Sneaky Snake: Assessing metacognitive behavior in kindergarten children with an unsolvable task

Florian Jonas Buehler & Niamh Oeri

University of Bern, Switzerland

Conflicts of Interest

The author declares no conflict of interest.

Corresponding Author

Florian Jonas Buehler: florian.buehler@unibe.ch

Draft version. This paper has not been peer-reviewed.

Abstract

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 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 intervals than in the unsolvable intervals. The current study contributes to the research methodology to capture metacognitive skills in action by introducing an unsolvable behavioral metacognitive task.

Keywords: Metacognitive behavior, monitoring, control, kindergarten

Metacognitive processes are typically assessed verbally. In these tasks, children are often asked to answer questions such as "*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 for preschool 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 (Bryce and Whitebread, 2012; Marulis and Nelson, 2021). Such problem-solving tasks allow observing metacognition closer related to real-life scenarios. 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 skills in preschoolers.

The theoretical foundations of metacognition

Most research on procedural metacognition is based on the influential framework by Nelson and Narens (1990). The metamemory framework differentiates between metacognitive monitoring and control. 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 three-year-olds can monitor their performance by reporting higher confidence in correct than incorrect trials (e.g., identifying degraded pictures; Coughlin et al., 2015; Lyons & Ghetti, 2011). From age four, 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 five, children show signs of metacognitive control as they are more likely to withdraw an incorrect answer than a correct answer (Destan et al., 2014; Destan and Roebers, 2015). 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 preschoolers, yielding biased results for children with low language skills. Thus, non-verbal metacognitive tasks provide insights into children's behaviorally occurring metacognitive skills. Studies focusing on metacognitive behavior rather than on explicit classical judgments/measures (Bryce and Whitebread, 2012; Marulis and Nelson, 2021) are scarce but reveal similar developmental patterns: From the age of three, 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).

Behavioral metacognition tasks

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 to 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 skills 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 monitoring or control behaviors. Results showed that metacognitive monitoring and control can be reliably observed at the age of three. 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 skills 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 skills. 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). 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 skills play out and which aspects might be especially challenging to exert metacognition successfully.

Methodological challenges in assessing metacognition behaviorally

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 four minutes (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. Even though the previous experience may impact task motivation and potentially more diverse strategies to approach 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 analyse 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 eight to eleven 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 one second and child B searches for 10 seconds, 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.

The current study

Building on the foundation of Bryce and Whitebread (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 to 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. The unsolvable nature of the task keeps task difficulty constant across participants, independent of previous experience. It prevents ceiling effects as no participant can fully complete the task (e.g., Dunlosky et al., 2016). Finally, the task's convergent validity was evaluated by comparing the observed metacognitive behavior (monitoring and control) with the BRIEF-P Plan / Organize scale (German version: Daseking and Petermann, 2013), as well as by exploring the relation between verbal assessment of children's metacognitive behavior in a ball-throwing task (Schneider, 1998).

The hypotheses were the following: (1) We expect the observed monitoring and control behaviors in kindergarten children to be negatively related to teacher ratings of self-regulated behavior (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 longer monitoring and control behavior in older 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).

Method

Participants

In the preregistered study (https://doi.org/10.17605/OSF.IO/ZX86H), we relied on a random subsample from a larger study on children's self-regulated learning (*N* = 219). The target sample size for the present study was *N* = 66 children and is based on previous studies (Bryce and Whitebread, 2012). 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 kindergartens. 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.

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. Trained experimenters individually tested participants. Testing took place in a quiet room at the kindergarten. Among all administered tasks, task order was counterbalanced.

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 trail version) consisted of seven 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 trail 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 five minutes or whenever the child ended the task. Children's behavior was videorecorded. We excluded ten children from the analyses because they were not able to successfully complete the practice trial, interrupted the task (i.e., going to the bathroom), were interrupted by an external person, or understood that the task was unsolvable (children who had systematically tried all the distractor pieces, $N = 1$).

