American Psychological Association

Developmental Psychology

0012-1649/19/$12.00

2019, Vol. 55, No. 7, 1389–1399
http://dx.doi.org/10.1037/dev0000739

The Role of Inhibitory Control in Strategy Change: The Case of Linear Measurement

Kexin Ren
Temple University

Ying Lin
University of Rochester

Elizabeth A. Gunderson
Temple University

Elementary school students often lack a conceptual understanding of linear measurement, which is revealed by their poor performance when the object to be measured is not aligned with the start of the ruler. Instead of correctly counting the units that correspond to the object (e.g., inches or centimeters), children often use 1 of 2 incorrect strategies: reading off the number that corresponds to the end of the object (the least-mature, read-off strategy) and counting the hash marks that flank the object (a more mature, but still incorrect, hash-mark strategy). We hypothesized that shifting to a more mature linear measurement strategy would require the ability to inhibit less-mature prepotent responses, such as read-off and hash-mark responses. In the present study, we predicted that children with better inhibitory control would be more likely to improve in their linear measurement strategy use over one year. Participants ($n = 317$) were in 1st through 3rd grades when they completed a linear measurement task that required measuring objects that were not aligned with the start of the ruler; they also completed an inhibitory control task and control measures (visuospatial working memory, arithmetic calculations, and number line estimation). One year later, they repeated the linear measurement task. Students with higher initial inhibitory control were more likely to adopt a more mature strategy over time. Moreover, inhibitory control was a significant predictor of strategy improvement over and above other cognitive measures, including visuospatial working memory and arithmetic calculation skill.

Keywords: linear measurement, inhibitory control, mathematical cognition, executive function

Children's ability to perform mathematics is correlated with executive functions (EF; Brock, Rimm-Kaufman, Nathanson, & Grimm, 2009). Inhibitory control (IC), the ability to suppress prepotent responses (Diamond, 2013), is an important EF skill that itself correlates with mathematics achievement (for a review, see Bull & Scerif, 2001). However, it is important to move beyond broad correlations between IC and mathematics to identify specific mathematical skills and strategies for which IC is particularly important. Theoretically, strong IC skills may help children suppress the use of a less-sophisticated but perceptually salient, well-practiced strategy in favor of a less-familiar but more adaptive strategy (Cragg & Gilmore, 2014). One important mathematical area in which children have difficulty transitioning beyond a perceptually salient, incorrect strategy is linear measurement (Solomon, Vasilyeva, Huttenlocher, & Levine, 2015). Children find linear measurement to be challenging in elementary school and this difficulty persists in middle school and high school, especially when measuring objects that are misaligned with the ruler (Kamii, 2006). We propose that this difficulty may be due to the inability to inhibit perceptually salient, well-practiced responses.

Linear measurement is important in mathematical and spatial domains in which a solid conceptual background contributes to a clear understanding of geometry, spatial relations, and real numbers (Clements, 1999). Furthermore, strong conceptual knowledge of linear measurement is important in mathematical areas that are imperative for scientific reasoning, such as mass and relational measures (for a review, see Lehrer, 2003). Although children are taught about linear measurement at a young age, their conceptual understanding is often shallow (Kamii, 2006). According to the NAEP and TIMSS, measurement content is especially difficult, and U.S. students in the 4th and 8th grade performed poorly in measurement compared to other countries (Wilson, Blank, & Rolf,

This article was published Online First April 22, 2019.

Kexin Ren, Department of Psychology, Temple University; Ying Lin, Department of Brain and Cognitive Sciences, University of Rochester; Elizabeth A. Gunderson, Department of Psychology, Temple University.

Kexin Ren and Ying Lin contributed equally.

This research was supported by NSF CAREER DRL-1452000 to Elizabeth A. Gunderson. We thank the children and teachers who gave their time to this research and the research assistants who helped carry it out: Synclaire Arthur, Conner Black, Jennifer Bradley, Riley Brown, Yesha Dave, Olivia Dermody, John Durison, Lillian Ham, Lindsey Hildebrand, Maya Johnson, Jenny Jung, Christin Kim, Nicole Lee, Lexi Liu, Victoria McLaughlin, Samantha McLaughlin, Sumaiya Nusrath, Jessica Palmarini, Gabriella Riccardo, Jaeyong Sung, Brittany Worthington, Audrey Wrobel, and Najah Young.

Correspondence concerning this article should be addressed to Elizabeth A. Gunderson, Department of Psychology, Temple University, 1701 North 13th Street, Philadelphia, PA 19122. E-mail: liz.gunderson@temple.edu

1999). In fact, middle school students' measurement performance has remained low since the 1990s (Thompson & Preston, 2004). One explanation for this persistent poor performance may be the way linear measurement is taught in schools. Classroom instruction and curriculum materials usually concentrate on measurement procedures but not concepts (Kamii & Clark, 1997). Typically, children are taught to align an object with the left edge of the ruler (the 0-point mark) and read off the number that corresponds to the right end of the object. Teachers are guided to correct students if they do not align the ruler and the object (Kamii, 2006). Although this read-off procedure yields correct responses, many children only gain procedural knowledge from the activity, without conceptual understanding of spatial units (Clements & Stephan, 2004). Because this procedural strategy yields correct responses, it is difficult for teachers to recognize their students' lack of conceptual understanding.

Students' lack of conceptual knowledge of linear measurement is revealed when the ruler and the object are misaligned (referred to as a "shifted ruler problem"), for example, when a crayon that is 3 units long extends from the 2-point mark to the 5-point mark of the ruler (see Figure 1). This is a persistent problem even among older students: on this type of measurement problem, a nationally representative sample found that fewer than one fourth of 4th-grade students chose the correct answer, a rate that rose only to two thirds of 8th graders and four-fifths of 12th graders (Kenney & Silver, 1997). When asked to measure objects that are misaligned with the ruler, students tend to use one of two incorrect strategies: the read-off strategy (reading the number that corresponds to the right end of the object; Figure 1d) and the hash-mark strategy (counting the hash marks that flank the object; Figure 1c; Solomon et al., 2015). Although the hash-mark strategy yields an incorrect response, it is more advanced than the read-off strategy; counting hash marks requires noticing that the object is not aligned with the start of the ruler and that counting strategies are needed. Further evidence that the hash-mark strategy is more advanced than the read-off strategy comes from a recent study that trained first graders on shifted ruler problems (Congdon, Kwon, & Levine, 2018). They found that some students who predominately used the read-off strategy during pretest moved to using the hash-mark strategy after training. In contrast, none of the students switched from the hash-mark to the read-off strategy, which suggests a progression of strategies, with the read-off strategy being the least sophisticated, the hash-mark strategy being more advanced, and the correct strategy (counting units) being the most advanced.

In sum, the evidence to date suggests that, on the shifted ruler task, a task designed to tap conceptual understanding of linear measurement, children's strategies progress from the read-off strategy to the more-sophisticated (but still incorrect) hash-mark strategy, to the correct counting-units strategy (Congdon et al., 2018; Solomon et al., 2015). Having identified this trajectory, it is important to provide (to our knowledge) the first longitudinal picture of this trajectory and to investigate how strategy change occurs in this domain.

