8624.2011.01649.x

Marulis, L. M., Baker, S. T., and Whitebread, D. (2020). Integrating metacognition and executive function to enhance young children's perception of and agency in their learning. *Early Child Res. Q.* 50, 46–54. doi: 10.1016/J.ECRESQ.2018.12.017

Marulis, L. M., and Nelson, L. J. (2021). Metacognitive processes and associations to executive function and motivation during a problem-solving task in 3–5 year olds. *Metacogn. Learn.* 16, 207–231. doi: 10.1007/s11409-020-09244-6

Marulis, L. M., Palincsar, A. S., Berhenke, A. L., and Whitebread, D. (2016). Assessing metacognitive knowledge in 3–5 year olds: the development of a metacognitive knowledge interview (McKI). *Metacogn. Learn.* 11, 339–368. doi: 10.1007/s11409-016-9157-7

Nelson, T. O., and Narens, L. (1990). Metamemory: a theoretical framework and new findings. *Psychol. Learn. Motivat. Adv. Res. Theory* 26, 125–173. doi: 10.1016/S0079-7421(08)60053-5

Oeri, N., and Roebers, C. M. (2021). Task persistence in kindergarten children: disentangling age from schooling effects. *Br. J. Dev. Psychol.* 39, 217–230. doi: 10.1111/BJDP.12358

OpenAI. (2024). ChatGPT (October 2024 version) [Large language model]. Available at: https://chat.openai.com/.

R Core Team (2021). R: a language and environment for statistical computing [computer software]. Available at: https://www.r-project.org/ (Accessed October 21, 2024).

Roebers, C. M. (2017). Executive function and metacognition: towards a unifying framework of cognitive self-regulation. *Dev. Rev.* 45, 31–51. doi: 10.1016/j.dr.2017.04.001

Schneider, W. (1998). Performance prediction in young children: effects of skill, metacognition and wishful thinking. *Dev. Sci.* 1, 291–297. doi: 10.1111/1467-7687.00044

van Loon, M., Orth, U., and Roebers, C. (2024). The structure of metacognition in middle childhood: evidence for a unitary metacognition-for-memory factor. *J. Exp. Child Psychol.* 241:105857. doi: 10.1016/J.JECP.2023.105857

Vygotsky, L. S. (1978). "Interaction between learning and development" in Readings on the development of children. eds. M. Gauvain and M. Cole (New York, US: Scientific American Books), 34–40.

Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., et al. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacogn. Learn.* 4, 63–85. doi: 10.1007/s11409-008-9033-1

Xia, M., Poorthuis, A. M. G., and Thomaes, S. (2023). Why do young children overestimate their task performance? A cross-cultural experiment. *J. Exp. Child Psychol.* 226:105551. doi: 10.1016/j.jecp.2022.105551

Zimmerman, B. J., and Moylan, A. R. (2009). "Self-regulation: where metacognition and motivation intersect" in Handbook of metacognition in education. eds. B. J. Zimmerman and A. R. Moylan (New York, US: Routledge).