Coding scheme for metacognitive behavior

We coded children’s behavior similarly to Bryce and Whitebread (2012) coding scheme. We coded for monitoring, control, making mistakes, and off-task behavior. To differentiate between monitoring and control behavior, we relied on Nelson and Naren's framework (1990), describing monitoring as a bottom-up process accumulating 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). In contrast, making mistakes describes monitoring failure (e.g., building in an incorrect piece) and off-task behavior describes control failure (e.g., walking away from the task).

We assessed the occurrences and duration of each behavior. Table 1 shows an overview of the coded behaviors, including their occurrences and durations. After the first coding round, we had to drop four behaviors (*“Comparing a single piece with the plan”, “Comparing pieces”, “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. See Table 1 for the final coding scheme.

Because the total test time varied between subjects, we divided the duration and occurrences per behavior through the total minutes spent on the task. 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/minute = .84; ICC duration/minute = .63), control (ICC occurrence/minute = .92; *ICC* duration/minute = .97), making mistakes (*ICC* occurrence/minute = .93; ICC duration/minute = .96), and off-task behavior (*ICC* occurrence/minute = .94; *ICC* duration/minute = .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.

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.

Ball-throwing

The task based on Schneider (1998) measures overconfidence. Participants were asked to throw ten 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 ten 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.

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 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.

Results

Descriptives

Children spent M = 247.84 seconds (SD = 67.48) on the Sneaky Snake task. Regarding the solvable and the unsolvable parts, participants worked on the task for M = 94.03 seconds (SD = 47.45) during the solvable part and for M = 156.21 seconds (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/minute

M = 4.2; duration/minute *M* = 11.25), followed by inspecting a piece (occurrences/minute *M* = 2.69; duration/minute *M* = 4.76). The most prevalent control behavior was seeking (occurrences/minute *M* = 4.44; duration/minute *M* = 24.78), followed by adjustments (occurrences/minute *M* = 0.13; duration/minute *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/minute) was monitoring, the longest (duration/minute) behavior was control.

As indicated in Table 2, normed t-scores (normed for age and gender) on the BRIEF Plan/Organize scale and the Emotional Control scale were normally distributed, indicated by mean scores close to 50 and standard deviations close to ten. 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).

Validating the metacognitive behavior codings

We expected negative correlations between the observed monitoring and control behaviors (occurrences and duration of the behaviors) and teacher ratings of self-regulated behavior in the BRIEF 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 Emotional Control scale. While controlling for age, the results showed a negative correlation between monitoring occurrences and the Emotional Control scale ($\rho = -.27$), 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 scale ($\rho = -.25$, $p = .054$), 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 explored the relationship 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. However, we found a trend toward a positive relation between duration off-task and metacognitive accuracy ($\rho = .25$, $p = .073$), indicating that longer off-task behavior is associated with less accurate metacognitive judgments of one's performance in the ball-throwing task. All other behaviors (making mistakes, occurrences and duration, and off-task occurrences) were unrelated to metacognitive accuracy.

Age differences in metacognitive behavior

We compared first- and second-year kindergarten children on occurrences and duration of monitoring, control, making mistakes, and off-task behavior. A MANOVA revealed a trend towards a significant kindergarten year difference (*Pillai's trace* = 0.24, F (8, 53) = 2.05, $p = .058$). Following up with Mann-Whitney U Tests revealed that first-year kindergarten children showed less control behavior (duration: $U = 265$, $p = .002$; $d = -0.77$), more off-task behavior (duration: $U = 632.5$, $p = .033$; $d = 0.35$), and more mistakes (duration: $U = 630.5$, $p = .034$; $d = 0.69$) than second-year kindergarten children. Mann-Whitney U Tests for all other kindergarten year comparisons (monitoring occurrences and duration, control occurrences, making mistake occurrences, off-task occurrences) were not significant. See Table 3 for means and Figures 2a and 2b 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.

