Metacognitive Monitoring and Help-Seeking Decisions on Mathematical Equivalence Problems

Lindsey J. Nelson & Emily R. Fyfe, Indiana University

Article accepted at *Metacognition and Learning,* August 19, 2019

Word count (primary text + abstract) = 192 + 7950

Author Note

Lindsey J. Nelson and Emily R. Fyfe, Department of Psychological and Brain Sciences, Indiana University. Support for this research was provided by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under Award Number T32HD007475. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank Nicholas Vest for help with data collection as well as the teachers and students at the participating schools.

Address correspondence to Lindsey J. Nelson, Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th Street, Bloomington IN 47405. Email: nelsonlj@iu.edu.

Abstract

Metacognition is central to children's cognitive development. However, there is conflicting evidence about children's ability to accurately monitor their performance and subsequently control their behavior. This is of particular interest for mathematics topics on which children exhibit persistent misconceptions—that is, when children's knowledge of a topic is inaccurate, yet resistant to change. This study investigated elementary school children's metacognitive regulation on mathematical equivalence problems (N = 52, ages 6.7 – 9.8 years), including their ability to accurately monitor their certainty and their ability to control their behavior by making strategic help-seeking decisions. Results revealed that children were exceedingly confident—even when their answers were incorrect—resulting in relatively low accurate monitoring scores. However, their help-seeking decisions were largely strategic—reflecting children's tendency to not ask for help when feeling confident—resulting in relatively high control scores. Additionally, individual differences in accurate monitoring and in strategic control were positively correlated with children's comprehensive knowledge of mathematical equivalence, and the correlation with accurate monitoring held up after controlling for baseline accuracy, certainty, and help-seeking. Collectively, these results suggest that children may face unique, but critically important, metacognitive challenges when solving mathematical equivalence problems.

Key Words: Metacognition, Monitoring, Help-Seeking, Mathematical Equivalence

Metacognitive Monitoring and Help-Seeking Decisions on Mathematical Equivalence Problems

Metacognition—the knowledge, monitoring, and regulation of cognition (Flavell, 1976)—plays a central role in cognitive development and is at the heart of self-regulated learning (e.g., Boekaerts, 1997). It has been and continues to be conceptualized in various ways, though there is consensus that it consists of two primary components: metacognitive knowledge and metacognitive regulation (e.g., Flavell, 1976; Nelson & Narens, 1990). The former is conceptualized as declarative and the latter as procedural; both are critical to learning, though procedural metacognitive regulation is thought to play a more direct role in ongoing learning processes (e.g., Roebers & Spiess, 2017). The focus of the current study is on procedural metacognitive regulation, including the ability to *monitor* and *control* one's thinking and actions. Specifically, we investigate elementary school children's procedural metacognitive regulation in the context of mathematics problem solving. Our goals are (1) to assess elementary school children's ability to accurately monitor their certainty (e.g., "how sure do you feel that you solved this problem correctly?"), (2) to assess their ability to control their actions strategically (e.g., "would you like to request help on this problem?"), and (3) to examine whether their monitoring and control predict performance. We investigated these questions in the context of mathematical equivalence, a fundamental concept in both arithmetic and algebra.

Monitoring and Control

In their original conceptualization of procedural metacognition, Nelson and Narens (1990) explained that monitoring consists of continuous observation and evaluation of one's thinking and behavior (*"I don't know how to solve this problem")*. Control, on the other hand, consists of the ability to make informed decisions at the task-level (*"Since I don't know how to solve this problem I am going to skip it"*). Importantly, Nelson and Narens (1990) suggested that

monitoring and control processes communicate continuously and dynamically, and it is this communication (e.g., Roebers & Spiess, 2017) that is of the greatest interest for learning.

Despite increasing research on metacognition across age groups and domains (e.g., Baten, Praet, & Desoete, 2017; Dignath & Buttner, 2008; Marulis, Palincsar, Berhenke, & Whitebread, 2016; Roebers, 2017; Whitebread & Coltman, 2010), research addressing metacognition in areas of problem solving—especially for young children—is limited. In particular, research is needed that addresses metacognition in problem-solving domains in which children exhibit misconceptions—when prior knowledge is not lacking, but is misaligned or incorrect. This was central to the current study; we investigated elementary school children's ability to accurately *monitor* their performance (i.e., feelings of certainty) on a series of mathematical equivalence problems as well as their ability to strategically *control* their decision-making when presented with the option of requesting a hint.

Monitoring and control have been assessed in many ways (see Gascione, Higgins, & Wall, 2017 for a review). In the current study, we view monitoring as children's ability to distinguish between solutions that are correct versus incorrect and we asses this monitoring by asking children to report feelings of certainty (e.g., De Clercq, Desoete, & Roeyers, 2000; Desoete, 2008). This, along with confidence judgments, judgements of learning, or mental effort ratings, is common practice across age groups and domains (e.g., Raaijmakers, Baars, Schaap, Paas, & van Gog, 2017; van Loon, de Bruin, van Gog, van Merrienboer, 2013).

Further, we conceptualize control as children's help-seeking decisions in response to feelings of certainty, as done by Coughlin, Hembacher, Lyons, and Ghetti (2015). Traditionally viewed as a deficiency in development, help-seeking has more recently been conceptualized as a strategic learning skill that children can employ when facing an impasse in their learning

(Karabenick & Berger, 2013; Newman, 2000). Help-seeking can represent a child's awareness of not knowing coupled with a plan to remedy that knowledge deficit by asking for help (Newman, 2000). The literature on children's strategic help-seeking decisions is sparse, especially when compared with other measures of metacognitive control such as allocation of study time (e.g., Metcalfe, 2009; Thiede, Anderson, & Therriault, 2003) or withdrawal of answers thought to be incorrect (e.g., Roebers & Spiess, 2017). This is surprising, given the fact that asking for help is a common practice in formal and informal learning environments (Jarvela, 2011); this study represents an important contribution to the literature on metacognitive control in children.

The current study was inspired by mixed findings on young children's accurate monitoring and control processes. For example, some prior studies suggest that these skills emerge as early as preschool (i.e., Coughlin et al., 2015; Lyons & Ghetti, 2013). These researchers found that 3-, 4-, and 5-year-olds accurately monitored their certainty during a perceptual identification task and responded strategically. That is, when presented with two degraded images and asked to select the one that represents a target object (e.g., a lemon), preschoolers recognized when they knew the answer, and if they did not, they tended to skip the question (Lyons & Ghetti, 2013) or ask for help (Coughlin et al., 2015). However, other work relays that preschoolers and older children tend to overestimate their abilities, resulting in poor monitoring and control (e.g., Rinne & Mazzocco, 2014; Roebers & Spiess, 2017; Shin, Bjorklund, & Beck, 2007). For example, Lipko, Dunlosky, and Merriman (2009) found that 4- and 5-year-olds persistently overestimated their abilities when asked to predict how many pictures they thought they could recall. A similar trend was observed with 8- and 9-year-olds in the context of a spelling task. When asked to rate their certainty after spelling a series of words, second-graders were largely overconfident (Roebers & Spiess, 2017). Furthermore, this

overconfidence led children to respond inefficiently when prompted to cross out words they thought they had not spelled correctly (Roebers & Spiess, 2017).

Thus, some studies show that children's metacognitive abilities are accurate and early emerging, and others consistently show poor calibration and overconfidence. Although some research suggests there is a gradual development in the ability to differentiate between correct and incorrect answers (e.g., Destan et al., 2014; Destan, Spiess, de Bruin, van Loon, Roebers, 2017; Lockl & Schneider, 2003; Lyons & Ghetti, 2013; Roderer & Roebers, 2010), this appears to function differently across tasks and domains (e.g., Baars, van Gog, de Bruin, & Paas, 2018; Erikson & Heit, 2015; Lucangeli & Cornoldi, 1997; Roebers, von der Linden, & Howie, 2007; Schneider, 1998; Thiede, Anderson, & Therriault, 2003). Thus, additional research is needed to assess monitoring and control for children across a variety of academic measures.

