Developmental Psychology
2016, Vol. 52, No. 10, 1493–1502

0012-1649/16/$12.00 http://dx.doi.org/10.1037/dev0000180

Children Can Accurately Monitor and Control Their Number-Line Estimation Performance

Jenna L. Wall, Clarissa A. Thompson, John Dunlosky, and William E. Merriman
Kent State University

Accurate monitoring and control are essential for effective self-regulated learning. These metacognitive abilities may be particularly important for developing math skills, such as when children are deciding whether a math task is difficult or whether they made a mistake on a particular item. The present experiments investigate children's ability to monitor and control their math performance. Experiment 1 assessed task- and item-level monitoring while children performed a number line estimation task. Children in 1st, 2nd, and 4th grade ($N = 59$) estimated the location of numbers on small- and large-scale number lines and judged their confidence in each estimate. Consistent with their performance, children were more confident in their small-scale estimates than their large-scale estimates. Experiments 2 ($N = 54$) and 3 ($N = 85$) replicated this finding in new samples of 1st, 2nd, and 4th graders and assessed task- and item-level control. When asked which estimates they wanted the experimenter to evaluate for a reward, children tended to select estimates associated with lower error and higher confidence. Thus, children can accurately monitor their performance during number line estimation and use their monitoring to control their subsequent performance.

Keywords: number-line estimation, metacognition, monitoring, control, confidence judgments

Sarah's first grade class completed a lesson on numerical magnitude. Their assignment was to estimate the location of numbers on a number line. At first, Sarah was confident in her performance, announcing "that's easy" as she estimated "2" on a 0–10 line. As the scale and to-be-estimated numbers increased, however, Sarah struggled. For example, when estimating "78" on a 0–100 line, she responded slowly and had trouble deciding where the number should go. However, is Sarah aware that she has more difficulty with the larger numeric range? Does she know when she has difficulty placing some numbers within each range? Metacognitive awareness, or the ability to monitor one's performance, is essential for self-regulated learning. Metacognition refers to people's thinking about their cognition (Flavell, 1979) and includes multiple components, including people's monitoring—or awareness—of ongoing progress and performance on a task as well as their control of task performance (for detailed discussion of metacognition, see Dunlosky & Metcalfe, 2009). For Sarah, this awareness could occur at the task level, where estimating in a smaller numeric range (0–10) is recognized as easier than estimating in a larger range (0–100). This awareness may also occur at the item level, where Sarah is more confident about the numbers that she estimates more accurately. Sensitivity at either level may allow Sarah to control the accuracy of her performance, such as by withholding answers from evaluation that she believes are incorrect (Koriat & Goldsmith, 1996; Winne & Hadwin, 1998).

Nothing is known about metacognitive monitoring and control skills in the domain of numerical estimation. Given the paucity of research on metacognition and numerical estimation, the current studies sought to answer three questions: At the task level, are children aware they are less accurate on a larger range? At the item level, are children aware they are less accurate when they have difficulties placing some numbers? Finally, can children use this awareness to control subsequent performance? We answered these questions in the context of a number line estimation task for two reasons. First, the quality of children's estimates is related to proficiency in arithmetic, memory for numbers, and math achievement (Booth & Siegler, 2006, 2008; Laski & Siegler, 2007; Schneider, Grabner, & Paetsch, 2009; Siegler, Thompson, & Schneider, 2011; Siegler & Thompson, 2014; Thompson & Siegler, 2010). Second, with age and experience, children's estimates become more accurate as they progress from a less advanced to more advanced representation (see Siegler, Thompson, & Opfer, 2009). This representational shift takes place by kindergarten for 0–10 estimates (Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010), by 2nd grade for 0–100 estimates (Geary, Hoard, Nugent, & Byrd-Craven, 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003), 4th grade for 0–1,000 estimates (Booth & Siegler, 2006; Siegler & Opfer, 2003), and 6th grade for 0–100,000 estimates (Thompson & Opfer, 2010). For example, a 1st grader might overestimate the location of "7" on a 0–100 line, but place it accurately on a 0–10 line. Although researchers agree that devel-

This article was published Online First August 22, 2016.

Jenna L. Wall, Clarissa A. Thompson, John Dunlosky, and William E. Merriman, Department of Psychological Sciences, Kent State University.

We thank Hannah Herc, Melissa Bishop, Mallory Klein, Morgan Buerke, Veronica Musser, Lisa Sukenaga, and Drew Norberg for their help with data collection and coding. We would also like to thank Annette Kratcoski for her help with participant recruitment at the Kent State University AT&T classroom and the teachers, parents, and students from the following school districts: Stow-Munroe Falls, Lippman, Streetsboro, and Ravenna.

Correspondence concerning this article should be addressed to Jenna L. Wall, Department of Psychological Sciences, Kent State University, 600 Hilltop Drive, Kent, OH 44240. E-mail: jwall4@kent.edu

opmental changes exist in numerical estimation abilities across ages and numerical ranges, the mechanism(s) responsible for these developmental changes are debated (Barth & Paladino, 2011; Barth, Slusser, Cohen, & Paladino, 2011; Chesney & Matthews, 2013; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Dackermann, Huber, Bahnmueller, Nuerk, & Moeller, 2015; Huber, Moeller, Nuerk, & Willmes, 2013; Hurst, Monahan, Heller, & Cordes, 2014; Link, Nuerk, & Moeller, 2014; Opfer, Siegler, & Young, 2011; Opfer, Thompson, & Kim, 2016; Rips, 2013; Slusser, Santiago, & Barth, 2013). The current studies focus on children's awareness of their difficulties estimating numbers in large versus small numerical ranges rather than the processes underlying these developmental changes in numerical estimation. Specifically, we asked whether children are aware that their small-scale estimates tend to be more accurate than their large-scale ones.

Although nothing is known about children's metacognitive skills during number line estimation, evidence from the memory domain suggests that this awareness may be in place in elementary school. For example, by 8 years of age, children are able to monitor their memory performance. That is, children accurately predict their memory span and judge whether they would be able to recall recently studied items after a delay (for a review, see Schneider & Loffler, 2016). Several recent studies suggest these skills may even be present as early as preschool (Lipowski, Merriman, & Dunlosky, 2013). For example, while identifying perceptually degraded images, 3- to 5-year-olds reported feeling less confident on trials where they had erred and were more likely to seek help or skip these trials if given the opportunity (Coughlin, Hembacher, Lyons, & Ghetti, 2015; Lyons & Ghetti, 2013). These findings suggest that preschoolers monitor the accuracy of their performance and use this skill to guide decision-making. Hembacher and Ghetti (2014) reported a similar finding. Preschoolers studied a series of images and then completed a delayed recognition test. On each trial, children judged which of two images they had studied and then rated their confidence on a 3-point scale. Finally, children sorted their answers into one of two boxes. They were told that the answers placed in the *open-eyes* box would be evaluated for a reward, whereas answers placed in the *closed-eyes* box would not. Overall, 4- and 5-year-olds were more confident in their correct answers and placed them in the open-eyes box.