Prior research on the development of linear measurement skill has focused on the importance of gradually coordinating multiple measurement concepts, including concepts of spatial units, partitioning, iteration, and the inverse relation between a unit's length and the number of units (Lehrer, 2003; Stephan & Clements, 2003). We suggest that, in addition to coordinating these important concepts, development also requires the ability to inhibit old strategies to successfully apply more advanced strategies. This proposal draws on overlapping waves theory (Chen & Siegler, 2000), which argues that the strategy a child uses is dependent on characteristics of the individual, the situation, and the current problem. Importantly, when a new strategy emerges, it often coexists with preexisting strategies, such that multiple strategies are available to an individual at any given time. When individuals gain more relevant experience and conceptual knowledge, they are

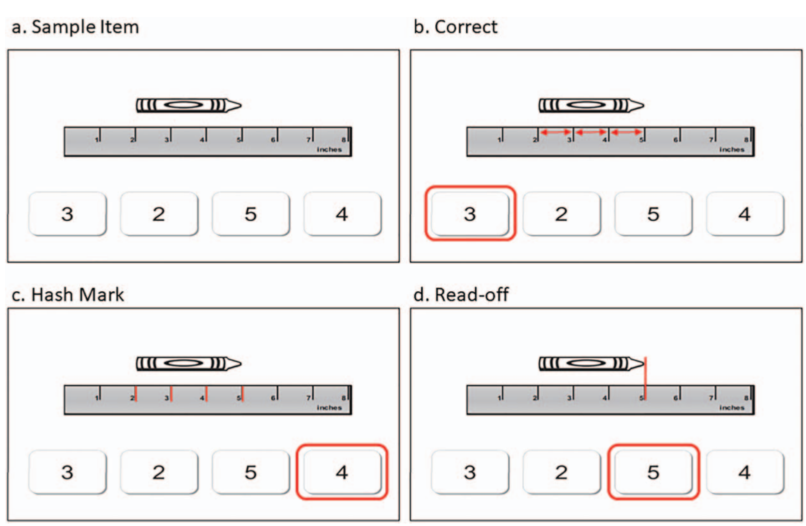

Figure 1. Shifted ruler task sample item. The task as presented is shown in Panel a. Panels b–d illustrate each strategy: (b) the correct answer is 3, (c) the hash mark answer is 4, and (d) the read-off answer is 5. Adapted from "Minding the gap: Children's difficulty conceptualizing spatial intervals as linear measurement units," by Solomon, Vasilyeva, Huttenlocher, and Levine, 2015, p. 1567. Copyright, 2015 by the American Psychological Association or one of its allied publishers. See the online article for the color version of this figure.

more likely to choose more sophisticated strategies; however, less-sophisticated strategies often remain available, even to adults. We suggest that correctly selecting a more-sophisticated strategy when an incorrect, less-sophisticated strategy remains cognitively available requires IC. This should be the case especially when the less-sophisticated strategy is both well-practiced and easy to use. In the context of linear measurement, the read-off strategy has these qualities: it is well-practiced in schools, and very easy to determine by simply reading the number that aligns with the end of the to-be-measured object. In fact, it is unlikely that the read-off strategy ever disappears from an individual's strategy repertoire in linear measurement, given that it is the correct strategy when the object is aligned with the zero mark. Given these properties of the shifted ruler task, we expect that, to move beyond the read-off strategy to use the hash-mark or counting-units strategies, children must inhibit the prepotent read-off response. Further, although the hash-mark strategy is more advanced than the read-off strategy, it may also have the property of being well-practiced and easy to determine, given that counting objects (i.e., hash-marks) is a highly practiced skill among children (much more so than counting spaces between objects). Thus, to select the counting-units strategy, we expect that children must inhibit both the hash-mark and read-off strategies.

To test this hypothesis, we examined whether elementary-school children with stronger IC would be more likely to transition to a more-sophisticated strategy on the shifted ruler linear measurement task over time. We focus on linear measurement because the understanding of linear measurement is essential for later higher mathematics and for everyday life (Congdon et al., 2018; Lehrer, Jaslow, & Curtis, 2003). We note that we are not arguing that linear measurement is the only aspect of mathematics to be impacted by IC. Rather, linear measurement provides a useful case in which to test the role of IC in promoting strategy change in mathematics, because it is a case in which progress to a more advanced strategy ought to involve the inhibition of prepotent responses. This is a novel approach, as few previous studies have examined specific links between IC and mathematical strategy change, and to our knowledge, none have focused on strategy change over time in linear measurement and whether IC is predictive of strategy change in this domain.

We also sought to determine the specificity of the predicted relation between IC and linear measurement by examining whether other cognitive skills (visuospatial working memory) and mathematical skills (arithmetic calculation and number line estimation) predicted improvement in linear measurement strategies, over and above IC. Although working memory and IC are both components of EF, they are still two separate processes (Begolli et al., 2018). Our analyses controlled for working memory to parse apart the variance that is specific to IC, over and above other EF processes.

Furthermore, we investigated whether IC would predict linear measurement improvement over and above general mathematics skills. Because arithmetic calculation skills are a focus on mathematics instruction in elementary school, they can serve as a reasonable proxy for general school mathematics learning (Bottge, 1999). Further, exact arithmetic calculations typically involve memorized math facts and procedures that are specific to arithmetic and different from the concepts and procedures needed for linear measurement. Thus, we chose an arithmetic calculation task that involves exact calculations (the Woodcock-Johnson IV Calculation subtest), and we expected that arithmetic skill would not strongly relate to strategy improvement in linear measurement. Importantly, if IC predicts linear measurement skills over and above arithmetic calculation, this provides support for the proposition that IC is not serving as a proxy for general math skill at this age.

Finally, we examined whether number line estimation would predict improvement in linear measurement strategies, and whether IC would predict measurement improvement over and above number line estimation. Linear measurement has some similarities to number line estimation (Siegler & Opfer, 2003), as both involve connecting numbers to horizontal extent in a linear manner, with the smaller numbers on the left and larger numbers on the right. The number line task, however, taps into estimation skills, an understanding of the magnitudes of large numbers, and the ability to place a target number proportionally relative to other numbers on the line (Barth & Paladino, 2011; Opfer & Siegler, 2007). In contrast, success on the shifted ruler linear measurement task is related to a conceptual understanding that measurement involves spatial units, the ability to count small numbers of units, and the ability to inhibit perceptually salient, learned strategies such as the read-off strategy (Newcombe, Levine, & Mix, 2015). Because of these differences in the conceptual and procedural bases of the two tasks, we expected that number line estimation skills would not predict improvement in linear measurement skill over time. However, we note that it is possible that, because both number line estimation and linear measurement involve left-to-right spatial-numeric relations, performance may be related between these tasks; thus, we consider our prediction of no relation to be exploratory. Importantly, we hypothesized that IC would predict improvement in linear measurement skill over and above number line estimation.