Metacognition and Mathematics

This research is particularly important in the domain of mathematics because (1) students need to learn a multitude of concepts and procedures and apply them appropriately (Lucangeli & Cornoldi, 1997), and (2) performance in mathematics in early childhood has been identified as a key predictor of later academic achievement (Duncan et al., 2007). Yet, metacognitive research in mathematics is sparse in comparison to other domains (see Schneider & Artelt, 2010 for a review). As in the broader literature, the existing evidence is mixed.

Many studies have showcased children's metacognitive skillfulness in mathematics. For example, Vo, Kornell, Pouget, and Cantlon (2014) found that children as young as 5 years of age were able to identify which of two dot clusters was larger, and their confidence judgments on these items predicted school-based mathematics knowledge. Similarly, Desoete and colleagues found that third-graders' certainty ratings on a series of arithmetic problems were relatively

accurate and were correlated with measures of procedural calculation and number fact retrieval (Desoete, Roeyers, & Huylebroek, 2006). Further, they were able to use these metacognitive measures to distinguish between students of different problem-solving abilities, including those with learning disabilities (Desoete, Roeyers, & Buysse, 2001; Desoete, et al., 2006).

However, many other studies have highlighted children's metacognitive deficits in mathematics. A recent study found that fifth- and sixth-grade students showed poor monitoring on mathematics word problems and were largely overconfident (García, Rodríguez, González-Castro, González-Pienda, & Torrance, 2016). Similarly, Baars, van Gog, de Bruin, and Paas (2014) found that fifth-graders' restudy selections did not align with their certainty monitoring on mathematics word problems. And most recently, Baars and colleagues (2018) revealed that third graders could distinguish between various difficulty levels of arithmetic problems (i.e., their judgments of learning were lower for more difficult items), but their performance monitoring was poor when assessed with a slight delay.

These mixed findings suggest that metacognitive skills are sensitive to domain-specific and task-specific demands. As such, it is imperative to consider the nature of the problems children are asked to solve as well as their prior knowledge of those problems. This is particularly true for domains in which children exhibit misconceptions—when prior knowledge is not lacking but is misaligned or incorrect. Several mathematics concepts pose unique challenges for children in that their knowledge of the concepts is inaccurate, yet resistant to change. Decades of research have identified symbolic mathematical equivalence as one such concept (see McNeil, 2014 for a review). Mathematical equivalence is the understanding that both sides of the equal sign are interchangeable and represent the same amount.

Mathematical Equivalence and Change-Resistance

There are several key reasons to focus on children's metacognition in the context of mathematical equivalence. For example, mathematical equivalence is considered a fundamental concept for both arithmetic and algebra (e.g., Carpenter, Franke, & Levi, 2003; Kieran, 1981), and it is included in the Common Core State Standards beginning in first grade (NGA Center & CCSSO, 2010). Also, knowledge of mathematical equivalence is related to mathematics achievement. For example, knowledge of the equal sign in second grade has been found to predict global mathematics achievement in third grade (McNeil, Hornburg, Devlin, Carrazza, & McKeever, 2017) as well as algebra knowledge in fourth grade (Matthews & Fuchs, 2018).

Most critically, elementary school children often exhibit misconceptions of mathematical equivalence (e.g., Byrd, McNeil, Chesney, & Matthews, 2015; Powell & Fuchs, 2010; Rittle-Johnson, 2006), that sometimes persist into middle school, high school, and beyond (e.g., Fyfe, Matthews, & Amsel, 2017; Kieran, 1981; McNeil et al., 2010). These misconceptions represent an ideal way to investigate metacognition in a context in which children have prior knowledge that is misaligned or incorrect. For example, when asked to solve a mathematical equivalence problem with operations on both sides of the equal sign (e.g., 5 + 9 = 4 + __), approximately 70-80% of elementary school children answer incorrectly (e.g., Fyfe & Rittle-Johnson, 2016; McNeil & Alibali, 2005). Most use simple arithmetic strategies and add up all the numbers (answering 18) or add up all the numbers before the equal sign (answering 14). Similarly, when asked to define the equal sign, many children provide operational definitions such as "get the answer" or "find the total," rather than relational definitions that indicate sameness (e.g., Byrd et al., 2015).

The leading theory in this area – termed the change-resistance account – suggests that children's misconceptions arise from prior learning experiences (McNeil, 2014). The argument is

that children pick up on routine patterns in their arithmetic practice; for example, noticing that the equal sign is almost always at the end of the problem ($a + b = c$, see Powell, 2012). Then, difficulties occur when children overgeneralize their knowledge of these patterns to mathematical equivalence problems (e.g., using the "add-all" strategy that works well on a standard arithmetic problem on an equivalence problem). The theory suggests that these misconceptions are resistant to change, in part, because children are confident that their strategies are correct (see McNeil & Alibali, 2005). Indeed, it is not just the lack of a correct understanding that matters, but the presence of arithmetic-specific misconceptions that hinders learning (McNeil et al., 2017).

Surprisingly, despite decades of research on children's understanding of mathematical equivalence, empirical investigations of children's metacognition in this area are lacking. Although one study included children's certainty ratings of their solutions to mathematical equivalence problems, the authors did not report on these ratings or their alignment with children's accuracy (McNeil & Alibali, 2005). Thus, it remains unclear the extent to which children can accurately monitor their understanding of mathematical equivalence and whether individual differences in this monitoring are associated with learning and performance. Research in this area is needed to provide empirical evidence on the well-developed theory and to more systematically consider the role of prior knowledge in children's metacognition.

The Current Study

In the current study, we asked the following research questions: (1) How well do elementary school children monitor their certainty while solving mathematical equivalence problems? (2) How well do they control their actions by making strategic help-seeking decisions? (3) Do individual differences in accurate monitoring and strategic control predict

performance? We answered these questions by obtaining independent measures of children's prior knowledge, monitoring, control, and performance, so that the variables would not directly influence one another. This study helps fill critical gaps in the current literature; it provides evidence on young children's metacognitive skills in a mathematics topic with consistent and well-documented misconceptions, it uses an understudied measure of metacognitive control, and it systematically considers the role of domain-specific prior knowledge.

Method

Participants

Participants were recruited from five classrooms in four schools (one Montessori school, one public school, and two small private schools) in a mid-sized city in the Midwest. Our aim was to recruit 50-60 children, as a power analysis indicated that our primary analyses predicting performance had approximately 80.3% power to detect a small effect with a sample size of 55 children. Original participants were 54 children with parent consent from first-grade, second-grade, and third-grade classrooms. Two children were excluded because they did not complete all components of the study. The final sample included 52 children (M = 8.1 years, SD = 0.8; 48% female): 14 first graders (M = 7.2 years, SD = 0.4), 22 second graders (M = 8.0 years, SD = 0.4), and 16 third graders (M = 9.1 years, SD = 0.4). Of the 52 participants, 24 were Caucasian, 1 was Hispanic, 1 was Asian, 6 were of mixed ethnic background, and 20 did not report.

Materials

Baseline Assessment. The baseline assessment of prior knowledge was a measure used in previous research that included three mathematical equivalence tasks (McNeil, Fyfe, Petersen, Dunwiddie, & Brletic-Shipley, 2011). For *problem-solving*, children solved two mathematical equivalence problems to assess their knowledge of correct strategies. For *problem-encoding*,

children reconstructed four mathematical equivalence problems from memory to assess their mental representation of the problem structure. Children were shown a problem for five seconds and then asked to write exactly what they saw. For *equal-sign definition*, children provided a written definition of the equal sign to assess their explicit relational knowledge of the equality symbol. Children were assigned a baseline score out of seven (Cronbach's $\alpha = .79$).

Target Problems. To assess monitoring and control, we created two sets of five target mathematical equivalence problems (see Table 1). We opted to use a modest number of items given that performance on these items tends to be bimodal, with most children in this age range doing poorly, some children near mastery, and only a handful of children in the middle (e.g., Fyfe, Evans, et al., 2017; McNeil et al., 2012). Each problem in the first set was matched to a corresponding problem in the second set on three criteria: (1) structurally equivalent, (2) addends within plus or minus one, and (3) solutions within plus or minus two. These target problems were presented in two matched sessions—a standard session and a help session. The purpose of the standard session was to assess children's accurate monitoring. That is, to assess whether children's certainty ratings matched their performance (e.g., feeling uncertain on a problem solved incorrectly). The purpose of the help session was to assess children's strategic decision-making. That is, to assess whether children's help-seeking decisions in the help session matched their certainty ratings in the standard session (e.g., asking for help on problem A in the help session in response to feeling uncertain on problem A in the standard session).