These findings suggest that metacognitive monitoring and control skills are present in preschool, but there are several reasons why this awareness may not be evident during math-related problem-solving. First, even adults, who supposedly have fully developed monitoring skills (for a review, see Dunlosky & Metcalfe, 2009), have difficulty estimating the location of very large numbers (Landy, Silbert, & Goldin, 2013). Similarly, even after years of instruction, middle school students continue to use faulty strategies when estimating the location of fractions on a number line (Siegler et al., 2011). Second, although the Common Core, a set of curriculum guidelines adopted by many states in the U.S., has recognized the number line as an important tool for visualizing numerical magnitudes, it's unlikely that children are being asked to reflect on their estimation performance during mathematics lessons. Finally, early control skills are fragile and continue to develop throughout childhood. For example, when deciding how long to study a list of items, elementary schoolchildren failed to allot more time for longer lists (Flavell, Friedrichs, & Hoyt, 1970; Gettinger, 1985; Leal, Crays, & Moely, 1985) or more difficult items (Dufresne & Kobasigawa, 1989).

Given (a) that both adults and children struggle on some number line tasks despite years of instruction on number concepts, (b) children's limited experience with monitoring their confidence in their number line estimates, and (c) the fragility of early control skills, children may find it difficult to monitor and/or control their estimation performance. Thus, the goal of the current studies was to examine whether these skills are present during the elementary school years.

To begin filling this gap in the literature, we assessed the generality and potential development of this awareness in several age groups. We adapted Hembacher and Ghetti's (2014) memory monitoring/control instructions to assess whether children show task- and item-level awareness during a numerical estimation task. Children in 1st, 2nd, and 4th grade estimated the location of numbers within a small and large scale and judged their confidence in each estimate. The scales differed by grade and, on the basis of previous research, were appropriate for discriminating between less advanced and more advanced representations.

To assess task- and item-level monitoring, Experiment 1 asked two questions. First, at the task level, are children aware they are less accurate on a larger (i.e., more difficult) numerical scale? If so, children should report higher confidence in small- compared to large-scale estimates. Second, at the item level, are children aware that the numbers they have trouble placing tend to be less accurate? If so, children should report higher confidence in estimates placed closer to the correct location on the line compared to those farther away.

In addition to replicating Experiment 1, Experiment 2 assessed both task- and item-level control. After providing confidence judgments for each estimate, children were promised a reward for perfect performance. The open-eyes and closed-eyes boxes were introduced, and the experimenter said she would only evaluate estimates in the open-eyes box. To increase their chances of receiving the reward, children decided which scale they wanted the experimenter to evaluate, and then sorted each estimate into one of the boxes. Thus, the purpose of Experiment 2 was to examine whether children use their awareness to control their final performance. If so, children should put the scale and estimates they were more confident in into the open-eyes box and those they were less confident in into the closed-eyes box. Finally, Experiment 3 sought to replicate this finding with immediate, rather than delayed, control judgments.

Experiment 1

Participants

Participants were 60 children from public elementary schools in Northeast Ohio. Parental permission forms were distributed to all children in one classroom at each grade level; all children whose parents consented were tested. The sample was predominately Caucasian (83%) and consisted of 19 1st graders (9 males; mean age = 7.02 years *SD* = 4.59 months), 24 2nd graders (12 males; mean age = 8.09 years, *SD* = 3.97 months), and 17 4th graders (8 males; mean age = 10.09 years, *SD* = 3.74 months). One 1st grader failed to use the whole number line and was excluded from analyses. Each child received a sticker for participating.

Tasks and Procedure

Number-line estimation. Children completed two packets, each consisting of 18 number lines from either a small or large scale. These scales differed by grade (1st grade: 0–10 and 0–100; 2nd grade: 0–100 and 0–1,000; 4th grade: 0–1,000 and 0–100,000) and were chosen based on the expectation that children would make more accurate estimates on the small scale and less accurate estimates on the large scale. Thus, estimating numbers on the small scale should be easier and result in more accurate performance than estimating on the large scale. Each number line was displayed on its own page with a to-be-estimated number in the top left-hand corner (see Figure 1; the to-be-estimated numbers for each scale can be found in Appendix). The number line estimation task was explained using the same instructions from typical number line tasks (e.g., Opfer & Siegler, 2007; Opfer & Thompson, 2008; Thompson & Opfer, 2008). As a warm-up, children were asked to point to the left and right endpoints of the number line (e.g., 0 and 10). Children rarely erred during the initial instruction phase, but if a child did err, he or she was corrected, and the instructions were explained again. On each trial, children were instructed to estimate the location of a number on the line. For example, "*If '0' is here and '10' is here, where does '7' go?*" Children responded by making a vertical mark through the line with their pen. No feedback was provided. Two random orders of the to-be-estimated numbers were randomly assigned to each packet. Some children completed the small scale first, whereas others completed the large scale first. No significant order effects were found, so all of our analyses were collapsed across order.

Confidence judgments. After making each estimate, children reported how confident they were that they marked the line in the correct location. Hembacher and Ghetti's (2014) 3-point confidence scale was displayed below each line (see Figure 1). Using a modified version of Hembacher and Ghetti's script, children were instructed to circle the picture on the left when they were "not so sure" about their estimate, the picture in the middle when they were "kind of sure," and the picture on the right when they were "really sure." Children were asked to identify the meaning of each picture prior to beginning the study (e.g., "*Which picture would you circle if you are not so sure you marked the number line in the correct place?*"). Any errors were corrected, and the scale was explained again.