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 seconds before the missing piece (solvable interval) with the 60 seconds 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: $W = 983.5$, $p = .003$; $d = 0.43$; duration: $W = 1049$, $p = .020$; $d = 0.28$), less control behavior (duration: $W = 372$, $p = .001$; $d = -0.48$), made more mistakes (duration: $W = 560$, $p < .001$; $d = 0.53$), and less off-task behavior (duration: $W = 139$, $p = .012$; $d = -0.38$). 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 Figures 3a and 3b for boxplots. In summary, children show more monitoring, less control, and less off-task behavior in the solvable than unsolvable interval.

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 kindergarten children. Building on existing behavioral metacognitive tasks (Bryce and Whitebread, 2012), 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. Third, to understand metacognitive skills 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 how long the behaviors were shown.

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

Evaluating the unsolvable metacognitive task

The Sneaky Snake task was developed to address existing constraints in capturing behavioral metacognitive skills in young children in a real-life, familiar play context. We included the BRIEF 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. The plan and 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. 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 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.

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. However, metacognitive control failure captured through off-task behavior was related to the ball-throwing task. More precisely, children who inaccurately estimated their performance in the ball-throwing task also showed more off-task behavior during the Sneaky Snake task. Thus, in terms of metacognitive failure, poorer performance predictions and behavioral control failure were related. However, these are initial findings. Further research using multimethodological approaches is needed to fully understand the relationship between verbal and behavioral metacognitive processes.

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 behavior. 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.

Metacognitive skills captured through behavioral observation

As expected, we found age differences in metacognitive behavior. 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. The present findings suggest that age differences are reflected in more consistent control behavior (longer durations) and not in the frequency of control behavior (occurrences). The differing pattern of results for occurrences and duration emphasizes the importance of including both measurements in future studies. Contrary to our expectations and different than Bryce and Whitebread (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 five-year-olds with seven-year-olds, whereas we compared first (*M* = five years) and second-year kindergarten children (*M* = five years and nine months). 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).

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. 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 past 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 skills 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 skills in action.

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. Furthermore, the focus of the BRIEF questionnaire made it difficult to validate the task. The zero relations between the metacognitive behavior and the teacher ratings limit our understanding of how well task captures metacognitive behavior in preschool children. 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.

Conclusion

We investigated young children’s metacognitive behavior in an unsolvable task. The task was designed to gain insight into how children's metacognitive skills operate in a problem-solving task that mimics real-life scenarios. 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 behavior depending on whether a task is solvable or unsolvable. We observed more monitoring and less control behavior in the solvable than unsolvable part of the task. Although the task still needs further improvement, the unsolvable nature of the task assesses metacognitive skills 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 this edge, potentially providing insight into metacognitive skills during a crucial moment in the learning process. The current study contributes to the research methodology to capture metacognitive skills in action by introducing an unsolvable behavioral metacognitive task.

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)
Mistakes (monitoring failure)	Child builds in an incorrect piece.	0.58 (0.73)	5.15 (5.6)
Off-task (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.	0.26 (0.23)	3.63 (4.3)
Sum score off-task	-	1.01 (0.79)	7.43 (6.28)

Note. 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.

Table 2

Mean scores BRIEF and Ball-Throwing

Scale	*M* (*SD*)	Range
BRIEF		
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
Accuracy	2.77 (2.43)	0-9

Note.

Table 3

Sneaky Snake mean scores for first and second kindergarten year.

Score	Kindergarten 1 $N = 31$		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 [sec/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 [sec/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 [sec/min]	7 (6.71)	0-24.14	3.3 (3.4)	0-10.95
Off-task				
Occurrences [occ/min]	1.2 (0.88)	0-3.6	0.82 (0.63)	0-2.75
Duration [sec/min]	8.64 (5.98)	0-28.86	6.22 (6.43)	0-31.54

Note. Accuracy = absolute value of performance – prediction; Solvable duration = time to solvable piece; Unsolvable duration = time from unsolvable piece to end. Significant differences between kindergarten 1 and 2 are bold ($p < .05$).