We deliberately assessed children's monitoring and control using two separate sessions so as not to influence children's monitoring and control behaviors in the process. If assessed concurrently, certainty ratings could have affected children's propensity to seek help and vice-versa. This design was taken from a previous study (Coughlin et al., 2015) and "reflects the idea

that participants will experience consistent levels of uncertainty on individual [items] across sessions" (Coughlin et al., 2015, p. 959). We counterbalanced the Set 1 problems and Set 2 problems between sessions and presented the problems in a random order within each session. The standard session and the help session were both administered on a computer; this allowed us to more easily randomize the order of item presentation.

Metacognitive Certainty Ratings. In the standard session, children rated their certainty on each of the five mathematical equivalence problems using a 4-point scale ranging from "I know I got this problem right" to "I know I got this problem wrong" (see Figure 1). Children were assigned an objective *certainty* score on each item ranging from 1 to 4 with higher scores indicating higher certainty of correctness. Each point on the scale was represented by a traffic light and a statement, which was adapted directly from De Clercq, Desoete, and Roeyers' (2000) Evaluation and Prediction Assessment (i.e., EPA2000). The EPA2000 measures children's off-line metacognition on domain-general mathematics problems and it has been validated for use with primary school children (e.g., De Clercq et al., 2000; Desoete, 2008).

Hints. In the help session, children could request hints on each of the five mathematical equivalence problems by pressing a "Hint" button on the screen. The hints were determined ahead of time, were specific to each problem, and were based on instructional scripts used in previous research (e.g., Fyfe & Rittle-Johnson, 2016). The hints did not provide the correct answer but provided guidance toward correct strategies. For example, the hint provided for the problem 8 = 3 + __ was: *This problem is eight is equal to three plus blank (point to each number/symbol as you read it out loud). Ready for a hint? This equal sign (point to equal sign) means both sides of the equal sign have to be the same amount. Can you use that hint to try to solve this one again? If you're still not sure, that's okay. Just give it your best guess*. Children

were assigned a help-seeking score from 0 to 5 based on the number of problems on which they asked for help.

Comprehensive Performance Assessment. This assessment was a broader measure of mathematical equivalence knowledge adapted from Rittle-Johnson, Matthews, Taylor, and McEldoon (2011). It consisted of six *problem-solving* items that tapped children's use of correct strategies to solve mathematical equivalence problems, as well as ten *conceptual* items that tapped children's knowledge of the equal sign (e.g., which definition of the equal sign is the best?) and problem structure (e.g., is 3 = 3 true or false?). Children were assigned a performance score out of 16 (Cronbach's $\alpha = .93$).

Coding

We assessed children's accuracy on three different measures: the baseline assessment, the standard session, and the comprehensive performance assessment. Most of the items on these measures required students to solve open-ended mathematical equivalence problems (e.g., 5 + 9 = 4 + __). To code children's accuracy on these *problem-solving* items, we coded children's numerical solutions. Solutions within one of the correct answer were coded as correct to allow for a small margin of arithmetic error. On the comprehensive performance assessment, children provided both a numerical solution and a verbal report; thus, responses were coded based on the solutions and the verbal reports. For example, for the problem 3 + 4 = __ + 5, a report of "three plus four is seven and two plus five is seven" was coded as correct, whereas a report of "I added the numbers 3, 4, and 5 and put 12 in the blank" was coded as incorrect. The first author coded all items and the second author coded 30% of the responses. Interrater agreement on correct responses across all problem-solving items was high (98%). Additionally, several items on the baseline assessment and comprehensive performance assessment were conceptual in nature. For

problem-encoding items, we coded children's accuracy in recalling the problem structure (i.e., the placement of the addends, operations, and equal sign). For open-ended *equal-sign items*, we coded whether children's responses were operational (e.g., "find the answer," or "the sum"), relational (e.g., "same as," or "same amount"), or both; relational responses were coded as correct. Interrater agreement was high ranging from 93% to 98% depending on the item type.

In addition to measuring accuracy, we also coded children's monitoring and control. Based on a measure used in previous research (see Baars, van Gog, de Bruin, & Paas, 2018), we calculated an accurate monitoring score that reflected the calibration between children's certainty ratings on a four-point scale and their standard session accuracy (see Table 2). The monitoring score on each item varied between 0 and 1 (i.e., 0, .33, .67, 1) based on each combination of certainty rating (1 to 4) and accuracy (right or wrong). As shown in Table 2, lower certainty ratings combined with a wrong answer resulted in higher monitoring scores, and lower certainty ratings combined with a right answer resulted in lower monitoring scores. Conversely, higher certainty ratings combined with a wrong answer resulted in lower monitoring scores, and higher certainty ratings combined with a right answer resulted in higher monitoring scores. We averaged children's monitoring scores across the five items in the standard session.

In a similar way, we calculated a strategic control score that reflected the calibration between children's help-seeking decisions and their standard session certainty (see Table 2). The control score on each item varied between 0 and 1 (i.e., 0, .33, .67, 1) based on each combination of certainty rating (1 to 4) and help-seeking (yes or no). As shown in Table 2, lower certainty ratings combined with a decision to ask for help resulted in higher control scores, and lower certainty ratings combined with a decision to not ask for help resulted in lower control scores. Conversely, higher certainty ratings combined with a decision to ask for help resulted in lower

control scores, and higher certainty ratings combined with a decision to not ask for help resulted in higher control scores. We averaged children's control scores across the five items.

Procedure

Children participated in two sessions (see Figure 2). On day 1, they completed the baseline assessment, which was conducted in a classroom setting and lasted 10-20 minutes. On day 2—held within one week of day 1—an experimenter met with each child for a single session that lasted 30-40 minutes. During this one-on-one session, children completed the standard session, the help session, and the performance assessment. The baseline assessment and performance assessment were administered with pencil and paper. The standard session and help session were administered on the computer using Qualtrics, an online survey platform.

Standard session. During the standard session, children first completed a simple practice problem (*Which number is bigger? 29 or 31*) to get acquainted with the computer as well as the metacognitive certainty rating scale (Figure 1). Children were prompted by the researcher to type their response and press the "next" button to continue. After responding to the practice question (i.e., typing 29 or 31), they were asked: *How sure do you feel that you solved this problem correctly?* They were instructed to click on the traffic light that matched their certainty. After the practice problem, children solved five mathematical equivalence problems, presented one at a time. On each one, they were asked by the experimenter to *figure out the number that should go in the box to make the problem true*. After typing a response into the box, they proceeded to rate their metacognitive certainty (Figure 1). Children completed this procedure for all five problems and did not receive any feedback.

Help session. After completing the standard session, children were given an opportunity to take a short break and then complete the help session. The experimenter said: *this session will*

be similar to the last one except this time you will have the option to ask for help. If you would like to ask for help, you can press a "Hint" button on the screen. Children then solved five mathematical equivalence problems, presented one at a time. On each one, they could enter an answer or press the hint button. If a child pressed the hint button, the experimenter said: *I've marked this problem so we can come back to it. Can you try the next one? Hit the "next" button.* The hint was delayed to avoid providing children with additional instruction during the session, which could affect the propensity to ask for help on later problems. Thus, pressing the hint button was akin to skipping the problem until the end of the session, at which point the hint was provided by the experimenter. After all five problems had been presented, the experimenter provided hints for any problem on which they pressed the hint button and were given another chance to submit a response. We were interested in assessing children's strategic help-seeking decisions rather than examining how the provision of hints facilitates later performance. After the help session, children completed the comprehensive performance assessment.