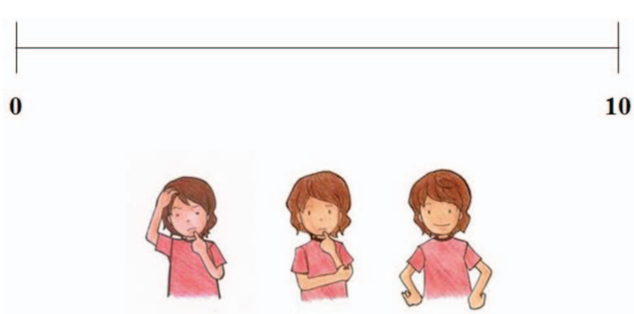

Figure 1. Example number line estimation trial (0–10 scale) and confidence rating scale. Left: "*Not so sure*"; Middle: "*Kind of sure*"; Right: "*Really sure.*" Adapted with permission from "Don't Look At My Answer: Subjective Uncertainty Underlies Preschoolers' Exclusion of Their Least Accurate Memories," by E. Hembacher, S. Ghetti, 2014, *Psychological Science*, *25*, p.1768. Copyright, 2014 by Sage. See the online article for the color version of this figure.

Counting assessment. Children completed two counting tasks (adapted from Barth, Starr, & Sullivan, 2009). Results from the counting tasks were not central to our main questions, and preliminary results indicated that average confidence and error (percent absolute error; PAE) did not differ for children who made a counting error versus those who did not. Thus, results of the assessment will not be discussed further.

Results

Task-level monitoring. Although performance on the estimation task was not our focus, one goal was to evaluate whether children's confidence judgments differ depending on the quality of their estimates. Thus, we first describe how children's estimates were classified as a less- versus more-advanced representation.

For each scale, the quality of a child's representation was determined by the best fitting function for his or her estimates. If the estimates increased linearly with magnitude, the child understood the equal-interval property of the line and was coded as *advanced*. Children whose estimates were better fit by the logarithmic than the linear function were coded as *less advanced*. Although the majority of children showed the expected developmental pattern—that is, advanced on the small scale and less advanced on the large scale—some children were advanced on both scales (see Table 1). To isolate the role representation plays in confidence, we analyzed our main questions overall and at the representational level.[1]

Confidence was scored on a 3-point scale. *Not so sure* judgments were coded as 1; *kind of sure* judgments were coded as 2; and *really sure* judgments were coded as 3. To assess monitoring at the task level, mean confidence was computed by averaging the 18 confidence judgments for each scale (see Table 2). A 2 (scale: small, large) × 3 (grade: 1st, 2nd, 4th) analysis of variance (ANOVA) revealed a main effect of scale, $F(1, 56) = 31.31$, $p < .01$, $\eta^2 = .35$. All grades reported significantly higher confidence in small- compared to large-scale estimates. No main effect of grade, $F(2, 56) = 2.36$, $p > .05$, or interaction, $F(2, 56) = 1.79$, $p > .05$, was found. We then limited our analyses to children who showed the expected representation pattern (middle columns of Table 2). A 2 × 3 ANOVA revealed a main effect of scale, $F(1, 28) = 12.12$, $p < .01$, $\eta^2 = .30$, and a trend toward a main effect of grade, $F(2, 28) = 3.06$, $p = .06$, $\eta^2 = .18$. No significant interaction was found, $F(2, 28) = .10$, $p > .05$. Although the main effect of scale suggests children are sensitive to their difficulties on the large scale, the association between confidence and scale was not dependent upon the quality of their estimates. Even children who possessed an advanced representation for both scales reported higher confidence in their small-scale performance (right-hand columns of Table 2). A main effect was found for scale, $F(1, 22) = 13.19$, $p < .01$, $\eta^2 = .32$, but not grade, $F(2, 22) = .60$, $p > .05$,

[1] Given the small sample size for the Less Advanced-Less Advanced and Less Advanced-Advanced patterns, we did not analyze these groups.

Table 1
Number of Children Possessing Each Representational Pattern

Grade	Quality of Representation for Small and Large Scale: Advanced-Less Advanced	Advanced-Advanced	Less Advanced-Less Advanced-	Less Advanced-Advanced
Experiment 1				
1st	12	6	0	0
2nd	14	9	1	0
4th	5	10	0	2
Experiment 2				
1st	7	7	1	1
2nd	11	9	1	0
4th	6	9	0	1

and an interaction approached significance, $F(2, 22) = 3.16$, $p = .06$, $\eta^2 = .18$. Simple effects tests revealed that this difference was only significant for the 4th graders, $t(9) = 3.49$, $p = .01$, $d = 1.02$.

Item-level monitoring. Although item-level error was not our focus, one goal was to compare a child's confidence with their accuracy for each estimate. Accuracy was measured with percent absolute error (PAE): PAE = [(|Participant's Estimate—Correct Answer|)/Scale of the Number Line] × 100 (Siegler & Booth, 2004). Note that higher PAE indicates less accurate estimates. For example, if a child was asked to estimate the location of "7" on a 0–10 line but marked the location corresponding to "9", the PAE would be [(|9–7|)/10] × 100, or 20%. Each participant's mean PAE was computed separately for the small and large scales (see Table 3). A 2 × 3 ANOVA revealed a main effect of scale, $F(1, 56) = 14.85$, $p < .01$, $\eta^2 = .18$, a main effect of grade, $F(1, 56) = 17.41$, $p < .01$, $\eta^2 = .06$, and an interaction, $F(2, 56) = 5.44$, $p = .01$, $\eta^2 = .13$. Although all children showed lower error when estimating on the small compared to large scale, this difference was only significant for 2nd, $t(23) = 9.09$, $p < .01$, $d = 1.80$, and 4th, $t(16) = 3.56$, $p < .01$, $d = .86$, graders.

To assess monitoring at the item level, we examined the relationship between confidence and PAE across all numbers within each scale. If children are aware that they had difficulties estimating certain numbers and that these estimates are likely to be less accurate than others, confidence should be lower for numbers placed far away from the correct location and higher for numbers placed closer. For each participant, a gamma correlation was computed between his or her confidence and PAE (see Table 4). This correlation is a measure of relative accuracy that ranges from −1.0 to 1.0, where 0 represents no predictive accuracy and higher values indicate higher levels of accuracy (Nelson, 1984).[2] Gamma is a nonparametric correlation that is used extensively in metacognitive research to estimate judgment accuracy because it does not assume equal intervals between levels of a measure (which is an inappropriate assumption for judgment scales, as per Figure 1). In the present case, the correlation represents children's ability to monitor their performance while making estimates. In particular, mean values that are significantly greater than zero indicate that children's judgments have above-chance accuracy at discriminating between estimates that show greater (vs. less) error. As shown in Table 4, across all grades and conditions, the mean values were positive, with the majority being significantly greater than 0. Thus, children have some ability to monitor the accuracy of their estimates.

Discussion

Experiment 1 demonstrated that elementary schoolchildren have some ability to accurately monitor their math performance at both the task and item level. However, these results are preliminary and need to be replicated. Moreover, whether children use their monitoring ability to control their performance remains unknown.