Table 4

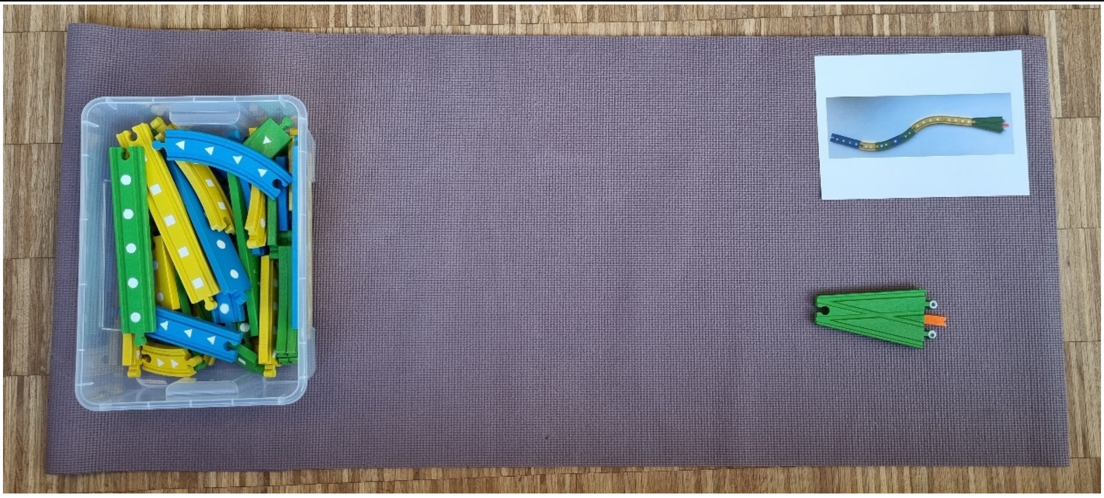
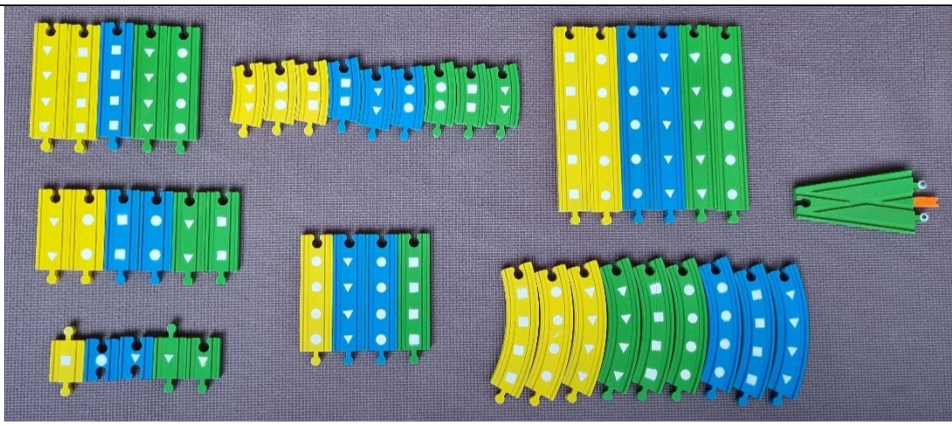
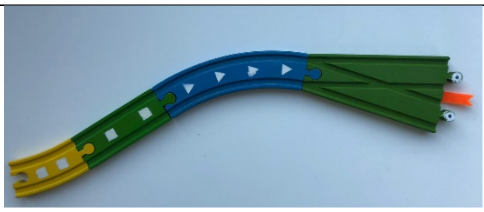
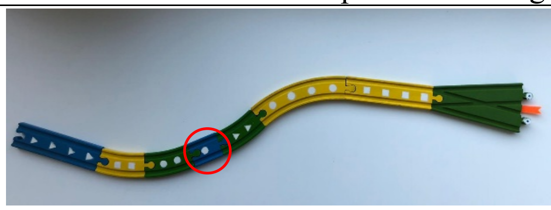
Mean scores for behavior in the solvable and unsolvable task part.