Results

Preliminary Analyses

First, we confirmed that set order (i.e., solving the Set 1 problems vs. the Set 2 problems in the standard session) did not influence performance. There was not a significant difference in problem-solving accuracy, $F(1, 49) = 0.39$, $p = .53$, $\eta_p^2 = .01$, or in certainty, $F(1, 49) = 2.28$, $p = .14$, $\eta_p^2 = .04$, in the standard session between Set 1 and Set 2 when we controlled for baseline scores; this variable was excluded from subsequent analyses. Second, we explored the roles of gender and prior knowledge. Preliminary analyses revealed that gender was not significantly related to our primary variables of interest, including certainty, help-seeking, and performance scores; this variable was excluded from subsequent analyses. However, preliminary analyses

revealed that prior knowledge at baseline was relevant. Overall, children exhibited low knowledge of mathematical equivalence at baseline (M = 2.35 out of 7, SD 2.09), and nearly 40% of children earned a score of 0 or 1, which is consistent with low levels of performance reported in prior work in this age range (e.g., McNeil et al., 2012). In subsequent analyses, we examined variables of interest as a function of baseline performance. Specifically, we split children into a low knowledge group (scored 0 or 1 at baseline, n = 19) and a high-knowledge group (scored 2 or higher at baseline, n = 33). This allowed us to better examine the relation between prior knowledge and metacognition.

Research Question 1: How well do elementary school children monitor their certainty while solving mathematical equivalence problems?

We addressed this question by examining the relation between children's accuracy and certainty during the standard session. On average, children solved 2.35 problems correctly (out of 5, [47%] SD = 2.16), and a full 52% of children solved 0 or 1 problems correctly. Not surprisingly, accuracy differed significantly by prior knowledge, $F(1, 50) = 24.20$, $p < .001$, $\eta_p^2 = .33$, such that children with high knowledge at baseline solved more problems correctly (M = 3.27, SD = 2.02) than children with low knowledge at baseline (M = 0.74, SD = 1.28).

Though performance was low to moderate, children's average certainty across the five problems was quite high (M = 3.23 on a 4-point scale, SD = 0.54), meaning they were mostly certain they were correct. Indeed, only six children (out of 52) had an average certainty rating of 2.50 or below, indicating general *un*certainty in the correctness of their responses. Further, when we looked at how often children selected each of the four responses on the certainty scale, children *knew* or *thought* they were correct on 84% of problems (they *knew* on 41% and *thought* on 43%) and children rarely *knew* or *thought* that they were wrong (they *knew* on 2% and

thought on 14%). Certainty ratings were relatively similar across items (ranging from 3.1 to 3.4) and certainty ratings did not differ significantly as a function of prior knowledge, $F(1, 50) = 1.93$, $p = .17$, $\eta_p^2 = .03$; high-knowledge children had an average certainty rating of 3.30 ($SD = 0.50$) and low-knowledge children had an average certainty rating of 3.10 ($SD = 0.59$). Thus, even when knowledge of mathematical equivalence was low, certainty in correctness was high.

To examine the extent to which children *accurately* monitored their certainty, we calculated an accurate monitoring score (ranging from 0 to 1) that reflected the calibration between children's certainty ratings on a four-point scale and their standard session accuracy (see Table 2). Children's average monitoring score was 0.55 ($SD = 0.29$), and scores spanned the full range from 0 to 1. Accurate monitoring scores were higher on Problem A ($M = 0.63$) relative to Problems B through E (Ms = 0.51-.54), likely because of the simpler structure of Problem A (see Table 1). Accurate monitoring scores differed significantly by prior knowledge, $F(1, 50) = 25.44$, $p < .001$, $\eta_p^2 = .34$; children with high knowledge at baseline had higher monitoring scores ($M = 0.67$, $SD = 0.25$) than children with low knowledge at baseline ($M = 0.33$, $SD = 0.21$).

We descriptively examined monitoring separately for problems solved correctly and problems solved incorrectly. Table 3 presents children's certainty ratings along the four-point scale aligned with their accuracy. As shown in the table, children were equally likely to select certainty ratings of 3 on correct and incorrect problems (21% vs. 22%) and were equally likely to select certainty ratings of 4 on correct and incorrect problems (23% vs. 18%), thus showing little calibration. If we descriptively dichotomize the certainty ratings into two categories (green lights versus red lights), then children "matched" their certainty with their accuracy 57% of the time (i.e., solved it correctly and selected a certainty rating of 3 or 4, or solved it incorrectly and selected a certainty rating of 1 or 2). These dichotomized match rates were strikingly higher for

items solved correctly than incorrectly (94% vs. 25%). Indeed, when a mismatch occurred, it was almost always (93% of the time) the result of children feeling certain that their answers were correct when in fact they were not.

Research Question 2: How well do elementary school children control their actions by making strategic help-seeking decisions?

We addressed this question by examining the relation between children's help-seeking decisions during the help session and their certainty ratings during the standard session. On average, children asked for help on 0.85 out of 5 problems (SD = 1.11); this is only 17% of problems. Due to the scarcity of help-seeking, we also quantified help-seeking as a dichotomous variable: children who asked for help on at least one problem (n = 27, 52% of children) and children who did not ask for help at all (n = 25, 48% of children). Rates of help-seeking were relatively similar across items (ranging from 13% to 17%), and help-seeking did not differ significantly as a function of prior knowledge, χ^2 (1, N = 52) = 1.51, p = .22; the proportion of children seeking help at least once was 46% for children with high knowledge at baseline and 60% for children with low knowledge at baseline.

To examine the extent to which children made *strategic* help-seeking decisions, we calculated a strategic control score (ranging from 0 to 1) that reflected the calibration between children's standard session certainty ratings on a four-point scale and their help-seeking decisions during the help session (see Table 2). Children's average strategic control score was 0.71 (SD = 0.17), which was significantly higher than their average accurate monitoring score (M = 0.55), t(51) = 4.29, p < .001. Scores did not span the full range, but were quite high and went from 0.40 to 1.00. Strategic control scores were higher on Problem A (M = 0.80) relative to Problems B through E (Ms = 0.67-.73). Control scores did not differ significantly as a function of

prior knowledge, $F(1, 50) = 1.84$, $p = .18$, $\eta_p^2 = .04$; children with high knowledge at baseline had an average control score of 0.73 ($SD = 0.17$) and children with low knowledge at baseline had an average control score of 0.68 ($SD = 0.18$).

We descriptively examined control separately for problems on which they asked for help and problems for which they did not ask for help. Table 4 presents children's certainty ratings along the four-point scale aligned with their bids for help. If we descriptively dichotomize the certainty ratings into two categories (green lights versus red lights), then children made strategic help-seeking decisions 81% of the time (i.e., selected a certainty rating of 1 or 2 and asked for help, or selected a certainty rating of 3 or 4 and did not ask for help). These strategic decisions were substantially higher for items on which children did not ask for help versus items on which they did ask for help (89% vs. 41%). This is because when children did ask for help, they were equally likely to do so when they felt certain they were incorrect and when they felt certain they were correct (7% vs. 10%). Further, when children were uncertain, they were similarly likely to ask for help or not ask for help (8% vs. 9%), suggesting room for improvement.

Research Question 3: Do individual differences in accurate monitoring and strategic control predict performance?

We addressed this question by examining whether children's accurate certainty monitoring and strategic help-seeking decisions predicted their mathematical equivalence knowledge. On average, children answered 6.54 problems correctly out of 16 ($SD = 5.16$) on the comprehensive performance assessment. The raw bivariate correlations among accurate monitoring scores, strategic control scores, and comprehensive performance scores were all positive and significant (between monitoring scores and control scores, $r(50) = .33$, $p = .02$; between monitoring scores and performance scores, $r(50) = .82$, $p < .001$, between control scores

and performance scores, $r(50) = .46$, $p = .001$). Thus, both metacognitive monitoring and control were positively related to overall mathematical equivalence knowledge.

To better assess the unique relations between aspects of metacognition and overall performance, we ran two stepwise linear regression models predicting comprehensive performance scores. Model 1 focused on the predictive relation of monitoring and Model 2 focused on the predictive relation of control. Each model had two steps. In the first step, we entered baseline group (low versus high knowledge), standard session accuracy (out of 5), standard session certainty ratings (out of 4), and help-seeking (ever asked for help versus never asked for help). This step accounted for prior knowledge as well as overall levels of accuracy, certainty, and help-seeking. In the second step, we entered accurate monitoring scores (in Model 1) or strategic control scores (in Model 2) to examine whether these variables explained unique variance in the target outcome measure.