Experiment 2

To address these issues, children in Experiment 2 completed the same monitoring task from Experiment 1 along with an additional task designed to tap both task- and item-level control. To increase their chances of receiving a reward, children decided which scale they wanted the experimenter to evaluate (task level) and then sorted estimates they thought they got right into the open-eyes box and estimates they thought they got wrong into the closed-eyes box (item level).

Participants

Fifty-five children participated and were recruited in a similar manner as in Experiment 1; parental permission slips were distributed to all children in one 1st and one 4th grade classroom, and two 2nd grade classrooms. The sample was predominantly Caucasian (93%) and consisted of 17 1st graders (8 males; mean age = 7.24 years, $SD = 6.54$ months), 21 2nd graders (11 males; mean age = 8.42 years, $SD = 10.36$ months), and 17 4th graders (5 males; mean age = 10.16 years, $SD = 3.56$ months). One 4th grader requested to leave before the study was over and was thus excluded from analyses. Each child received two stickers for participating.

Tasks and Procedure

The procedure was identical to Experiment 1 except for the addition of a sorting task. After estimating all 36 numerals and

[2] Measures based on signal detection theory (e.g., d_a) arguably are even better suited (Masson & Rotello, 2009) but require a more fined-grained judgment scale (and often many more observations) to compute and hence will not be useful for investigations involving young children.

Table 2
Task-Level Monitoring

	Mean Confidence Judgments					
	Overall		Advanced-Less Advanced		Advanced-Advanced	
Grade	Small	Large	Small	Large	Small	Large
Experiment 1						
1st	2.71 (.29)	2.42 (.42)	2.68 (.32)	2.43 (.33)	2.77 (.21)	2.40 (.61)
2nd	2.43 (.33)	2.31 (.42)	2.40 (.34)	2.20 (.47)	2.52 (.31)	2.50 (.29)
4th	2.64 (.26)	2.40 (.24)	2.76 (.31)	2.53 (.20)	2.54 (.21)	2.32 (.22)
Experiment 2						
1st	2.62 (.30)	2.24 (.46)	2.55 (.35)	2.20 (.38)	2.69 (.27)	2.20 (.60)
2nd	2.46 (.36)	2.47 (.35)	2.30 (.40)	2.27 (.34)	2.67 (.20)	2.72 (.19)
4th	2.35 (.39)	2.18 (.43)	2.27 (.34)	2.20 (.52)	2.43 (.33)	2.15 (.42)
Experiment 3						
1st		2.58 (.30)				
2nd		2.25 (.40)				
4th		2.22 (.40)				

Note. Standard deviations are indicated in parentheses.

providing confidence judgments for each, two boxes were introduced; one had a smiley face and open eyes (*open-eyes* box), and the other had a smiley face with closed eyes (*closed-eyes* box; see Figure 2). Children were promised a reward for estimating every number correctly. They could choose which estimates the experimenter evaluated; only estimates placed in the open-eyes box would be evaluated for the reward. Thus, children were instructed to sort estimates they thought they got right into the open-eyes box and estimates they thought they may have gotten wrong into the closed-eyes box.

Children were asked, "*Which number line packet do you want to put in the open-eyes box?*" After the child made his or her selection, the experimenter placed the other scale face down in front of the closed-eyes box and said, "Let's go through the packet you put in the eyes-open box and see if you want me to look at all of your answers, or if there are some you don't want me to look at." Children were shown each estimate one at a time and decided which box to put them in. For example, "*You marked N here and were really sure. Do you want to put this in the open-eyes or closed-eyes box?*" After sorting all 18 estimates, children repeated the same process with the scale they placed in the closed-eyes box. The experimenter stated, "*Let's do the same thing for the packet you put in the eyes-closed box. There may be some answers in here that you want me to see*." Thus, the experimenter promised to look at/ignore individual estimates rather than the packets themselves. Children were required to demonstrate an understanding of the boxes before beginning (i.e., *"Which box would you put the number line in if you think you got it right and want me to look at it?"*) No other feedback was provided, and children received the reward regardless of their overall performance.

Results

Task-level monitoring. As in Experiment 1, children were classified as either *advanced* or *less advanced* (see Table 1). Each grade's mean confidence is listed in the left-hand column of Table

Table 3
Mean PAE Broken Down by Grade and Representational Pattern

	Mean Percent Absolute Error (PAE)					
	Overall		Advanced-Less Advanced		Advanced-Advanced	
Grade	Small	Large	Small	Large	Small	Large
Experiment 1						
1st	18.40 (6.41)	20.51 (9.33)	18.24 (7.41)	23.72 (4.38)	18.73 (4.33)	14.10 (13.39)
2nd	7.95 (3.07)	16.88 (5.69)	8.78 (3.17)	19.57 (4.16)	6.15 (1.83)	11.88 (4.20)
4th	8.40 (7.01)	15.56 (9.18)	7.50 (1.22)	24.57 (7.24)	5.36 (1.66)	9.26 (2.97)
Experiment 2						
1st	21.42 (8.93)	12.72 (5.50)	18.95 (2.83)	16.42 (2.77)	19.83 (11.26)	7.85 (1.78)
2nd	7.69 (3.24)	14.62 (7.61)	8.45 (3.45)	18.84 (6.45)	6.15 (2.09)	8.13 (1.69)
4th	8.62 (6.31)	19.31 (11.67)	9.93 (2.50)	28.70 (4.40)	5.50 (2.69)	10.96 (7.39)
Experiment 3						
1st		16.07 (10.00)				
2nd		25.79 (8.52)				
4th		15.10 (10.00)				

Note. Standard deviations are indicated in parentheses.

Table 4
Item-Level Monitoring

Grade	Mean Gamma Correlations (PAE & CJs)	
	Small	Large
Experiment 1		
1st	.30 (.21)**	.08 (.36)
2nd	.06 (.33)	.22 (.37)*
4th	.25 (.34)*	.28 (.31)**
Experiment 2		
1st	.20 (.49)	.16 (.38)
2nd	.12 (.28)	.20 (.27)*
4th	.27 (.22)**	.32 (.46)*
Experiment 3		
1st		.08 (.23)
2nd		.17 (.29)**
4th		.27 (.27)**

Note. Higher PAE represents a greater percentage of error. Thus, gammas were reverse coded for ease of interpretation; standard deviations are indicated in parentheses.
* $p < .05$. ** $p < .01$.