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 [sec/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 [sec/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 [sec/min]	8.85 (10.04)	0-39.88	3.12 (6.38)	0-28.15
Off-task				
Occurrences [occ/min]	0.49 (1.02)	0-6	0.75 (1.14)	0-5
Duration [sec/min]	1.34 (3.64)	0-20.96	4.37 (7.48)	0-40.92

Note. Scores for the solvable part are based on the 60 seconds before children reached the unsolvable piece, and scores for the unsolvable part are based on the 60 seconds after children reached the unsolvable piece. Significant differences between the solvable and unsolvable task part 2 are bold ($p < .05$).

Figure 1

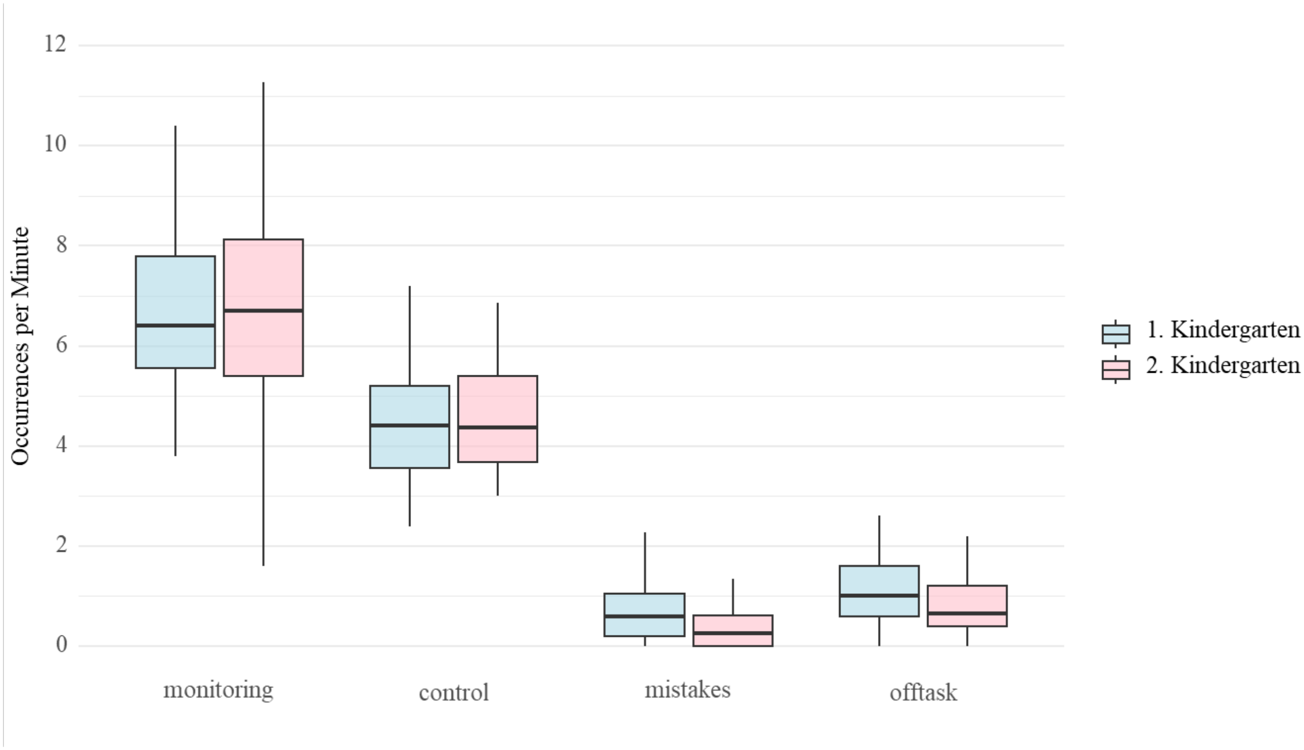
The Sneaky Snake Task

<table>