Because IC is necessary to inhibit prepotent responses, our main hypothesis was that IC would predict improvement in children's dominant linear measurement strategy over time. We focus on children's dominant strategy based on prior research showing that children's responses typically show a predominant strategy that is used consistently throughout the task (Deitz, Huttenlocher, Kwon, Levine, & Ratliff, 2009). We also hypothesized that the relation between IC and linear measurement strategy would be robust even after controlling for other potential predictors (visuospatial working memory, arithmetic calculation, and number line estimation). We examined this hypothesis among children in 1st through 3rd grades, based on prior work showing that 2nd graders have substantial difficulties with shifted ruler problems (Solomon et al., 2015). Our 1-year longitudinal design allows us to establish predictive relations (not just concurrent relations) between initial levels of IC and changes in children's linear measurement strategies over time. This research has both theoretical and practical import, with the potential to inform theories of why IC relates to math learning and to inform educational practices in the teaching of linear measurement.

Method

Participants

Participants were students (N = 317, 174 females) from 11 schools in a large city in the Eastern United States. Students were in 1st through 3rd grades at the start of the study (117 first graders, 108 second graders, and 92 third graders; age: M = 8.02 years, *SD* = .91, *n* = 304). Parents reported their child's race and ethnicity (*n* = 278); children were 45.7% Black or African American, 18.0% White, 10.1% Hispanic, 3.2% Asian or Asian American, 0.6% American Indian or Alaskan Native, 0.3% Native Hawaiian or other Pacific Islander, 9.5% multiple race/ethnicities, and 0.3% other race/ethnicities. The highest education received by either parent averaged 14.56 years (*SD* = 2.34, *n* = 282; ranging from less than high school to graduate degree). The average annual income reported ranging from $35,000 to 49,999 (*n* = 262; ranging from less than $15,000 to more than $100,000).

From a larger longitudinal study, we included all participants who were enrolled in the study by spring of Year 1, were in 1st through 3rd grades in Year 1,[1] and who completed at least one of the following measures: linear measurement (Year 1 spring, Year 2 spring), IC (Year 1 spring), calculation (Year 1 spring), number line estimation (Year 1 spring), and visuospatial working memory (Year 1 spring). The maximum sample included 317 students, and sample sizes for each task are shown in Table 2. The main aim of this study was not associated with the original aim of the larger longitudinal study, which was to examine links between spatial skills and numeracy. Further, our key measures of interest, linear measurement and IC, were added to the original longitudinal study to address additional questions.

Procedure

We examined two time points from a larger longitudinal study: Year 1 spring and Year 2 spring, on average one year apart. The larger study included 4 time points (Year 1 fall, Year 1 spring, Year 2 fall, and Year 2 spring). At each time point, participants completed a battery of tasks assessing cognitive skills and motivation. Participants completed these tasks during two one-on-one sessions with an experimenter lasting about 30 min each, in one of four pseudorandom orders.

IC was only assessed in Year 1 spring. Linear measurement, number line estimation, arithmetic calculation, and visuospatial working memory were each assessed at all 4 time points. Because IC was only measured in Year 1 spring, we considered Year 1 spring to be the initial time-point for the purposes of the present study. Therefore, we analyzed performance on each predictor (IC) and control measure (visuospatial working memory, arithmetic calculation, and number line estimation) as assessed at Year 1 spring. We examined change in linear measurement from Year 1 spring to Year 2 spring to give students the longest time interval possible to improve in linear measurement strategies, given their overall poor performance.

The study procedures were approved under Temple University IRB protocol 21935, "Cognitive and Emotional Bases of Math, Reading, and Spatial Development."

Measures

Linear measurement. The shifted ruler task assesses children's strategy use and conceptual understanding of linear measurement (Levine, Kwon, Huttenlocher, Ratliff, & Deitz, 2009). On each item, children were presented with an image of a ruler with a crayon misaligned with the start of the ruler (see Figure 1) and were asked to judge how many units long the crayon was by choosing one of four choices. The four choices were the correct answer, a read-off answer (the number that corresponds to the end of the crayon), a hash-mark answer (the number of hash marks that flank the object), and a foil. The length of the crayon ranged from 3 units to 8 units long. The task had 10 trials.

For each child at each time point, we calculated the percentage of trials on which they selected each type of response (correct, read-off, hash mark, or foil). If the child selected one response type on 6 or more of the trials, we categorized the child as using that strategy at that time point (e.g., if children chose the hash-mark answer 6 or more times in one test, they were categorized as using the hash-mark strategy). This cutoff was chosen because on a 10-item test with four response types, a child choosing the same response type on six or more items is unlikely to occur by chance based on the binomial distribution ($p < .05$; criteria also used by Congdon et al., 2018). We also asked children (*n* = 180) to describe how they got their answer to one shifted ruler item in spring of Year 2. As described in Appendix A, children's self-reported strategy responses were closely aligned with their strategy categorization based on the 10-item multiple-choice task, further validating this method of categorizing children's strategy use.

Children without a dominant strategy were excluded from our main analyses (9.8% of children who completed the task in Year 1 spring, 7.5% in Year 2 spring. Additional analyses that included these participants, as a check on the robustness of the main result, also showed that IC was a significant predictor of counting-based strategies on the linear measurement task; see Appendix B. (We note that the analyses in Appendix B examine a slightly different question—whether IC predicts counting-based responses in the linear measurement task—which differs slightly from our main research question regarding changes in children's dominant strategy.) Based on prior research (Levine et al., 2009), we considered the strategies to be ordered from most to least mature as follows: correct, hash mark, read-off, foil. Comparing children's strategies in Year 1 spring and Year 2 spring, we coded participants' strategy change as improved (strategy changed from less to more mature), no-change (stayed the same), or got-worse (changed to less mature strategy), and we analyzed this as an ordinal variable.

There were 25.5% of participants (*N* = 275) who only completed the linear measurement task in Year 1. They did not significantly differ from participants who completed the task in both years (*n* = 205) in terms of parental education, family income, IC, working memory performance, arithmetic calculation, or number line estimation.

IC. We measured IC using the hearts and flowers task (Wright & Diamond, 2014), administered using E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002). On each trial, children were presented with either a heart or flower on the left or right side of the screen. Children were asked to press a key on the same side as the heart and on the opposite side as the flower. A fixation cross was presented for 500 ms before each stimulus, and the stimulus disappeared after the participant responded. There were four practice trials with feedback, and children were expected to respond correctly to three out of four practice trials to move on to the test

[1] The larger study included children in prekindergarten through third grades in Year 1. However, the tasks varied by grade level, and only first through third graders were given our focal task assessing linear measurement.

phase; if not, participants repeated the practice block once. There were three test blocks: a congruent hearts-only block (12 trials), an incongruent flowers-only block (12 trials), and a mixed block (32 trials) that included both hearts and flowers in a fixed pseudorandom order.