2. A 2 × 3 ANOVA revealed a main effect of scale, $F(1, 51) = 21.42$, $p < .01$, $\eta^2 = .24$, and an interaction, $F(2, 51) = 8.94$, $p < .01$, $\eta^2 = .20$, but no main effect of grade, $F(2, 51) = 1.56$, $p > .05$. Replicating Experiment 1, 1st and 4th graders were significantly more confident in their small- compared to large-scale estimates, $t(16) = 3.86$, $p < .01$, $d = .94$, and $t(15) = 2.56$, $p < .05$, $d = .40$, respectively.

We then limited our analyses to children who showed the expected representational pattern (middle columns of Table 2). A 2 × 3 ANOVA revealed a main effect of scale, $F(1, 21) = 13.39$, $p < .01$, $\eta^2 = .27$, and a significant interaction, $F(2, 21) = 7.38$, $p < .01$, $\eta^2 = .30$, but no main effect of grade, $F(2, 21) = .23$, $p > .05$. All children reported higher confidence in small- compared to large-scale estimates, but this difference was only significant for 1st graders, $t(6) = 4.31$, $p = .01$, $d = .95$. Finally, for children who were advanced on both scales, a main effect was found for both scale, $F(1, 22) = 7.91$, $p = .01$, $\eta^2 = .22$, and grade, $F(2, 22) = 3.98$, $p < .05$, $\eta^2 = .27$, while an interaction approached significance, $F(2, 22) = 3.26$, $p = .06$, $\eta^2 = .18$. First and fourth graders demonstrated this pattern, yet it was only significant in the 4th graders, $t(8) = 3.11$, $p < .05$, $d = .75$.

Item-level monitoring. Mean PAE was computed separately for each scale (see Table 3). A 2 × 3 ANOVA revealed a main effect of scale, $F(1, 51) = 6.58$, $p = .01$, $\eta^2 = .06$, grade, $F(2, 51) = 4.26$, $p < .05$, $\eta^2 = .14$, and a significant interaction, $F(2, 51) = 25.12$, $p < .01$, $\eta^2 = .47$. As in Experiment 1, 2nd and 4th graders showed significantly lower error in the small compared to large scale, (2nd grade: $M = 7.69\%$ vs. 14.62%, $t(20) = 4.51$, $p < .01$, $d = 1.10$; 4th grade: $M = 8.62\%$ vs. 19.31%, $t(15) = 5.24$, $p < .01$, $d = .94$), whereas 1st graders showed significantly greater error on the small compared to the large scale ($M = 21.42\%$ vs. 12.72%), $t(16) = 3.52$, $p < .01$, $d = 1.17$. Mean gamma correlations for all grades and conditions were positive, and three values were significantly different than 0 (see Table 4). Overall, these values replicate Experiment 1, indicating children have some ability to monitor their performance at the item level.

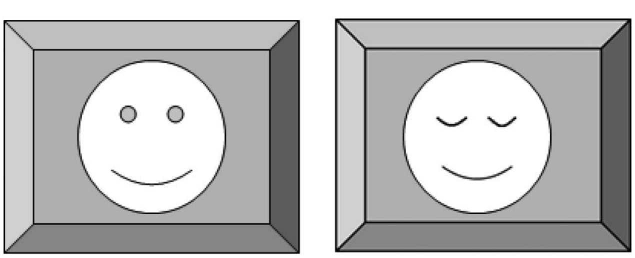

Figure 2. Open-Eyes and Closed-Eyes boxes used for the sorting task in Experiment 2.

Task-level control. To assess task-level control, we calculated the number of children that placed the small scale in the open-eyes box (see Table 5). Children selected the small scale above chance regardless of grade level (1st grade: $t(16) = 2.50$, $p < .05$; 2nd grade, $t(20) = 2.75$, $p = .01$; 4th grade, $t(15) = 4.39$, $p < .01$). Of the children showing the expected representational pattern, all 1st graders and the majority of 2nd and 4th graders selected the small scale. Finally, of the children that possessed an advanced representation for both scales, more than half of 1st and 2nd graders placed the small scale into the open-eyes box, but only 4th graders selected this scale above chance, $t(8) = 3.50$, $p = .01$. This finding is consistent with our task-level monitoring results.

Item-level control. Children sorted a total of 36 estimates into one of two boxes. Children were told that estimates placed in the open-eyes box would be evaluated for a reward. Overall, children sorted significantly more estimates into the open-eyes box than expected by chance (chance = 18): 1st grade: $M = 22.94$, $SD = 8.50$, $t(16) = 2.40$, $p < .05$; 2nd grade: $M = 24.52$, $SD = 6.10$, $t(20) = 4.90$, $p < .01$; 4th grade: $M = 22.19$, $SD = 4.75$; $t(15) = 3.53$, $p < .01$. The average number of open-eyes estimates did not differ by grade, $F(2, 53) = .61$, $p = .55$, or scale, $F(1, 51) = 1.85$, $p = .18$ (1st grade: $M = 11.59$, $SD = 4.50$ vs. $M = 11.35$, $SD = 4.61$; 2nd grade: $M = 12.57$, $SD = 2.93$ vs. $M = 11.95$, $SD = 3.47$; 4th grade: $M = 11.50$, $SD = 3.12$ vs. $M = 10.69$, $SD = 2.87$).

Item-level control was assessed in two ways. First, we calculated each participant's mean confidence for the estimates placed in each box (see Table 6). For each scale, we conducted two separate 2 (box placement: eyes-open, eyes-closed) × 3 (grade: 1st, 2nd, 4th) ANOVAs. For the small scale, there was a main effect of box placement, $F(1, 45) = 37.40$, $p < .01$, $\eta^2 = .42$, where estimates in the open-eyes box were associated with higher confidence compared to estimates in the closed-eyes box, and an interaction that approached significance, $F(2, 45) = 3.06$, $p = .06$, $\eta^2 = .07$. No main effect of grade was found, $F(2, 45) = .97$, $p >$

Table 5
Task-Level Control (Experiment 2)

Grade	Number of Children Who Put Small Scale In Open-Eyes Box		
	Overall	Advanced-Less Advanced	Advanced-Advanced
1st	13/17 (76.5%)*	7/7 (100%)†	4/7 (57%)
2nd	16/21 (76%)**	9/11 (82%)*	6/9 (67%)
4th	14/16 (88%)**	5/6 (83%)	8/9 (89%)**

† Due to the lack of variance, a statistic could not be calculated.
* $p < .05$. ** $p = .01$.

.05. The large scale showed a similar pattern. A main effect of box placement, $F(1, 48) = 53.39$, $p < .01$, $\eta^2 = .49$, and a marginally significant interaction were found, $F(2, 48) = 3.32$, $p = .05$, $\eta^2 = .06$, while the main effect of grade was not significant, $F(2, 48) = 2.05$, $p > .05$.