<tr><td colspan="2">Setup: Box (40 x 24 x 13cm) on the left with target and distractor pieces; plan in the top right corner (A4), Head of the snake (starting point) below the plan; mat (140 x 60cm)</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td colspan="2">Available Pieces: 44 target and distractor pieces in the box</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>Plan Practice Trial</td><td>Plan Test Trial: the circled piece is missing</td></tr>
<tr><td></td><td></td></tr>
</table>

Note. The Sneaky Snake for kindergarten children.

Figure 2a

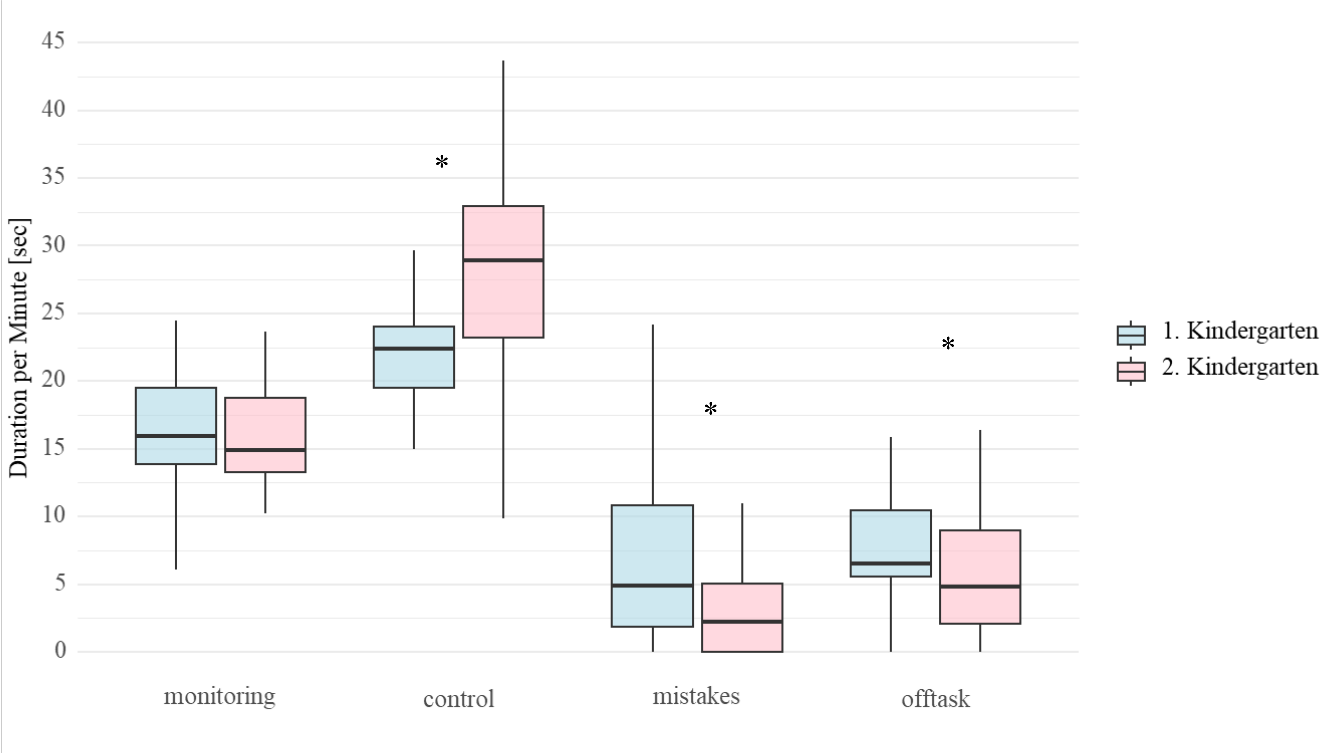
Boxplots for occurrences of metacognitive behavior for kindergarten 1 and kindergarten 2



Note. Children from kindergarten 1 and kindergarten 2 did not differ in occurrences of metacognitive behavior in the Sneaky Snake.