For our main analyses, we used accuracy on the switch trials within the mixed block (i.e., flower trials on which the previous trial was a heart, and vice versa), because these switch trials place the greatest demand on IC (Davidson, Amso, Anderson, & Diamond, 2006). Accuracy on mixed block switch trials (17 trials) was our primary index of IC because in prior work young children had a larger difference between switch trials and nonswitch trials on accuracy than on response time (RT; Davidson et al., 2006). Five participants were excluded due to software errors. We excluded trials with response times that were too fast [<200 ms] or too slow [≥10 s] from all analyses (1.37% of switch trials, 1.90% of hearts block trials, and 1.37% of flowers block trials). We calculated mean accuracy separately for switch trials, hearts block trials, and flowers block trials. We also calculated RT for correct switch trials and excluded RTs that were 3 standardized residuals beyond the participant's mean of each block as outliers (2.19% of switch trials).

Calculation. Children completed the Woodcock-Johnson IV Calculation subtest (Schrank, Mather, & McGrew, 2014), a nationally normed measure in which children complete written arithmetic problems of increasing difficulty (e.g., addition, subtraction, multiplication, and division). Basal and ceiling criteria were met if children completed the first six items correctly and the last six items incorrectly ($n = 25$ excluded because these criteria were not established due to experimenter error). We used the W score, a Rasch-scaled score that is appropriate for analyses of individual differences (McGrew, LaForte, & Schrank, 2014).

Number line estimation. Participants completed the 0–1,000 number line estimation task (Siegler & Opfer, 2003). This task was administered on the computer (Fazio, Bailey, Thompson, & Siegler, 2014) and participants were asked to click a location on the number line that corresponded to the number above the line (e.g., "Where does 131 go on the number line?"). There were 18 items (3, 19, 103, 158, 240, 297, 391, 438, 475, 525, 562, 609, 703, 760, 842, 897, 981, 997), and item order was randomized. We calculated the percent absolute error (PAE): (|correct location—participants' estimate|)/number line range. PAE is a widely used measure of accuracy in number line estimation (Siegler, Thompson, & Schneider, 2011; Simms, Clayton, Cragg, Gilmore, & Johnson, 2016).

Visuospatial working memory. Children completed the Dot Matrix task from the Automated Working Memory Assessment (Alloway et al., 2005). Children saw a series of 4 × 4 grids on the computer screen. On each trial, a red dot appeared sequentially in one or more squares of the grid and then disappeared. Children were asked to point to the location(s) of the dot. After successfully completing one dot, the task increases in difficulty by increasing the number of dots in sequential order. This task contains three practice trials with feedback and terminates after three wrong answers within a block. We analyzed the child's standard score, normed based on the child's age group.

Analytic Approach

We first provide descriptive statistics and relations between variables, then conduct a series of ordinal logistic regression to test our main hypothesis, that IC predicts linear measurement strategy change, over and above other covariates.

We examined the pattern of missing data using Little's (1988) MCAR test in SPSS missing values analysis. This test was nonsignificant ($p = .496$), indicating that missingness can be characterized as missing completely at random (MCAR). We handled missing data using listwise deletion, which is appropriate when

Table 1
Frequencies of Linear Measurement Strategy Use and Strategy Change Within Each Gender and Grade

Strategy	Total ($n = 176$)	Gender: Female ($n = 97$)	Gender: Male ($n = 79$)	Year 1 grade: First ($n = 62$)	Year 1 grade: Second ($n = 65$)	Year 1 grade: Third ($n = 49$)
Year 1 spring dominant strategy						
Correct	4.5%	5.2%	3.8%	3.2%	6.2%	4.1%
Hash mark	20.5%	17.5%	24.1%	22.6%	23.1%	14.3%
Read off	75.0%	77.3%	72.2%	74.2%	70.8%	81.6%
Year 2 spring dominant strategy						
Correct	6.8%	2.1%	12.7%	8.1%	6.2%	6.1%
Hash mark	29.0%	29.9%	27.8%	25.8%	33.8%	26.5%
Read off	64.2%	68.0%	59.5%	66.1%	60.0%	67.3%
Strategy change						
Improvement	18.2%	15.5%	21.5%	17.7%	20.0%	16.3%
No change	76.1%	76.3%	75.9%	74.2%	72.3%	83.7%
Got worse	5.7%	8.2%	2.5%	8.1%	7.7%	.0%

Note. Percentages are within either gender or grade. Dominant strategy was the answer choice the student selected on 60% or more of the items. Students with no dominant strategy at either time point (i.e., no answer choice was selected on 60% or more of the items) are excluded from this table ($n = 21$ had no dominant strategy in Year 1 spring only, $n = 12$ in Year 2 spring only, and $n = 4$ at both time points). Only one participant consistently chose the foil response (in Year 2 spring only), but that participant was already excluded because they had no dominant strategy in Year 1 spring. Four participants used the correct strategy at both time points and were excluded from the main regression analyses but are included here.

Table 2
Descriptive Statistics and Correlations Among Key Variables

Variable	*M* (*SD*)	*n*	1	2	3	4	5	6	7	8	9
1. Mixed block switch trial percent accuracy	.77 (.18)	267	—								
2. Mixed block switch trial RT (ms)	1,150.06 (396.72)	267	$.36^{***}$	—							
3. Hearts (congruent) block accuracy	.97 (.08)	267	$.14^{*}$	.04	—						
4. Flowers (incongruent) block accuracy	.85 (.21)	267	$.48^{***}$	$.26^{***}$	$.20^{**}$	—					
5. Visuospatial WM standardized score	94.79 (15.53)	270	$.16^{*}$	−.03	$.13^{*}$	$.22^{**}$	—				
6. Calculation W score	455.48 (19.48)	249	$.32^{***}$	$-.16^{*}$	.01	$.23^{***}$	$.20^{**}$	—			
7. NL estimation PAE	.21 (.09)	269	$-.40^{***}$	−.06	$-.16^{*}$	$-.35^{***}$	$-.30^{***}$	$-.39^{***}$	—		
8. Linear measurement strategy change	.13 (.47)	176	$.21^{**}$	.12	.02	.08	−.08	.06	$-.16^{*}$	—	
9. Gender (0 = female, 1 = male)	N/A	317	.04	$-.13^{*}$	.04	.08	.01	−.09	$-.21^{**}$	.12	—
10. Grade	1.92 (.81)	317	$.22^{***}$	$-.19^{**}$	.07	$.15^{*}$	.12	$.76^{***}$	$-.27^{***}$	.06	−.01

Note. RT = response time; WM = working memory; NL = number line; PAE = percent absolute error. Missing data by task was due to factors including attrition from to Year 1 spring to Year 2 spring, student absence on the day of testing, student refusal to complete the task, computer error, and experimenter error.
* $p < .05$. ** $p < .01$. *** $p < .001$.

data are MCAR. We used listwise deletion separately for each analysis; therefore, the sample size varies by analysis.