To assess item-level control, gamma correlations were computed between box placement (open-eyes = "1"; closed-eyes = "0") and PAE (left-hand columns of Table 7), and between box placement and confidence (right-hand columns of Table 7). Box placement was related to confidence more strongly than accuracy. A 2 (measure: confidence correlation, PAE correlation) × 2 (scale: small, large) × 3 (grade: 1st, 2nd, 4th) ANOVA revealed a main effect of measure: $F(1, 40) = 89.84$, $p < .01$, $\eta^2 = .53$, but no main effect of grade, $F(2, 40) = 1.06$, $p > .05$, suggesting that children's control behavior is based more on their perceived, rather than actual, performance. None of the two- or three-way interactions were significant.

Discussion

Experiment 2 replicated and extended Experiment 1; children showed some ability to monitor and control their estimation performance. Because item-level control judgments were made retrospectively (and not immediately after each magnitude estimate had been made), it is possible that these judgments were driven less by introspective evaluations of performance and more by a simple heuristic (e.g., "high confidence estimates go in the open-eyes box; low confidence estimates go in the closed-eyes box"). Our methodological decision to investigate children's monitoring and control judgments in separate blocks was necessary for assessing task-level awareness and for providing evidence of direct replication; however, this decision necessarily introduced a time delay between children's monitoring and control decisions.

Experiment 3

To address this issue, children in Experiment 3 made both a monitoring and control judgment immediately after each estimate was made. Moreover, only the large-scale number line was used; this scale is associated with more variability in underlying representation (i.e., individual children are likely to show either an *advanced* or *less-advanced* representation), providing the best chance for detecting this awareness. As a result, only item-level judgments were collected in the current experiment. Furthermore, this revised methodology no longer required that children make item-level confidence and monitoring decisions embedded within an easy (small numerical range) versus more difficult (large numerical range) context. In this way, the small versus large numeric context would not provide cues to amplify differences in children's item-level monitoring.

Table 6
Item-Level Control

Grade	Mean Confidence Judgments: Small EO	Small EC	Large EO	Large EC
Experiment 2				
1st	2.79 (.23)	2.13 (.52)	2.44 (.46)	1.95 (.50)
2nd	2.52 (.31)	2.29 (.53)	2.58 (.30)	2.26 (.44)
4th	2.46 (.33)	2.07 (.41)	2.31 (.43)	2.09 (.43)
			Large EO	Large EC
Experiment 3				
1st			2.67 (.29)	2.15 (.54)
2nd			2.37 (.40)	1.85 (.54)
4th			2.43 (.31)	1.69 (.38)

Note. EO = Eyes-Open box; EC = Eyes-Closed box; Standard deviations in parentheses.

Table 7
Item-Level Control

Grade	Mean Gamma Correlations: Box Placement PAE Small	Box Placement PAE Large	Box Placement Confidence Small	Box Placement Confidence Large
Experiment 2				
1st	.31 (.43)*	.18 (.34)	.64 (.61)**	.56 (.35)**
2nd	.08 (.41)	.34 (.40)**	.35 (.67)*	.56 (.37)**
4th	.11 (.33)	.37 (.42)**	.51 (.48)**	.38 (.58)*
		Large		Large
Experiment 3				
1st		.21 (.35)*		.50 (.65)*
2nd		.10 (.32)		.61 (.53)**
4th		.27 (.34)**		.94 (.13)**

Note. Higher PAE represents a greater percentage of error. Thus, gammas were reverse coded for ease of interpretation; Standard deviations indicated in parentheses.
* $p < .05$. ** $p < .01$.

Participants

Eighty-five children participated and were recruited in a similar manner as the previous experiments; parental permission slips were distributed to all children in two classrooms at each grade level. The sample was primarily Caucasian (69%) and consisted of 18 1st graders (8 males; mean age = 6.72 years, *SD* = 6.63 months), 31 2nd graders (13 males; mean age = 7.89 years, *SD* = 4.90 months), and 36 4th graders (16 males; mean age = 9.91 years, *SD* = 3.86 months). Each child received two stickers for participating.

Tasks and Procedure

The procedure was identical to Experiment 2 except for two changes. First, because children only received the large-scale number line packet, they made a total of 18 rather than 36 estimates. Second, item-level control judgments were made immediately after item-level monitoring judgments. That is, after children made their estimate and reported their confidence, they were asked, "*You marked N here and were not so sure/kind of sure/really sure. Do you want to put this in the open-eyes or closed-eyes box?*"

Results

As in Experiments 1 and 2, children were classified as either *advanced* or *less advanced*. Half (50%) of 1st graders who made estimates on the 0–100 scale and the majority (69%) of 4th graders who made estimates on the 0–100,000 scale were classified as "advanced," whereas the majority (84%) of 2nd graders who made estimates on the 0–1,000 scale were "less advanced."

Item-level monitoring. Mean confidence and PAE are reported in Tables 2 and 3, respectively. Mean gamma correlations between confidence and PAE are displayed in Table 4; all are positive and two are significantly different than 0. These findings replicate Experiments 1 and 2, providing further evidence that children have some ability to monitor their performance at the item level.

Item-level control. Children sorted significantly more estimates into the open-eyes box than expected by chance (chance = 9): 1st grade: $M = 13.67$, $SD = 3.87$, $t(17) = 5.12$, $p < .01$; 2nd grade: $M = 13.61$, $SD = 3.40$, $t(30) = 7.55$, $p < .01$; 4th grade: $M = 13.28$, $SD = 3.36$; $t(35) = 7.64$, $p < .01$. The average number of open-eyes estimates did not differ by grade, $F(2, 84) = .11$, $p > .05$.

Mean confidence was computed for the estimates placed in each box (see Table 6). A 2 (box placement: open-eyes, closed-eyes) × 3 (grade: 1st, 2nd, 4th) ANOVA revealed a main effect of box placement, $F(1, 66) = 109.83$, $p < .01$, $\eta^2 = .61$, and a main effect of grade, $F(2, 66) = 5.07$, $p = .01$, $\eta^2 = .13$, but no significant interaction, $F(2, 66) = 2.08$, $p > .05$. Replicating Experiment 2, estimates in the open-eyes box were associated with higher confidence compared to estimates in the closed-eyes box.