Table 5 presents the full results from each model. Model 1 was significant and the final model explained 84% of the variance in overall performance scores. Of note, including accurate monitoring scores in the second step resulted in significant model improvement, $F(1, 46) = 4.86$, $p = .03$, as accurate monitoring scores were a unique, significant predictor of performance scores, $\beta = .27$, $p = .03$, over and above the other predictors. Model 2 was also significant and the final model explained 82% of the variance in overall performance scores. However, including strategic control scores in the second step did not result in model improvement, $F(1, 46) = 1.10$, $p = .30$, as strategic control scores were not a significant predictor of performance scores, $\beta = .13$, $p = .30$, over and above the other predictors. Rather, the significant predictors in Model 2 were baseline knowledge, $\beta = .29$, $p < .001$, and standard session accuracy, $\beta = .65$, $p < .001$.

Discussion

In the current study, we investigated 52 elementary school children's ability to accurately *monitor* their performance on a series of mathematical equivalence problems as well as their propensity to *control* their actions strategically when presented with the option of requesting help. Overall, elementary school children struggled to solve mathematical equivalence problems correctly, yet they were often certain that they had. Thus, their monitoring scores were low. Children's help-seeking decisions were largely strategic; overall control scores were higher than monitoring scores, although this was primarily due to children not seeking help in response to feelings of certainty. Lastly, individual differences in children's monitoring and control were both positively correlated with their comprehensive performance, and the relation with monitoring remained significant after accounting for prior knowledge, certainty, and help-seeking. We outline the contributions of this work and discuss limitations and future directions.

First, the current results contribute to a growing body of literature investigating metacognitive skills in childhood. Consistent with previous research, we found that children were only modestly able to monitor their performance. Further, children struggled more with accurate monitoring of *incorrect* responses (e.g., Destan et al., 2017; Roebers & Spiess, 2017). Indeed, when children solved a problem incorrectly, they were much more likely to feel certain they were correct than incorrect; children used the upper end of the certainty scale on over 80% of trials. Although this could be a function of our sample size and scale, previous work suggests it is an accurate characterization of children's confidence. For example, in Roebers and Spiess, (2017), 119 7-year-olds used a 7-point likert scale on a 15-item spelling task and average certainty ratings were 6.4 for correct trials and 6.2 on incorrect trials, indicating children were primarily using the upper-end of the scale. This suggests that elementary school children's certainty and knowledge are not well calibrated; rather, they tend to overestimate their abilities.

Second, the present study extends previous research by examining metacognition in a domain with persistent, entrenched misconceptions. Mathematical equivalence is a foundational topic that relates to future mathematics performance (e.g., Matthews & Fuchs, 2018). However, children exhibit persistent misconceptions that can last well beyond elementary school (e.g., Fyfe et al., 2017). Children's accurate monitoring of equivalence knowledge appears to be at the lower bound relative to other topics. For example, the average accurate monitoring score in the current study was 55%; previous studies have reported accurate monitoring scores ranging from 56-72% for third-graders solving arithmetic problems (Baars et al., 2018; Desoete et al., 2001; Desoete et al., 2006; Desoete, 2008), and from 80%-91% of problems for second- and third-graders solving place value problems (Garrett, Mazzocco, & Baker, 2006). In contrast, rates of overconfidence on mathematical equivalence appear to be at the upper bound. Garrett et al. (2006) reported that second-graders were overconfident 14% of the time (certain on incorrect trials) on place value problems, compared to 40% of the time in the current study (see Table 3).

Children in the current study may have struggled to monitor their performance because of specific misconceptions. For example, when solving equivalence problems, children often rely on basic "add-all" strategies that are effective for simple arithmetic problems. These arithmetic problems share many surface features with equivalence problems (e.g., numbers, plus sign, equal sign), but have critical structural differences. Thus, children's familiarity with the surface features may lead them to think they know how to solve equivalence problems correctly. Further, most children who employ basic "add-all" strategies can add the numbers correctly, potentially contributing to their overconfidence in a strategy that leads to the wrong answer. This is consistent with the change-resistance account of mathematical equivalence, which suggests that children are relatively certain in incorrect strategies (McNeil & Alibali, 2005), though this is the

first empirical report to quantify it. Additionally, there are other possible explanations related to the study design. For example, children's overconfidence could have resulted from low feelings of mental effort because of fluency with the task (Baars et al., 2018; Koriat, Nussinson, & Ackerman, 2014), the absence of external standards (e.g., a worked example) (Raaijmakers et al., 2019), or the social pressure of having a novel experimenter present during problem solving.

Third, the present findings provide novel insights into the role of strategic help-seeking decisions as a mechanism of metacognitive control. Strategic help-seeking is adaptive and well-planned; it is a sign that children recognize that they are uncertain and need help to move forward. In this study, help-seeking was extremely rare. We acknowledge that the low rates of help-seeking may represent floor effects thus limiting our interpretation. However, these rates may also be a meaningful characterization of children's behavior on this task. In fact, other metrics of control used in this age range have shown low frequency rates as well. For example, Roebers and Spiess (2017) had 7-year-olds spell words, rate whether they thought they were correct, and then cross out words they thought they had spelled incorrectly. The crossing-out was their measure of metacognitive control and children rarely crossed any words out; of all the words spelled incorrectly, only 10% were crossed out.

More importantly, children may have only asked for help on rare occasions because they were often feeling certain they were correct. Indeed, this behavior contributed to high strategic control scores, which contrasts with children's low monitoring scores described above. Certainly, children's help-seeking decisions were far from perfect. For example, children were equally likely to ask for help when they felt certain or uncertain. However, there may be multiple factors other than confidence that prevented them from asking help. The novel experimenter, the use of a hint button on a computer screen, or the fact that hints were provided after a delay may have

suppressed help-seeking (although, the fact that rates of help-seeking were consistent across problems and did not decline over time suggests the last factor was a non-issue). Despite these alternative possibilities, the current results suggest children's control behavior was relatively well-attuned and largely due to children's decisions *not* to seek help in response to feelings of certainty. Further, the evidence indicates that children's monitoring and control processes are indeed connected within this domain, even though their monitoring was faulty.

Finally, the positive correlations between monitoring, control, and performance as well as the unique predictive power of monitoring on performance—above and beyond prior knowledge, certainty ratings, and help-seeking—contribute to the growing body of literature that shows that the predictive strength of metacognition is robust across many age groups and academic domains (e.g., Dignath & Buttner, 2008; Wang, Haertel, & Walberg, 1990). The current study extends this work to include domains in which children exhibit persistent misconceptions and potentially suggests that the monitoring aspect of metacognitive regulation may be more predictive than the control aspect, at least with the type of measures used in this study. Further, given that monitoring and control appear to be well connected, it is also possible that improvements in metacognitive monitoring would lead to improvements in metacognitive control, though future research is needed. The predictive power of prior knowledge, both for accurate monitoring and comprehensive performance, is also consistent with prior work (e.g., van Loon et al., 2013).

Currently, interventions aimed at improving mathematical equivalence understanding have focused on children's *content knowledge* of arithmetic strategies and equivalence concepts (e.g., Blanton et al., 2015; McNeil, Fyfe, & Dunwiddie, 2015); however, the current research suggests that exploring ways to focus on children's metacognitive regulation in this topic may be a fruitful endeavor. This is consistent with the literature that illustrates metacognitive

interventions can be beneficial for mathematics achievement more broadly (e.g., Baten et al., 2017; Mevarech & Fridkin, 2006; Raaijmakers et al., 2019; and see Schneider & Artelt, 2010 for a review). For example, researchers have found that elementary school children who received contextualized, interactive metacognitive training (e.g., direct instruction of metacognitive strategies, metacognitive regulation checklists, and metacognitive questioning from teachers) made significant gains on measures of metacognitive knowledge and regulation as well as mathematics performance (Ozsoy & Ataman, 2009; Pennequin, Sorel, Nanty, & Fontaine, 2010). Additionally, Mevarech and Kramarski (2003) found that providing eighth-graders with metacognitive training—largely consisting of guided self-questioning—was more effective than providing worked-out examples. Further, positive effects of metacognitive scaffolding have been found in children as young as 3, 4, and 5 years of age (Whitebread & Coltman, 2010; see also Baten et al., 2017) and children with learning disabilities (e.g., Desoete et al., 2003; Garrett et al., 2006; Kramarski & Mevarech, 2003), suggesting these interventions may be broadly applicable.