We also examined the extent to which clustering of children within schools and classrooms accounted for variance in our key dependent variable, strategy change from Year 1 to Year 2 (an ordinal variable). We tested this using HLM 7.0 (Raudenbush & Bryk, 2002), with separate two-level models to examine nesting within the school level and classroom level. The ICC (intraclass correlation coefficient) estimate for classroom was .088 and the ICC estimate for school was .070. Because ICCs were lower than 10%, indicating that less than 10% of the variance was due to clustering of students within classrooms or schools, multilevel modeling was not required, and for the sake of simplicity we opted to use ordinal logistic regression (generalized linear model; Vajargah & Nikbakht, 2015).

Results

Preliminary Analyses

Table 1 shows frequencies of linear measurement strategy use at each time point, as well as strategy change, by grade and gender. Performance was generally poor on this task, with 60% or more of participants consistently using the least mature read-off strategy at each time point. However, students did improve over time, with 18.2% of the participants improving to a more mature strategy from Year 1 spring to Year 2 spring. Treating dominant strategy as an ordinal variable, there was significant overall change in strategy from Year 1 spring to Year 2 spring (Wilcoxon signed-ranks test, $Z = -3.00$, $p = .003$). Among the participants who improved ($n = 32$), 75.0% improved from the read off to the hash mark strategy, 9.4% improved from the read off to the correct strategy, and 15.6% improved from the hash mark to the correct strategy. Among the participants who got worse ($n = 10$), 60.0% changed from the hash mark to the read off strategy, 20.0% changed from the correct to the read off strategy, and 20.0% changed from the correct to the hash mark strategy.

However, grade level was not significantly associated with more advanced strategy use in Year 1 spring, $r_s(176) = -.05$, $p = .478$, or Year 2 spring, $r_s(176) = .01$, $p = .909$, nor to strategy change from Year 1 spring to Year 2 spring, $r_s(172) = .052$, $p = .502$. Gender was also not related to dominant strategy in Year 1 spring (Mann–Whitney $U = 3655.5$, $p = .488$) or in Year 2 spring ($U = 3381.0$, $p = .112$), nor was gender related to whether students' strategy changed from Year 1 spring to Year 2 spring, $\chi^2(2) = 3.42$, $p = .181$.

Descriptive statistics and correlations among all measures are shown in Table 2. Switch trial accuracy, which we used to index IC, was significantly correlated with all other variables, suggesting that it was a more stable and reliable indicator of IC than other variables such as switch trial RT. Switch trial accuracy was also the only variable from the hearts and flowers task that was significantly correlated with linear measurement strategy change.

Main Analyses

We tested our main hypotheses using strategy change from Year 1 spring to Year 2 spring as the outcome variable in ordinal logistic regression models (see Table 3). Model 1 included switch trial accuracy, grade, and gender as predictors of linear measurement strategy change (see Table 3). Switch trial accuracy ($B = 2.71$, $SE = 1.10$, $p = .013$) and gender ($B = -.74$, $SE = .38$, $p = .050$) significantly predicted measurement strategy change from Year 1 spring to Year 2 spring, but grade ($B = .07$, $SE = .23$, $p = .778$) did not.[2] Furthermore, when we controlled for visuospatial working memory (Table 3, Model 2), mixed block switch trial accuracy remained a significant predictor of linear measurement strategy change ($B = 3.43$, $SE = 1.22$, $p = .005$), and gender was also a significant predictor ($B = -.97$, $SE = .41$, $p = .019$), whereas visuospatial working memory was not ($B = -.03$, $SE = .01$, $p = .053$). This suggests that the results are specific to IC rather than EF more generally.

To control for children's attentiveness to the hearts and flowers task, we examined whether accuracy on the hearts block (congruent trials) or flowers block (incongruent trials) would predict change over time in linear measurement strategies, when mixed

[2] We also ran the regression using mixed block switch trial RT instead of accuracy. Mixed block switch trial RT was not a significant predictor of change in linear measurement ($B = .00$, $SE = .00$, $p = .067$), consistent with our expectation that accuracy would be a better measure of IC at this age.

Table 3
Ordinal Logistic Regressions Predicting Linear Measurement Strategy Change

Variable	Child strategy change on linear measurement task from Y1 to Y2					
	Unstandardized parameter estimate (*SE*)					
	Model 1 (*n* = 159)	Model 2 (*n* = 143)	Model 3 (*n* = 159)	Model 4 (*n* = 159)	Model 5 (*n* = 140)	Model 6 (*n* = 149)
Grade	.07 (.23)	.07 (.25)	.07 (.23)	.07 (.23)	.38 (.38)	.10 (.25)
Gender	−.74 (.38)*	−.97* (.41)	−.75* (.38)	−.77* (.38)	−.72 (.42)	−.64 (.40)
Switch accuracy	**2.71* (1.10)**	**3.43* (1.22)**	**2.71* (1.10)**	**2.93* (1.21)**	**2.53* (1.26)**	2.15 (1.26)
Spatial WM		−.03 (.01)				
Hearts (congruent block) accuracy			−.09 (1.91)			
Flowers (incongruent block) accuracy				−.44 (1.04)		
Calculation (W score)					−.01 (.02)	
NL estimation (PAE)						−1.94 (2.81)

Note. Y1 = Year 1 spring; Y2 = Year 2 spring; WM = working memory; NL = number line; PAE = percent absolute error. Bold type indicates significance of the key predictor (switch accuracy).
* $p \leq .05$.

block switch trial accuracy was also included in the model. We entered hearts block accuracy and flowers block accuracy into two separate models, controlling for grade, gender, and mixed block switch trial accuracy. Hearts block (congruent) accuracy ($B = -.09$, $SE = 1.91$, $p = .962$) was not a significant predictor, whereas switch trial accuracy ($B = 2.71$, $SE = 1.10$, $p = .014$) and gender ($B = -.75$, $SE = .38$, $p = .050$) remained significant (Table 3, Model 3). In a separate model (Table 3, Model 4), flowers block (incongruent) accuracy ($B = -.44$, $SE = 1.04$, $p = .669$) was not a significant predictor, and mixed block switch trial accuracy ($B = 2.93$, $SE = 1.21$, $p = .016$) and gender ($B = -.77$, $SE = .38$, $p = .046$) remained significant. These results suggest that the relation between IC and changes in strategy use was specific to switch trial accuracy.

Next, we examined whether other mathematical skills could explain the relation between IC and measurement strategy change. In a model with both calculation and switch trial accuracy (Table 3, Model 5), calculation was not a significant predictor of linear measurement change ($B = -.01$, $SE = .02$, $p = .521$), and switch trial accuracy remained significant ($B = 2.53$, $SE = 1.26$, $p = .044$). Finally, in a model with both number line estimation PAE and switch trial accuracy as predictors (Table 3, Model 6), number line estimation was not a significant predictor ($B = -1.94$, $SE = 2.81$, $p = .491$), however switch trial accuracy was trending toward significance ($B = 2.15$, $SE = 1.26$, $p = .089$).[3]

Discussion

The relation between EF and mathematics has been studied using broad measures, yet advancing theory and practice requires understanding specific links between EF components and specific mathematical concepts. We predicted that one component of EF—IC—would be particularly important in mathematical contexts that require inhibiting a salient incorrect strategy in favor of a more advanced strategy, such as in the case of linear measurement. Understanding the development of linear measurement is important because it is widely used in everyday life (e.g., carpentry, making clothing, and determining distances while traveling) and forms an essential foundation for higher mathematics such as geometry, ratios, and fractions (National Council of Teachers of Mathematics, 1989).