Gamma correlations were computed between box placement and PAE (left-hand columns of Table 7), and between box placement and confidence (right-hand columns of Table 7). A 2 (measure: confidence correlation, PAE correlation) × 3 (grade: 1st, 2nd, 4th) ANOVA revealed a main effect of measure: $F(1, 66) = 148.01$, $p < .01$, $\eta^2 = .65$, and a significant interaction, $F(2, 66) = 6.31$, $p < .01$, $\eta^2 = .06$, but no main effect of grade, $F(2, 66) = 2.64$, $p > .05$. Replicating Experiments 1 and 2, box placement was more strongly related to confidence than accuracy.

Discussion

Experiment 3 revealed two important findings. First, children showed an ability to control their performance at the item level even when control decisions were made immediately after they placed numbers on the number line. Second, children's control decisions were most heavily influenced by their prior monitoring judgment. This finding replicates Experiment 2 and suggests that control decisions are based more on perceived, rather than actual, performance.

General Discussion

The current experiments examined children's ability to monitor and control their performance during number line estimation. These metacognitive abilities are essential for effective self-regulated learning and may be particularly important for mastering mathematics. At the task level, children who are sensitive to the difficulty of a particular scale may spend more time thinking about their estimates or adopt a strategy that results in accurate performance. Similarly, item-level awareness might encourage children to ask for help, withhold potentially inaccurate estimates, or retry estimates they are less certain about. Children may use the estimates they are most confident about to help estimate more difficult items. Thus, awareness at either level has the potential to enhance overall performance.

With respect to monitoring, even our youngest age group accurately monitored their estimation performance at both the task and item level. Hembacher and Ghetti (2014) reported early *memory* monitoring skills in children as young as 4 years of age (see also Lipowski et al., 2013). Perhaps with an age-appropriate task (e.g., magnitude comparison), even preschoolers could monitor their math performance. The mental processes children monitor during estimation remain unclear. Children may be sensitive to the difficulty of a task or of individual items, perhaps by monitoring the speed of their performance. Consistent with this idea, Siegler et al. (2011) found that children were slower to estimate the location of fractions as compared to whole numbers. If children are sensitive to the relative difficulty of fractions, their longer response times may reflect a strategy for reducing their chances of making errors.

Alternatively, children may be sensitive to their experience with numbers for each scale. For example, 1st graders may have reported feeling more confident in their small-scale estimates because numbers within this range are more familiar. Similarly, 4th graders may have reported feeling less confident in their large-scale estimates because these numbers are less common and thus seem less familiar. Two of our findings suggested that task-level awareness may arise from monitoring familiarity rather than difficulty. First, children who were advanced on both scales reported higher confidence on the small scale. Second, in Experiment 2, 1st graders reported higher confidence on the small scale even though it was associated with greater error. Thus, familiarity may have greater influence on children's judgments than accuracy of performance.

With regard to control, Experiments 2 and 3 suggested that although this skill may be present in elementary school, it is far from perfect and may continue to develop throughout childhood and adolescence. Children's control behavior was driven by confidence more so than actual performance (see Table 7). Because children were reminded of their confidence judgment immediately before sorting, they may have been more likely to consider this information when making control decisions. Future research should examine whether children (1) engage in spontaneous control processes without explicit reminders of their confidence judgments, and (2) make similar control decisions when the introspective process is eliminated (e.g., making control decisions about other children's estimates). Moreover, although Experiment 3 demonstrated that control decisions were not influenced by their temporal proximity to confidence judgments, we cannot rule out the possibility that these decisions were based on a heuristic (e.g., "all items with low confidence should be disregarded") rather than introspective processes. Future research should encourage strategy reporting to identify such heuristics. Alternatively, interspersing items from the small and large number line scales may shed light on whether children simply represent smaller number lines as "easier" than larger ones. Finally, Experiment 2 revealed one unexpected finding: About half of 1st graders possessed an advanced representation for the small scale, yet showed greater error (PAE) for these estimates compared to those on the large scale. High PAE in the 0–10 range has been reported in other studies (Berteletti et al., 2010; Lanfranchi, Berteletti, Torrisi, Vianello, & Zorzi, 2015). Although it does not directly affect our results, future work will need to replicate and examine the cause of this high PAE.

One limitation of our study is the relatively small sample size at each grade level within any given experiment. However, we found consistent support for metacognitive awareness on a number line estimation task across three experiments that included a total of 198 participants. A second limitation is that asking children to rate their confidence after every estimate could impact children's subsequent estimates. That is, it is possible that children used their confidence judgments as a type of feedback that informed more careful future estimates. This positive impact of judgments on estimation accuracy would have important applied implications if it does occur. This issue has not been addressed by prior work (e.g., Hembacher & Ghetti, 2014; Lyons & Ghetti, 2013), so to evaluate this provocative possibility, additional empirical work could compare the estimation performance of children who do and do not make immediate confidence judgments after each estimate.

The current experiments also have educational implications. Elementary schoolchildren are sensitive to math tasks and items that are more challenging, so teachers could capitalize on this awareness by encouraging children to spend more time thinking about their answers or to ask questions when in doubt. This type of control behavior may promote insights into effective strategies. For example, second graders overestimate the location of 150 on a number line ranging from 0–1,000. When second graders received feedback about the correct location of 150, this feedback improved their estimation accuracy (Opfer & Siegler, 2007; Opfer & Thompson, 2008; Thompson & Opfer, 2008). Seeking feedback on difficult problems when needed may be supported by children's metacognitive awareness and may promote a more advanced understanding of numerical magnitude.

References

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14,* 125–135. http://dx.doi.org/10.1111/j.1467-7687.2010.00962.x

Barth, H., Slusser, E., Cohen, D., & Paladino, A. (2011). A sense of proportion: Commentary on Opfer, Siegler, and Young. *Developmental Science, 14,* 1205–1206. http://dx.d0oi.org/10.1111/j.1467-7687.2011.01081.x

Barth, H., Starr, A., & Sullivan, J. (2009). Children's mappings of large number words to numerosities. *Cognitive Development, 24,* 248–264. http://dx.doi.org/10.1016/j.cogdev.2009.04.001

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology, 46,* 545–551. http://dx.doi.org/10.1037/a0017887

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42,* 189–201. http://dx.doi.org/10.1037/0012-1649.41.6.189

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79,* 1016–1031. http://dx.doi.org/10.1111/j.1467-8624.2008.01173.x

Chesney, D. L., & Matthews, P. G. (2013). Knowledge on the line: Manipulating beliefs about the magnitudes of symbolic numbers affects the linearity of line estimation tasks. *Psychonomic Bulletin & Review, 20,* 1146–1153. http://dx.doi.org/10.3758/s13423-013-0446-8

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review, 18,* 331–338. http://dx.doi.org/10.3758/s13423-011-0059-z