Given that metacognitive skills are sensitive to tasks and domains, interventions that include domain-specific metacognitive training may be beneficial. For topics on which children experience misconceptions, it may be important to make children aware of the accuracy and relevance of their prior knowledge as well as common errors (e.g., the “add-all” error) when intervening on mathematical equivalence knowledge. Similarly, future work might investigate the effects of feedback as a way of intervention. Recently, Raaijmakers and colleagues (2019) investigated the effect of self-assessment feedback on secondary students’ problem solving and found that feedback on performance may not be as helpful as feedback on self-assessment accuracy. Thus, future research might investigate different ways of providing feedback to

address children's overconfidence of mathematical equivalence, specifically comparing the benefits of feedback on accuracy versus metacognitive abilities.

Despite the stated contributions of the current research, several methodological limitations suggest directions for future research. Importantly, our study had a limited number of participants and the sample was spread across three grade levels; future work should include a sample size large enough to examine developmental trends, both cross-sectionally and longitudinally. Indeed, research indicates that metacognitive abilities improve in the age range studied here (e.g., Destan et al., 2017). Additionally, our measures of children's metacognitive regulation were limited to their feelings of certainty after problem solving and their requests for help on one type of mathematics problem. Moreover, children were near ceiling performance on their certainty ratings and near floor performance on their help seeking. Future work might include different tasks at various levels of complexity as well as more fine-tuned measures of metacognitive abilities to provide nuanced insights into the range of children's responses.

Further, researchers should consider measuring metacognitive abilities at multiple time points during the learning process. Ackerman and colleagues (2015) recently coined the term "meta-reasoning" to highlight the metacognitive processes that are specific to reasoning and problem-solving, as opposed to memory and retrieval. Problem-solving is complex and usually takes place over an extended period of time—in comparison to fact retrieval, for example—and therefore requires more complex metacognitive monitoring and control processes. A meta-reasoning lens specifically calls for analysis of metacognitive abilities prior to, during, and after complex problem solving. Finally, our results are correlational; thus, we cannot be certain whether more accurate monitoring caused children to perform better on the comprehensive

performance assessment. Intervention work focused on the malleability of accurate monitoring and its relation to performance is needed to determine causal relations.

In sum, this study is the first to quantify children's metacognitive monitoring and control in the domain of mathematical equivalence. Results revealed that children struggled to solve problems correctly, their ability to monitor their certainty was often faulty, and help-seeking decisions were largely strategic, though had clear room for improvement. Importantly, individual differences in children's accurate monitoring predicted performance suggesting a critical relation between metacognition and mathematical equivalence understanding.

Compliance with Ethical Standards

Acknowledgements. The authors thank Nicholas Vest for help with data collection as well as the teachers and students at the participating schools.

Funding. Support for this research was provided by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under Award Number T32HD007475. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Informed Consent. Informed consent was obtained from participants' parents and all children assented to participate.

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

References

Baars, M., Van Gog, T., de Bruin, A., & Paas, F. (2018). Accuracy of primary school children's immediate and delayed judgments of learning about problem-solving tasks. *Studies in Educational Evaluation*, *58*, 51-59. doi: 10.1016/j.stueduc.2018.05.010

Baars, M., Van Gog, T., de Bruin, A., & Paas, F. (2014). Effects of problem solving after worked example study on primary school children's monitoring accuracy. *Applied Cognitive Psychology*, *28*(3), 382-391. doi: 10.1002/acp.3008

Baten, E., Praet, M., & Desoete, A. (2017). The relevance and efficacy of metacognition for instructional design in the domain of mathematics. *ZDM*, *49*, 613-623. doi: 10.1007/s11858-017-0851-y

Behr, M., Erlwanger, S., & Nichols, E. (1980). How children view the equal sign. *Mathematics Teaching, 92*, 13-15.

Blanton, M., Stephens, A., Knuth, E., Gardiner, A. M., Lsler, I., & Kim, J. S. (2015). The development of children's early algebraic thinking: The impact of a comprehensive early algebra intervention in third grade. *Journal for Research in Mathematics Education, 46*, 39-87. doi:10.5951/jresematheduc.46.1.0039

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7,* 161-186. doi: 10.1016/S0959-4752(96)00015-1

Byrd, C. E., McNeil, N. M., Chesney, D. L., & Matthews, P. G. (2015). A specific misconception of the equal sign acts as a barrier to children's learning of early algebra. *Learning and Individual Differences, 38*, 61-67. doi: 10.1016/j.lindif.2015.01.001

Carpenter, T. P., Franke, M. L. & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school.* Portsmouth, NH: Heinemann.

Carr, M., Alexander, J., & Folds-Bennett, T. (1994). Metacognition and mathematics strategy use. *Applied Cognitive Psychology*, *8*, 583-595. doi: 10.1002/acp.2350080605

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

De Clercq, A., Desoete, A., & Roeyers, H. (2000). EPA2000: A multilingual, programmable computer assessment of off-line metacognition in children with mathematical-learning disabilities. *Behavior Research Methods, Instruments, & Computers*, *32*(2), 304-311. doi: 10.3758/BF03207799

Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities*, *34*(5), 435-447. doi: 10.1177/002221940103400505

Desoete, A., Roeyers, H., & De Clercq, A. (2003). Can offline metacognition enhance mathematical problem solving? *Journal of Educational Psychology*, *95*(1), 188. doi: 10.1037/0022-0663.95.1.188

Desoete, A., Roeyers, H., & Huylebroeck, A. (2006). Metacognitive skills in Belgian third grade children (age 8 to 9) with and without mathematical learning disabilities. *Metacognition and Learning*, *1*(2), 119-135. doi: 10.1007/s11409-006-8152-9

Desoete, A. (2007). Evaluating and improving the mathematics teaching-learning process through metacognition. *Electronic Journal of Research in Educational Psychology, N. 13 5(3)*, 705-730. ISSN: 1696-2095.

Desoete, A. (2008). Multi-method assessment of metacognitive skills in elementary school children: How you test is what you get. *Metacognition and Learning*, *3*(3), 189. doi: 10.1007/s11409-008-9026-0

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

Destan, N., Spiess, M. A., de Bruin, A., van Loon, M., & Roebers, C. M. (2017). 6-and 8-year-olds' performance evaluations: Do they differ between self and unknown others? *Metacognition and Learning*, *12*(3), 315-336. doi: 10.1007/s11409-017-9170-5

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, *3*(3), 231-264. doi: 10.1007/s11409-018-9181-x

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Sexton, H. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428. doi: 10.1037/0012-1649.44.1.217

Erickson, S., & Heit, E. (2015). Metacognition and confidence: comparing math to other academic subjects. *Frontiers in Psychology*, *6*, 742. doi: 10.3389/fpsyg.2015.00742

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-236). Hillsdale: Erlbaum.

Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology, 108*(1), 82. doi: 10.1037/edu0000053

Fyfe, E. R., Matthews, P. G., & Amsel, E. (2017). College students' knowledge of the equal sign and its relation to solving equations. In E. Galindo & J. Newton (Eds.), Proceedings of the 39th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 279-282). Indianapolis, IN: Hoosier Association of Mathematics Teacher Educators.