Longitudinally over one year, higher initial IC was related to children's later tendency to avoid perceptually salient responses (directly reading off the number on the ruler that aligns with the endpoint of the object) and move on to use more cognitively taxing counting strategies (counting hash marks or spatial units). Furthermore, this relation was specific to IC, even when considering another EF factor, visuospatial working memory, confirming that IC—rather than general EF—uniquely predicted change in linear measurement strategies over time.

We also investigated whether IC specifically predicted strategy change when accounting for other measures of attention and performance on the same task (hearts and flowers; Wright & Diamond, 2014). We examined IC (accuracy on switch trials in the mixed-trial block), accuracy on a block of only congruent trials (which involves memory maintenance), and accuracy on a block of only incongruent trials (which involves memory maintenance and a light load of inhibition). Only IC significantly predicted linear measurement strategy change, and this result held even after controlling for accuracy on the congruent-only and incongruent-only blocks. This is consistent with previous studies, which have found that switch trials are the most demanding and require the greatest IC in the hearts and flowers task (Davidson et al., 2006). Further, this shows that individual differences in the ability to deal with spatial incongruency alone are unable to explain children's shift toward more adaptive linear measurement strategies.

To assess whether IC was serving as a proxy for general math achievement, we also controlled for arithmetic calculation and

[3] We also tested number line estimation linear R^2 and log R^2 as a predictor in Model 6 separately (replacing number line estimation PAE). Both linear R^2 and log R^2 were not significant predictors ($ps \geq .063$), whereas switch trial accuracy in both models remained significant ($ps \leq .035$). Gender and grade were not significant predictors in either model ($ps \geq .077$). We note that because logarithmic performance is characterized by overestimation of small numerosities, it is possible that our task included only two numbers in the 0-100 range may have contributed to the non-significant results related to log R^2.

number line estimation. When controlling for arithmetic calculation, IC was significant in predicting linear measurement strategy change, whereas arithmetic calculation was not a significant predictor, confirming our hypotheses. However, unexpectedly, after adding number line estimation to the regression model, neither IC nor number line estimation was a significant predictor. Given the high correlation between number line estimation and IC, it is possible that number line estimation partitioned the variance in the outcome variable that was originally explained by IC. This is supported by previous work showing that individual differences in IC predicted children's quality of number line estimation (Laski & Dulaney, 2015). Furthermore, the trending relation of IC to measurement strategy change when controlling for number line estimation may be due to insufficient power. To further understand the role of IC as a unique predictor of linear measurement concepts, future studies could include a larger sample size and probe the relation between linear measurement and number line estimation.

In addition to investigating the role of IC in predicting linear measurement, our study is the first, to our knowledge, to longitudinally investigate linear measurement strategies across one year. However, the majority of the students in our dataset (60% or more, even at the oldest grade level) still merely followed a procedural routine of reading off the number on the ruler that corresponded to the end of the object. They failed to recognize the concept of spatial intervals of the linear measurement, and instead overgeneralized the procedural rule they were taught in school to complete the items. About one quarter of students used the more-advanced (but still incorrect) hash-mark strategy, and fewer than 10% were consistently correct. Surprisingly, in our study of 1st through 3rd graders, students in higher grade levels did not use significantly more sophisticated strategies than those in lower grade levels. However, students did use significantly more advanced strategies at the second time point than the first, regardless of their grade level. Taken together, this hints that students may not have learned about concepts of linear units in school, but that mere exposure to our task, which showed objects that were misaligned with the ruler, may have led to improvement. This would be consistent with experimental work in which students improved in linear measurement when given experience solving misaligned ruler items, but not aligned items (Congdon et al., 2018), and with the fact that standard curricular approaches use only aligned items (Kamii, 2006). It is also consistent with our finding that less than 10% of the variance in strategy improvement in our study was related to the child's school or classroom. However, it would nevertheless be informative for future studies to survey teachers or directly observe classroom practices to assess whether or not students had classroom experiences with misaligned linear measurement items.

Moreover, our main analyses revealed that boys were more likely to improve their linear measurement strategies than girls. In our sample, boys also outperformed girls in number line estimation, and the gender difference in linear measurement strategy change was reduced to nonsignificance when we controlled for number line estimation skill. This may suggest that a gender difference in spatial-numeric associations, identified previously in adults' number line estimation (Bull, Cleland, & Mitchell, 2013), may also extend to linear measurement, a possibility that could be investigated in future studies.

These results set the stage for additional studies of linear measurement strategy change. Given the poor performance even among 3rd graders in our study, future research can follow students until the end of elementary school. This will provide insight into when students begin to fully understand linear measurement, and at which point in their schooling they typically switch to a more mature strategy.

These results also set the stage for additional studies unpacking the relation between IC and mathematics. A close analysis of other mathematical domains and skills is expected to reveal additional cases in which IC ought to be especially related to performance. For example, research done by Lubin, Vidal, Lanoë, Houdé, and Borst (2013) found that middle-school children needed to inhibit a misleading strategy in order to perform well on word problems involving language that was inconsistent with the operation to be performed (e.g., solving a word problem using the words "more than" and requiring the operation "subtraction"; Lubin et al., 2013). Similarly, learning about fractions and decimals often requires overcoming whole-number biases (Hurst & Cordes, 2016; Lortie-Forgues, Tian, & Siegler, 2015; Rinne, Ye, & Jordan, 2017), which may also require IC. Future research could test whether individual differences in IC predict improvement in these areas of mathematics, and also provide negative test cases in areas of mathematics that are not expected to require IC. Our study provides a first step toward understanding the circumstances in which IC promotes mathematical thinking, and much more work is needed to fully understand this relation.

The correlational design of this study prevents us from drawing strong conclusions about whether IC causes change in linear measurement strategy use, although we included several key control measures to reduce the possibility that this relation was explained by confounds such as general EF. Nevertheless, it will be important for future research to use experimental designs to determine whether improving students' IC—either in the short-term or in the long-term—leads to improvements in their linear measurement strategies.

Above all, as the first study that longitudinally examined factors that affect children's change in conceptual knowledge of linear measurement, we provide strong evidence that IC plays a prominent role in predicting this conceptual gain. Our results also indicate that although students begin learning about linear measurement as young as 1st grade, most still lack essential conceptual knowledge of linear units even at the end of 3rd grade. This can provide impetus for teachers to focus more on spatial units when using rulers, and to provide examples of measuring objects that are not aligned with the start of the ruler, to help students understand linear measurement conceptually and not just procedurally.