Cohen, D. J., & Sarnecka, B. W. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology, 50,* 1640–1652. http://dx.doi.org/10.1037/a0035901

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. http://dx.doi.org/10.1111/desc.12271

Dackermann, T., Huber, S., Bahnmueller, J., Nuerk, H.-C., & Moeller, K. (2015). An integration of competing accounts on children's number line estimation. *Frontiers in Psychology, 6,* 884. http://dx.doi.org/10.3389/fpsyg.2015.00884

Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology, 47,* 274–296. http://dx.doi.org/10.1016/0022-0965(89)90033-7

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Beverly Hills, CA: Sage.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new era of cognitive–developmental inquiry. *American Psychologist, 34,* 906–911. http://dx.doi.org/10.1037/0003-066X.34.10.906

Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology, 1,* 324–340. http://dx.doi.org/10.1016/0010-0285(70)90019-8

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33,* 277–299. http://dx.doi.org/10.1080/87565640801982361

Gettinger, M. (1985). Time allocated and time spent relative to time needed for learning as determinants of achievement. *Journal of Educational Psychology, 77,* 3.3–11.

Hembacher, E., & Ghetti, S. (2014). Don't look at my answer: Subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. *Psychological Science, 25,* 1768–1776. http://dx.doi.org/10.1177/0956797614542273

Huber, S., Moeller, K., Nuerk, H. C., & Willmes, K. (2013). A computational modeling approach on three-digit number processing. *Topics in Cognitive Science, 5,* 317–334. http://dx.doi.org/10.1111/tops.12016

Hurst, M., Monahan, K. L., Heller, E., & Cordes, S. (2014). 123s and ABCs: Developmental shifts in logarithmic-to-linear responding reflect fluency with sequence values. *Developmental Science, 17,* 892–904. http://dx.doi.org/10.1111/desc.12165

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490–517. http://dx.doi.org/10.1037/0033-295X.103.3.490

Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive Science, 37,* 775–799. http://dx.doi.org/10.1111/cogs.12028

Lanfranchi, S., Berteletti, I., Torrisi, E., Vianello, R., & Zorzi, M. (2015). Numerical estimation in individuals with Down syndrome. *Research in Developmental Disabilities, 36,* 222–229.

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development, 78,* 1723–1743. http://dx.doi.org/10.1111/j.1467-8624.2007.01087.x

Leal, L., Crays, N., & Moely, B. E. (1985). Training children to use a self-monitoring study strategy in preparation for recall: Maintenance and generalization effects. *Child Development, 56,* 643–653. http://dx.doi.org/10.2307/1129754

Link, T., Nuerk, H.-C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *The Quarterly Journal of Experimental Psychology, 67,* 1597–1613. http://dx.doi.org/10.1080/17470218.2014.892517

Lipowski, S. L., Merriman, W. E., & Dunlosky, J. (2013). Preschoolers can make highly accurate judgments of learning. *Developmental Psychology, 49,* 1505–1516. http://dx.doi.org/10.1037/a0030614

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

Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman-Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 509–527. http://dx.doi.org/10.1037/a0014876

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109–133. http://dx.doi.org/10.1037/0033-2909.95.1.109

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55,* 169–195. http://dx.doi.org/10.1016/j.cogpsych.2006.09.002

Opfer, J. E., Siegler, R. S., & Young, C. J. (2011). The powers of noise-fitting: Reply to Barth and Paladino. *Developmental Science, 14,* 1194–1204. http://dx.doi.org/10.1111/j.1467-7687.2011.01070.x

Opfer, J. E., & Thompson, C. A. (2008). The trouble with transfer: Insights from microgenetic changes in the representation of numerical magnitude. *Child Development, 79,* 788–804. http://dx.doi.org/10.1111/j.1467-8624.2008.01158.x

Opfer, J. E., Thompson, C. A., & Kim, D. (2016). Free versus anchored numerical estimation: A unified approach. *Cognition, 149,* 11–17. http://dx.doi.org/10.1016/j.cognition.2015.11.015

Rips, L. J. (2013). How many is a zillion? Sources of number distortion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1257–1264. http://dx.doi.org/10.1037/a0031143

Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: Their interrelations in grades 5 and 6. *Journal of Educational Psychology, 101,* 359–372. http://dx.doi.org/10.1037/a0013840

Schneider, W., & Loffler, E. (2016). The development of metacognitive knowledge in children and adolescents. In J. Dunlosky & S. Tauber (Eds.), *The Oxford handbook of metamemory*. Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199336746.013.10

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75,* 428–444. http://dx.doi.org/10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14,* 237–250. http://dx.doi.org/10.1111/1467-9280.02438

Siegler, R. S., & Thompson, C. A. (2014). Numerical landmarks are useful—Except when they're not. *Journal of Experimental Child Psychology, 120,* 39–58. http://dx.doi.org/10.1016/j.jecp.2013.11.014

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education, 3,* 143–150. http://dx.doi.org/10.1111/j.1751-228X.2009.01064.x

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62,* 273–296. http://dx.doi.org/10.1016/j.cogpsych.2011.03.001

Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General, 142,* 193–208. http://dx.doi.org/10.1037/a0028560

Thompson, C. A., & Opfer, J. E. (2008). Costs and benefits of representational change: Effects of context on age and sex differences in symbolic magnitude estimation. *Journal of Experimental Child Psychology, 101,* 20–51. http://dx.doi.org/10.1016/j.jecp.2008.02.003

Thompson, C. A., & Opfer, J. E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development, 81,* 1768–1786. http://dx.doi.org/10.1111/j.1467-8624.2010.01509.x

Thompson, C. A., & Siegler, R. S. (2010). Linear numerical-magnitude representations aid children's memory for numbers. *Psychological Science, 21,* 1274–1281. http://dx.doi.org/10.1177/0956797610378309

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in education theory and practice, 93,* 27–30.

Appendix

The To-be-Estimated Numbers From Each Scale

0–10: 1–9 (each numeral appeared twice)
0–100: 2, 5, 9, 12, 15, 16, 17, 18, 25, 34, 37, 49, 56, 61, 72, 78, 82, 94
0–1000: 2, 5, 9, 17, 34, 56, 78, 122, 150, 163, 179, 246, 366, 486, 606, 722, 818, 938
0–100,000: 200, 500, 900, 1700, 3400, 5600, 7800, 12200, 15000, 16300, 17900, 24600, 36600, 48600, 60600, 72200, 81800, 93800

Received August 31, 2015
Revision received June 2, 2016
Accepted June 9, 2016 ■