Fyfe, E. R., Evans, J. L, Matz, L., Hunt, K., & Alibali, M. W. (2017). Relations between patterning skill and differing aspects of early mathematics knowledge. *Cognitive Development, 44*, 1-11. doi:10.1016/j.cogdev.2017.07.003

García, T., Rodríguez, C., González-Castro, P., González-Pienda, J. A., & Torrance, M. (2016). Elementary students' metacognitive processes and post-performance calibration on mathematical problem-solving tasks. *Metacognition and Learning*, *11*(2), 139-170. doi: 10.1007/s11409-015-9139-1

Garofalo, J., & Lester, F. K. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal of Research in Mathematics Education*, *16*, 163–176. Stable URL: http://www.jstor.org/stable/748391

Garrett, A. J., Mazzocco, M. M., & Baker, L. (2006). Development of the metacognitive skills of prediction and evaluation in children with or without math disability. *Learning Disabilities Research & Practice*, *21*(2), 77-88. doi: 10.1111/j.1540-5826.2006.00208.x

Karabenick, S. A. & Berger, J. (2013). Chapter 8: Help seeking as a self-regulated learning strategy In Bembenutty, H., Cleary, T. J., & Kitsantas, A. (Ed.), *Applications of self-regulated learning across diverse disciplines. A tribute to Barry J. Zimmerman* (pp. 237-261). Charlotte, NC: IAP.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics, 12*, 317-326. doi:10.1007/BF00311062

Koriat, A., Nussinson, R., & Ackerman, R. (2014). Judgments of learning depend on how learners interpret study effort. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1624. doi: 10.1037/xlm0000009

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal*, *40*(1), 281-310. Stable URL: https://www.jstor.org/stable/3699433

Krebs, S. S., & Roebers, C. M. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *British Journal of Educational Psychology*, *80*(3), 325-340. doi: 10.1348/000709910X485719

Lester, F. K. (1982). Building bridges between psychological and mathematics education research on problem solving. In F. K. Lester & J. Garofalo (Eds.), *Mathematical problem solving* (pp. 55–85). Philadelphia: The Franklin Institute Press.

Lipko, A. R., Dunlosky, J., & Merriman, W. E. (2009). Persistent overconfidence despite practice: The role of task experience in preschoolers' recall predictions. *Journal of Experimental Child Psychology*, *103*(2), 152-166. doi: 10.1016/j.jecp.2008.10.002

Lockl, K., & Schneider, W. (2003). Metacognitive monitoring and self-control processes for children's allocation of study time. *Zeitschrift fur Padagogische Psychologie, 17,* 173-182. doi: 1010-0652.17.3.173

Lucangeli, D., & Cornoldi, C. (1997). Mathematics and metacognition: What is the nature of the relationship?. *Mathematical Cognition*, *3*(2), 121-139. doi: 10.1080/135467997387443

Lyons, K. E., & Ghetti, S. (2013). I don't want to pick! Introspection on uncertainty supports early strategic behavior. *Child Development*, *84*(2), 726-736. doi: 10.1111/cdev.12004

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

Matthews, P. G., & Fuchs, L. S. (2018). Keys to the gate? Equal sign knowledge at second grade predicts fourth-grade algebra competence. *Child Development.* doi:10.1111/cdev.13144

McNeil, N. M. (2014). A change-resistance account of children's difficulties understanding mathematical equivalence. *Child Development Perspectives, 8*, 42-47. doi: 10.1111/cdep.12062

McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development*, *76*(4), 883-899. doi: 10.1111/j.1467-8624.2005.00884.x

McNeil, N. M., Fyfe, E. R., Petersen, L. A., Dunwiddie, A. E., & Brletic-Shipley, H. (2011). Benefits of practicing 4= 2+ 2: Nontraditional problem formats facilitate children's understanding of mathematical equivalence. *Child Development*, *82*(5), 1620-1633. doi: 10.1111/j.1467-8624.2011.01622.x

McNeil, N. M., Chesney, D. L., Matthews, P. G., Fyfe, E. R., Petersen, L. A., & Dunwiddie, A. E. (2012). It pays to be organized: Organizing arithmetic practice around equivalent values facilitates understanding of math equivalence. *Journal of Educational Psychology, 104*(4), 1109-1121. doi:10.1037/a0028997

McNeil, N. M., Hornburg, C. B., Devlin, B. L., Carrazza, C. & McKeever, M. O. (2017). Consequences of individual differences in children's formal understanding of mathematical equivalence. *Child Development*. doi:10.1111/cdev.12948

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159-163. doi: 10.1111/j.1467-8721.2009.01628.x

Mevarech, Z., & Fridkin, S. (2006). The effects of IMPROVE on mathematical knowledge, mathematical reasoning and meta-cognition. *Metacognition and Learning*, *1*(1), 85-97. doi: 10.1007/s11409-006-6584-x

Mevarech, Z. R., & Kramarski, B. (2003). The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. *British Journal of Educational Psychology*, *73*(4), 449-471. doi: 10.1348/000709903322591181

NGA Center & CCSSO. (2010). Common core state standards for mathematics. Washington, DC: Author. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings, IN G. H. Bower (Ed.), *The psychology of learning and instruction: Advances in research and theory* (vol. 26, pp. 125-141). New York, NY: Academic Press.

Nelson-Le Gall, S. (1985). Help-Seeking Behavior in Learning. *American Educational Research Association, 12*, 55-90. doi: 10.2307/1167146

Newman, R. S. (2000). Social influences on the development of children's adaptive help seeking: The role of parents, teachers, and peers. *Developmental Review*, *20*(3), 350-404. doi: 10.1006/drev.1999.0502

Özsoy, G., & Ataman, A. (2017). The effect of metacognitive strategy training on mathematical problem solving achievement. *International Electronic Journal of Elementary Education*, *1*(2), 67-82. doi: 10.1111/cdev.12644

Pennequin, V., Sorel, O., Nanty, I., & Fontaine, R. (2010). Metacognition and low achievement in mathematics: The effect of training in the use of metacognitive skills to solve mathematical word problems. *Thinking & Reasoning*, *16*(3), 198-220. doi: 10.1080/13546783.2010.509052

Powell, S. R. (2012). Equations and the equal sign in elementary mathematics textbooks. *The Elementary School Journal, 112*, 627-648. doi: 10.1086/665009

Powell, S. R., & Fuchs, L. S. (2010). Contribution of equal-sign instruction beyond word-problem tutoring for third-grade students with mathematics difficulty. *Journal of Educational Psychology*, *102*, 381–394. doi:10.1037/a0018447

Raaijmakers, S. F., Baars, M., Paas, F., van Merriënboer, J. J., & van Gog, T. (2019). Effects of self-assessment feedback on self-assessment and task-selection accuracy. *Metacognition and Learning*, 1-22. doi: 10.1002/acp.2883

Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., & Van Gog, T. (2017). Effects of performance feedback valence on perceptions of invested mental effort. *Learning and Instruction*, *51*, 36-46. doi: 10.1016/j.learninstruc.2016.12.002

Renwick, E. (1932). Children's misconceptions concerning the symbols for mathematical equality. *The British Journal of Educational Psychology*, *2*, 173-183. doi: 10.1111/j.2044-8279.1932.tb02743.x

Rinne, L. F., & Mazzocco, M. M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PloS one*, *9*(7). doi: 10.1371/journal.pone.0098663

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77*, 1-15. doi:10.1111/j.1467-8624.2006.00852.x

Rittle-Johnson, B., Matthews, P. G., Taylor, R. S., & McEldoon, K. L. (2011). Assessing knowledge of mathematical equivalence: A construct-modeling approach. *Journal of Educational Psychology*, *103*(1), 85. doi: 10.1037/a0021334

Roderer, T., & Roebers, C. M. (2010). Explicit and implicit confidence judgments and developmental differences in metamemory: An eye-tracking approach. *Metacognition and Learning*, *5*(3), 229-250. doi: 10.1007/s11409-010-9059-z

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

Roebers, C. M., & Spiess, M. (2017). The development of metacognitive monitoring and control in second graders: A short-term longitudinal study. *Journal of Cognition and Development*, *18*(1), 110-128. doi: 10.1080/15248372.2016.1157079

Roebers, C. M., von der Linden, N., & Howie, P. (2007). Favourable and unfavourable conditions for children's confidence judgments. *British Journal of Developmental Psychology*, *25*(1), 109-134. doi: 10.1348/026151006X104392

Schneider, W., & Artelt, C. (2010). Metacognition and mathematics education. *ZDM*, *42*(2), 149-161. doi: 10.1007/211858-010-0240-2

Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science*, *1*(2), 291-297. doi: 10.1111/1467-7687.00044

Schoenfeld, A. H. (1987). What's all the fuss about metacognition? In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education*, (pp. 189-215). Hillsdale, NJL Lawrence Erlbaum Associates, Inc.

Shin, H., Bjorklund, D. F., & Beck, E. F. (2007). The adaptive nature of children's overestimation in a strategic memory task. *Cognitive Development*, *22*(2), 197-212. doi: 10.1016/j.cogdev.2006.10.001

Silver, E. A. (1982). Knowledge organization and mathematical problem solving. In F. K. Lester & J. Garofalo (Eds.), *Mathematical problem solving* (pp. 15–25). Philadelphia: Franklin Institute Press.