References

Alloway, T. P., Gathercole, S. E., Adams, A.-M., Willis, C., Eaglen, R., & Lamont, E. (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. *British Journal of Developmental Psychology, 23,* 417–426. http://dx.doi.org/10.1348/026151005X26804

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

Begolli, K. N., Richland, L. E., Jaeggi, S. M., Lyons, E. M., Klostermann, E. C., & Matlen, B. J. (2018). Executive function in learning mathematics by comparison: Incorporating everyday classrooms into the science of learning. *Thinking & Reasoning, 24,* 280–313. http://dx.doi.org/10.1080/13546783.2018.1429306

Bottge, B. A. (1999). Effects of contextualized math instruction on problem solving of average and below-average achieving students. *The Journal of Special Education, 33,* 81–92. http://dx.doi.org/10.1177/002246699903300202

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of "hot" and "cool" executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly, 24,* 337–349. http://dx.doi.org/10.1016/j.ecresq.2009.06.001

Bull, R., Cleland, A. A., & Mitchell, T. (2013). Sex differences in the spatial representation of number. *Journal of Experimental Psychology: General, 142,* 181–192. http://dx.doi.org/10.1037/a0028387

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19,* 273–293. http://dx.doi.org/10.1207/S15326942DN1903_3

Chen, Z., & Siegler, R. S. (2000). II. Overlapping waves theory. *Monographs of the Society for Research in Child Development, 65,* 7–11. http://dx.doi.org/10.1111/1540-5834.00075

Clements, D. H. (1999). Teaching length measurement: research challenges. *School Science and Mathematics, 99,* 5–11. http://dx.doi.org/10.1111/j.1949-8594.1999.tb17440.x

Clements, D. H., & Stephan, M. (2004). Measurement in preK to grade 2 mathematics. In D. H. Clements, J. Sarama & A.-M. DiBiase (Eds.), *Engaging young children in mathematics standards for early childhood mathematics education* (pp. 299–317). Mahwah, NJ: Lawrence Erlbaum Associates.

Congdon, E. L., Kwon, M.-K., & Levine, S. C. (2018). Learning to measure through action and gesture: Children's prior knowledge matters. *Cognition, 180,* 182–190. http://dx.doi.org/10.1016/j.cognition.2018.07.002

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education, 3,* 63–68. http://dx.doi.org/10.1016/j.tine.2013.12.001

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44,* 2037–2078. http://dx.doi.org/10.1016/j.neuropsychologia.2006.02.006

Deitz, K., Huttenlocher, J., Kwon, M.-K., Levine, S. C., & Ratliff, K. (2009). Children's understanding of ruler measurement and units of measure: A training study. *Proceedings of the Annual Meeting of the Cognitive Science Society, 31,* 2391–2395. Retrieved from https://escholarship.org/uc/item/5cz2r7vj

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64,* 135–168. http://dx.doi.org/10.1146/annurev-psych-113011-143750

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology, 123,* 53–72. http://dx.doi.org/10.1016/j.jecp.2014.01.013

Hurst, M., & Cordes, S. (2016). Rational-number comparison across notation: Fractions, decimals, and whole numbers. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 281–293. http://dx.doi.org/10.1037/xhp0000140

Kamii, C. (2006). Measurement of length: How can we teach it better? *Teaching Children Mathematics, 13,* 154–158.

Kamii, C., & Clark, F. B. (1997). Measurement of length: The need for a better approach to teaching. *School Science and Mathematics, 97,* 116–121. http://dx.doi.org/10.1111/j.1949-8594.1997.tb17354.x

Kenney, P. A., & Silver, E. A. (1997). *Results from the sixth mathematics assessment of the national assessment of educational progress*. Reston, VA: National Council of Teachers of Mathematics.

Laski, E. V., & Dulaney, A. (2015). When prior knowledge interferes, inhibitory control matters for learning: The case of numerical magnitude representations. *Journal of Educational Psychology, 107,* 1035–1050. http://dx.doi.org/10.1037/edu0000034

Lehrer, R. (2003). Developing understanding of measurement. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 179–192). Reston, VA: National Council of Teachers of Mathematics.

Lehrer, R., Jaslow, L., & Curtis, C. L. (2003). Developing an understanding of measurement in the elementary grades. In D. H. Clements & G. W. Bright (Eds.), *Learning and teaching measurement: 2003 yearbook* (pp. 100–121). Reston, VA: National Council of Teachers of Mathematics.

Levine, S. C., Kwon, M.-K., Huttenlocher, J., Ratliff, K., & Deitz, K. (2009). Children's understanding of ruler measurement and units of measure: A training study. *Proceedings of the Annual Meeting of the Cognitive Science Society, 31,* 2391–2395.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83,* 1198–1202. http://dx.doi.org/10.1080/01621459.1988.10478722

Lortie-Forgues, H., Tian, J., & Siegler, R. S. (2015). Why is learning fraction and decimal arithmetic so difficult? *Developmental Review, 38,* 201–221. http://dx.doi.org/10.1016/j.dr.2015.07.008

Lubin, A., Vidal, J., Lanoë, C., Houdé, O., & Borst, G. (2013). Inhibitory control is needed for the resolution of arithmetic word problems: A developmental negative priming study. *Journal of Educational Psychology, 105,* 701–708. http://dx.doi.org/10.1037/a0032625

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Woodcock-Johnson IV Technical Manual*. Rolling Meadows, IL: Riverside.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Newcombe, N. S., Levine, S. C., & Mix, K. S. (2015). Thinking about quantity: The intertwined development of spatial and numerical cognition. *WIREs Cognitive Science, 6,* 491–505. http://dx.doi.org/10.1002/wcs.1369

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

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rinne, L. F., Ye, A., & Jordan, N. C. (2017). Development of fraction comparison strategies: A latent transition analysis. *Developmental Psychology, 53,* 713–730. http://dx.doi.org/10.1037/dev0000275

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Pittsburgh, PA: Psychology Software Tools.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Achievement*. Rolling Meadows, IL: Riverside.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14,* 237–243. http://dx.doi.org/10.1111/1467-9280.02438

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

Simms, V., Clayton, S., Cragg, L., Gilmore, C., & Johnson, S. (2016). Explaining the relationship between number line estimation and mathematical achievement: The role of visuomotor integration and visuospatial skills. *Journal of Experimental Child Psychology, 145,* 22–33. http://dx.doi.org/10.1016/j.jecp.2015.12.004

Solomon, T. L., Vasilyeva, M., Huttenlocher, J., & Levine, S. C. (2015). Minding the gap: Children's difficulty conceptualizing spatial intervals as linear measurement units. *Developmental Psychology, 51,* 1564–1573. http://dx.doi.org/10.1037/a0039707

Stephan, M., & Clements, D. H. (2003). Linear and area measurement in prekindergarten to Grade 2. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement* (pp. 3–16). Reston, VA: National Council of Teachers of Mathematics.