Figure 2b

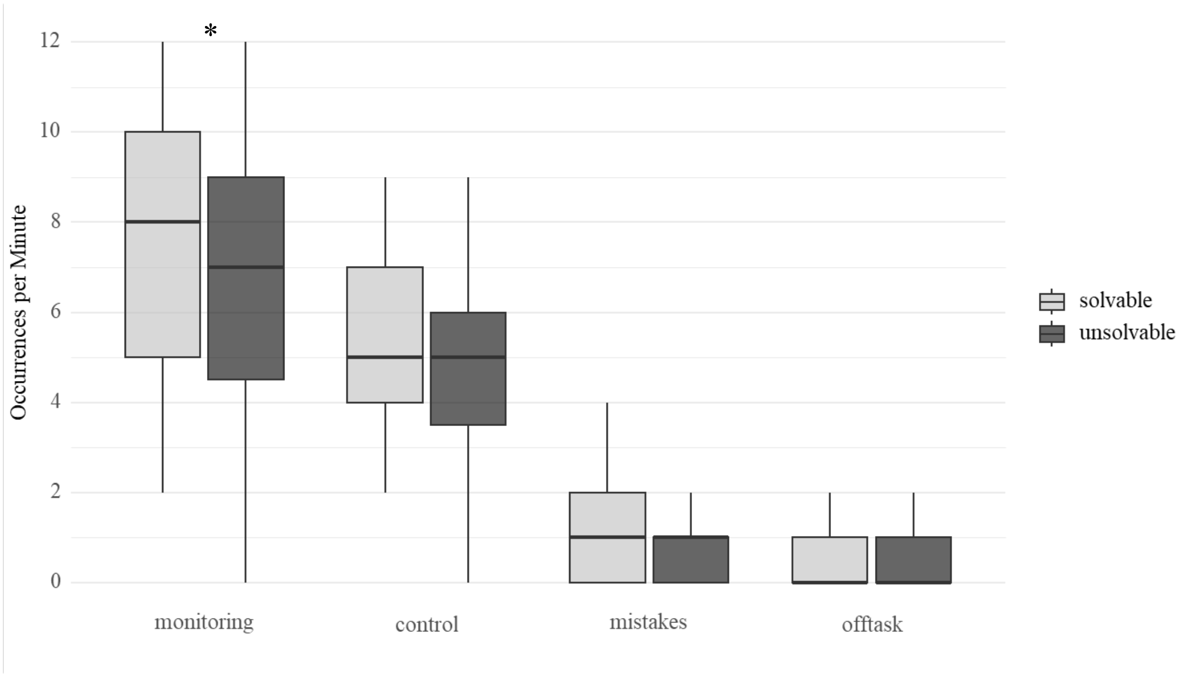
Boxplots for duration of metacognitive behavior for kindergarten 1 and kindergarten 2



Note. *Significant differences between kindergarten 1 and kindergarten 2.