Thiede, K. W., Anderson, M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of educational psychology*, *95*(1), 66. doi: 10.1037/0022-0663.95.1.66

van Loon, M. H., de Bruin, A. B., van Gog, T., & van Merriënboer, J. J. (2013). Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration. *Learning and Instruction*, *24*, 15-25. doi: 10.1016/j.learninstruc.2012.08.005

Veenman, M. V. J. (2006). The role of intellectual and metacognitive skills in math problem solving. In A. Desoete & M. Veenman (Eds.), *Metacognition in mathematics education*, (pp. 35-50). Haupauge, NY: Nova Science.

Vo, V. A., Li, R., Kornell, N., Pouget, A., & Cantlon, J. F. (2014). Young children bet on their numerical skills: metacognition in the numerical domain. *Psychological Science*, *25*(9), 1712-1721. doi: 10.1177/0956797614538458

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *The Journal of Educational Research*, *84*(1), 30-43. Stable URL: https://www.jstor.org/stable/40539680

Whitebread, D., & Coltman, P. (2010). Aspects of pedagogy supporting metacognition and self-regulation in mathematical learning of young children: evidence from an observational study. *ZDM*, *42*(2), 163-178. doi: 10.1007/s11858-009-0233-1

Tables and Figures

Table 1

Matched target problems in standard session and help session

Problem	Set 1	Set 2
A	7 = 4 + __	8 = 3 + __
B	3 + 7 = 3 + __	4 + 6 = 4 + __
C	9 + 4 = 3 + __	5 + 9 = 4 + __
D	3 + 7 + 8 = __ + 8	8 + 3 + 7 = __ + 7
E	5 + 3 + 4 = 5 + __	4 + 5 + 3 = 4 + __

Table 2

Accurate monitoring scoring system and strategic control scoring system

Accurate Monitoring Score		
	Accuracy on the Problem	
Certainty Rating	Right	Wrong
1 (I know I got this problem wrong)	0.00	1.00
2 (I think I got this problem wrong)	0.33	0.67
3 (I think I got this problem right)	0.67	0.33
4 (I know I got this problem right)	1.00	0.00
Strategic Control Score		
	Decision to Ask for Help	
Certainty Rating	Yes	No
1 (I know I got this problem wrong)	1.00	0.00
2 (I think I got this problem wrong)	0.67	0.33
3 (I think I got this problem right)	0.33	0.67
4 (I know I got this problem right)	0.00	1.00

Table 3

Relations between problem-solving accuracy and certainty ratings in the standard session

	Problem Correct	**Problem Incorrect**	**Total Match Rate**
(1) I know I got this problem wrong	<1% (Mismatch)	1% (Match)	80%
(2) I think I got this problem wrong	2% (Mismatch)	12% (Match)	86%
(3) I think I got this problem right	21% (Match)	22% (Mismatch)	49%
(4) I know I got this problem right	23% (Match)	18% (Mismatch)	56%
Total Match Rate	94%	25%	57%

Note. The values in each cell represent the percentage of problems in the standard session that met the specified criterion. For example, in the top left cell, <1% represents the percentage of all problems that children solved correctly and rated their certainty as a 1 ("I know I got this problem wrong"). Similarly, in the bottom left cell, 94% represents the percentage of all correct problems on which children matched their certainty rating by selecting a rating of 3 or 4.

Table 4

Relations between help-seeking decisions in the help session and certainty ratings in the standard session

TOTAL	**Seek Help**	**No Help**	**Total Match Rate**
(1) I know I got this problem wrong	<1% (Strategic)	1% (Non-Strategic)	20%
(2) I think I got this problem wrong	7% (Strategic)	8% (Non-Strategic)	47%
(3) I think I got this problem right	7% (Non-Strategic)	36% (Strategic)	84%
(4) I know I got this problem right	3% (Non-Strategic)	38% (Strategic)	93%
Total Strategic Rate	41%	89%	81%

Note. The values in each cell represent the percentage of problems in the help session that met the specified criterion. For example, in the top left cell, <1% represents the percentage of all problems in the help session on which children asked for help and rated their certainty as a 1 (“I know I got this problem wrong”) on the corresponding problem in the standard session.

Table 5

Summary of stepwise regression predicting posttest performance

MODEL 1	ß	*p*	ΔR^2
Step 1			.82
Baseline Knowledge (low vs. high)	.278	.001	
Standard Session Accuracy	.672	.000	
Standard Session Certainty	.015	.844	
Help Session Help-Seeking (yes vs. no)	-.104	.166	
Step 2			.02
Baseline Knowledge (low vs. high)	.237	.003	
Standard Session Accuracy	.457	.001	
Standard Session Certainty	.085	.287	
Help Session Help-Seeking (yes vs. no)	-.067	.363	
Accurate Monitoring Scores	.269	.033	
MODEL 2			
Step 1			.82
Baseline Knowledge (low vs. high)	.278	.001	
Standard Session Accuracy	.672	.000	
Standard Session Certainty	.015	.844	
Help Session Help-Seeking (yes vs. no)	-.104	.166	
Step 2			.01
Baseline Knowledge (low vs. high)	.286	.000	
Standard Session Accuracy	.654	.000	
Standard Session Certainty	-.061	.560	
Help Session Help-Seeking (yes vs. no)	-.052	.559	
Strategic Control Scores	.135	.300	

Note. Baseline knowledge was dichotomized into a low knowledge (0) and high knowledge group (1). Standard session accuracy was the number of equivalence solved correctly (out of 5). Standard Session certainty was measured on a scale ranging from 1 = "I know I got this problem wrong," to 4 = "I know I got this problem right." Help-seeking was quantified as a dichotomous variable consisting of asking for help on at least one problem versus not asking for help at all. Accurate monitoring scores assessed the calibration between certainty ratings and accuracy and ranged from 0 to 1 (averaged across five problems). Strategic control scores assessed the calibration between help-seeking decisions and certainty ratings and ranged from 0 to 1 (averaged across five problems).

Figure 1

Target problem and certainty rating scale in standard session.

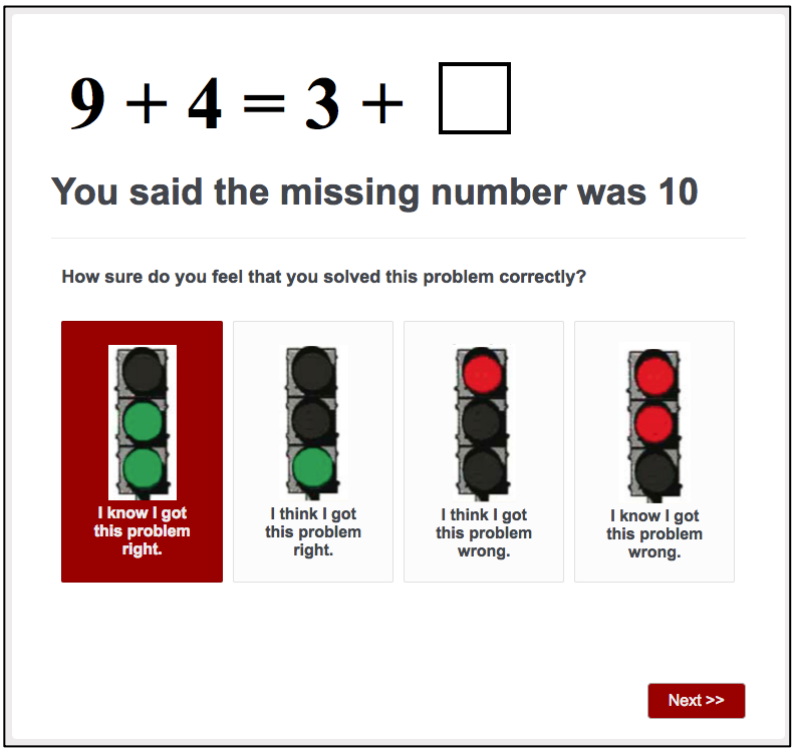

Figure 2.

The procedure.