Thompson, T. D., & Preston, R. V. (2004). Measurement in the middle grades: Insights from NAEP and TIMSS. *Mathematics Teaching in the Middle School, 9,* 514–519.

Vajargah, K. F., & Nikbakht, M. (2015). Application REML model and determining cut off of ICC by multi-level model based on Markov Chains simulation in health. *Indian Journal of Fundamental and Applied Life Sciences, 5,* 1432–1448.

Wilson, L. D., Blank, R. K., & Rolf, K. (1999). *Improving mathematics education using results from NAEP and TIMSS* (No. 978–1-884037–56-6). Washington, DC: Council of Chief State School Officers.

Wright, A., & Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in Psychology, 5,* 213. http://dx.doi.org/10.3389/fpsyg.2014.00213

Appendix A

Correspondence Between Children's Self-Reported Strategies and Behavioral Responses on the Shifted Ruler Task

Children were given one shifted ruler item in Year 2 spring and were asked to explain how they figured out their answer. The experimenter typed the child's verbal response in real-time. We created a coding scheme that included: whether the child mentioned the end of the crayon, whether the child mentioned hash marks, whether the child mentioned spaces and whether the child mentioned counting. We categorized these self-reported strategy responses as being consistent with (a) the correct strategy (mentioned spaces and counting), (b) the hash mark strategy (mentioned hash marks and counting but not spaces), or (c) the read off strategy (mentioned the end of the crayon but did not mention counting). We compared the self-reported strategy with children's behavioral responses (Table A1). Children's self-reported strategy was consistent with their choices in the shifted ruler task, especially for hash-mark and read-off strategies, which validated the effectiveness of our method of categorizing their strategy use in the main text (i.e., if they chose that strategy-consistent answer on six or more of 10 items).

Table A1
Comparison Between Children's Self-Report Strategy and Behavioral Response in Year 2

Self-reported strategy category in Year 2 (*n* = 180)	Example transcript	Strategy category from behavioral data in Year 2: Correct (*n* = 14)	Hash mark (*n* = 49)	Read off (*n* = 116)	Other (foil; *n* = 1)
Correct (*n* = 2)	"Count how many spaces it took up"	50.0%	50.0%	0%	0%
Hash mark (*n* = 21)	"Counted each line the crayon was on"	19.0%	81.0%	0%	0%
Read off (*n* = 69)	"When you look at the tip of the crayon and put your finger down, it goes to 4"	1.4%	4.3%	92.8%	1.4%
Unclassifiable (*n* = 88)	"Just measured it"	9.1%	31.8%	59.1%	0%

Note. Percentages are within each self-reported strategy category in Year 2. Students with no dominant strategy are excluded from this table (*n* = 12).

(*Appendices continue*)

Appendix B

Exploratory Analyses of Inhibitory Control Predicting Counting-Based Strategies on the Shifted Ruler Task

Table B1

Binomial Generalized Linear Model Predicting Linear Measurement Counting-Based Strategy Use in Year 2 Spring

	Child counting-based responses on linear measurement task in Year 2 spring					
	Unstandardized parameter estimate (*SE*)					
Variable	Model 1 (n = 192)	Model 2 (n = 173)	Model 3 (n = 192)	Model 4 (n = 192)	Model 5 (n = 168)	Model 6 (n = 178)
Y1 linear measurement counting-based responses	$.30^{***}$ (.01)	$.30^{***}$ (.02)	$.30^{***}$ (.01)	$.29^{***}$ (.01)	$.29^{***}$ (.02)	$.28^{***}$ (.02)
Grade	−.03 (.07)	.02 (.08)	−.03 (.07)	−.03 (.07)	$-.26^{*}$ (.12)	−.16 (.08)
Gender	$-.30^{**}$ (.11)	$-.51^{***}$ (.12)	$-.29^{*}$ (.11)	$-.28^{*}$ (.11)	$-.37^{**}$ (.13)	−.09 (.12)
Switch accuracy	**2.94^{***} (.32)**	**3.29^{***} (.35)**	**2.87^{***} (.32)**	**2.56^{***} (.34)**	**2.35^{***} (.37)**	**1.26^{***} (.36)**
Spatial WM		−.01 (.00)				
Hearts (congruent block) accuracy			1.61^{*} (.80)			
Flowers (incongruent block) accuracy				1.03^{**} (.35)		
Calculation (W score)					$.02^{***}$ (.01)	
NL estimation (PAE)						-7.75^{***} (.99)

Note. Counting-based responses include items where the child selected the hash-mark or correct response. Y1 = Year 1 spring; Y2 = Year 2 spring; WM = working memory; NL = number line; PAE = percent absolute error. Bold type indicates significance of the key predictor (switch accuracy).
* $p < .05$. ** $p < .01$. *** $p < .001$.

To better understand the role of inhibitory control in predicting children's strategy use across time, we also conducted exploratory analyses. In our main analyses, we categorized children based on their dominant linear measurement strategy at each time point. However, this required excluding a number of children who did not have a dominant response strategy at one or both time points. To include data from all children who completed the linear measurement tasks, we examined the frequency of counting-based strategies (i.e., correct or hash-mark strategies) on the measurement task in Year 2 spring, controlling for the frequency of those strategies in Year 1 spring. We chose to combine correct and hash-mark responses because the frequency of correct responses alone was very low (Year 1 spring: M = 0.77 trials [of 10], SD = 1.81; Year 2 spring: M = 1.07, SD = 2.36). When correct and hash-mark strategies were combined, these counting-based responses accounted for about 30% of trials in Year 1 spring and 40% of trials in Year 2 spring (Year 1 spring: M = 2.89 trials [of 10], SD = 3.95; Year 2 spring: M = 3.90, SD = 4.48).

Because the frequency of children's use of these counting-based strategies was not normally distributed, we used a binary logistic regression (generalized linear model) where the dependent variable was the number of shifted ruler items (out of 10) on which children used counting-based strategies in Year 2 spring. We included grade, gender, Year 1 spring shifted ruler trials with a counting-based strategy, and mixed block switch trial accuracy (our measure of IC) as predictors in the simultaneous regression model (n = 192; Table B1, Model 1). Results showed that children with better inhibitory control were significantly more likely to use counting-based strategies one year later (B = 2.94, SE = .32, $p <$ .001) after controlling for frequency of counting-based strategies in Year 1 spring (B = .30, SE = .01, $p < .001$). Gender was also a significant predictor of Year 2 spring counting-based strategies (B = −.30, SE = .11, p = .009), such that boys (estimated marginal mean = 43.55%, SE = 2.04%) were more likely to use counting-based strategies than girls (estimated marginal mean = 36.46%, SE = 1.76%). Grade was not a significant predictor (p = .727). In additional models controlling for spatial WM, hearts accuracy, flowers accuracy, calculation, and number line estimation (Table B1, Models 2–6), switch trial accuracy remained a significant predictor of counting-based strategies in Year 2 spring, although the control variables were also significant in each case.

Received June 28, 2018
Revision received March 1, 2019
Accepted March 14, 2019 ■