Figure 3a

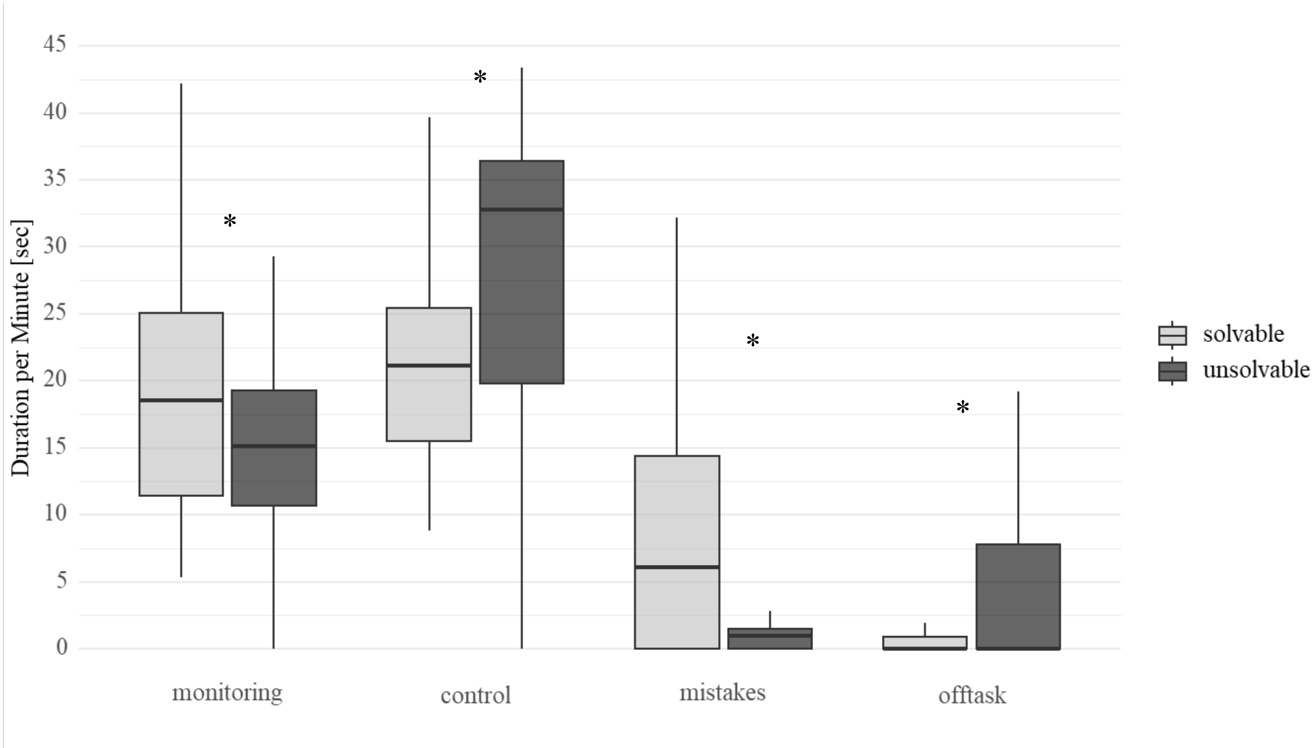
Boxplots for occurrences of metacognitive behavior in the solvable and unsolvable part.



Note. *Significant differences in metacognitive behavior between the solvable and unsolvable part.

Figure 3b

Boxplots for the duration of metacognitive behavior in the solvable and unsolvable part.



Note. *Significant differences in metacognitive behavior between the solvable and unsolvable part.

References

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

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)]*. 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.

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

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., 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

Nelson, T. O., and Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. *Psychology of Learning and Motivation - Advances in Research and Theory* 26, 125–173.

Oeri, N., and Roebers, C. M. (2021). Task persistence in kindergarten children: Disentangling age from schooling effects. *British Journal of Developmental Psychology* 39, 217–230. doi: 10.1111/BJDP.12358

R Core Team (2021). R: A language and environment for statistical computing [Computer software]. Available at: https://www.r-project.org/

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

Vygotsky, L. S. (1978). "Interaction between Learning and Development," in *Readings on the Development of Children*, eds. M. Gauvain and M. Cole (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/TABLES/3

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. doi: 10.1016/j.jecp.2022.105551