Journal of Cognition and Development

ISSN: 1524-8372 (Print) 1532-7647 (Online) Journal homepage: https://www.tandfonline.com/loi/hjcd20

# Assessing Young Children's Number Magnitude Representation: A Comparison Between Novel and Conventional Tasks

Erin E. Reid, Arthur J. Baroody & David J. Purpura

JOURNAL OF COGNITION AND DEVELOPMENT, 16(5):759–779

ISSN: 1524-8372 print/1532-7647 online
DOI: 10.1080/15248372.2014.920844

# Assessing Young Children's Number Magnitude Representation: A Comparison Between Novel and Conventional Tasks

Erin E. Reid
*Erikson Institute*

Arthur J. Baroody
*University of Illinois at Urbana-Champaign*

David J. Purpura
*Purdue University*

Previously, researchers have relied on asking young children to plot a given number on a 0-to-10 number line to assess their mental representation of numbers 1 to 9. However, such a (''conventional'') number-to-position (N-P) task may underestimate the accuracy of young children's magnitude estimates and misrepresent the nature of their number representation. The purpose of this study was to compare young children's performance on the conventional N-P task and a ''modified'' N-P task that is more consistent with a discrete-quantity view of number and with measures of theoretically related mathematical competencies. Participants ($n = 45$), ranging in age from 4;0 to 6;0, were administered both versions of the N-P task twice during 4 sessions in 1 of 2 randomly assigned and counterbalanced orders. Between and within conditions, children were significantly more accurate on the modified version than on the conventional task. The results indicate that the conventional task, in particular, may be confusing and that several simple modifications can make it more understandable for young children. However, when performance on theoretically related number tasks is taken into account, both the conventional and the modified N-P tasks appeared to underestimate competence.

By 4 years of age, children have begun to learn that the number words in the counting sequence represent different specific quantities of increasingly larger size—the ''increasing magnitude'' principle (Sarnecka & Carey, 2008). This principle, however, does not necessarily entail a linear representation of numerical magnitude—an understanding that each successive number increases by a constant amount (e.g., that the difference between 8 and 9 is the same as that between 2 and 3). Instead, some research indicates that young children initially have a logarithmic representation (e.g., their estimated difference between 8 and 9 is smaller than the difference

Correspondence should be sent to Erin E. Reid, Early Math Collaborative, Erikson Institute, 451 N. LaSalle Street, Chicago, IL 60654, USA. E-mail: ereid@erikson.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hjcd.

between 2 and 3; Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010) and shift to a linear representation with age and increasing familiarity with numbers (Laski & Siegler, 2007; Opfer & Siegler, 2007; Praet & Desoete, 2014; Siegler, Thompson, & Opfer, 2009).

Constructing a linear representation of number is critical to developing an understanding of number relations and estimating relative size and provides a solid foundation for school mathematics (Case & Okamoto, 1996; National Council of Teachers of Mathematics, 2006). For example, a more linear representation of number has been associated with higher levels of academic achievement (Booth & Siegler, 2006, 2008; Siegler & Booth, 2004). Conversely, children with a mathematics learning disability or with low mathematics achievement have a representation that is less linear than that of typically achieving students (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Geary, Hoard, Nugent, & Byrd-Craven, 2008).

Given the importance of a linear representation of number to general mathematical development, there has been growing interest in early intervention designed to promote it among young children. However, the conventional task used in prior intervention research (Ramani & Siegler, 2008, 2011; Siegler & Ramani, 2008, 2009; Whyte & Bull, 2008) may have limitations when used with younger children. Given the importance of accurately assessing young children's number magnitude representation, it is important to identify a valid and developmentally appropriate task for educational, as well as research, purposes. The primary purpose of this study was to compare the performance of 4- and 5-year-old children on the conventional linearity measure, a modified measure that may better fit young children's extant informal number knowledge (i.e., knowledge acquired through everyday activities outside of a classroom-type setting; Baroody, Lai, & Mix, 2006), and theoretically related number tasks.

## Conventional Number-Line Tasks

A *number-to-position* (N-P) task commonly is used to assess children's representation of number. The N-P task requires children to mark where a given number falls on a number line (Siegler & Opfer, 2003). The ''conventional'' version of this task for younger children involves presenting a number line with ''0'' as the left anchor and ''10'' as the right anchor and a ''target number'' (a numeral *n* 1–9) above the center of the line (refer to Frame A of Figure 1). After asking a child to identify the digit above the line, a tester inquires, ''If this is where 0 goes (pointing to its position on the number line) and this is where 10 goes (pointing to its position on the number line), where does *n* go?'' Children then mark their estimated location on the line.

## Potential Difficulties with the Conventional N-P Task

The conventional N-P task may not accurately reflect a preschooler's mental representation of number magnitude for the following four reasons:

1. A number line, which is a measurement (continuous-quantity) model of number, may be a foreign representation of numbers to young children. Verbal and written representations of number (number words and numerals, respectively) are typically introduced in the context of subitizing or counting collections, and thus, children's initial representation of number may be tied closely, even entirely, to discrete quantities

(Sarama & Clements, 2009). An initial discrete quantity-based view of the counting numbers would account for a number of common findings. Young children typically do not spontaneously break up continuous quantities into countable units (Fuson, 1988; Huntley-Fenner, 2001) and struggle to understand the process of measuring length well into elementary school (Kamii, 2006; Lehr, Jenkins, & Osana, 1998; Thompson & Preston, 2004). They consistently misconceive number lines as representing discrete quantities as evidenced by counting the numbers or hash marks on the number line, not the linear extents between hash marks. This misconception often leads to systematic errors such as counting 0 or its hash mark as ''1'' (Lehrer, 2003). The evidence of young children's struggle to comprehend number lines has led some educators to recommend postponing their use until after children have mastered the concept of linear units (Ernest, 1985; Fuson, 2009; National Research Council, 2001).

2. Accurately indicating a number's position on a number line requires more advanced knowledge than that needed to represent mentally the relative magnitude of counting numbers (numbers that represent discrete quantities). Specifically, the bounded nature of conventional number-line tasks requires making a judgment based on either proportional reasoning about linear extents (Barth & Paladino, 2011; Slusser, MacDonald, & Barth, 2011; Slusser, Santiago, & Barth, 2013) or equal partitioning of a length (Steffe, von Glasersfeld, Richards, & Cobb, 1983). The accurate application of either strategy would seem to require children to comprehend the measurement concept of a linear unit, a concept that typically develops later than countable (discrete) units (Cohen & Sarnecka, in press; Sarama & Clements, 2009). Indeed, Saxe et al. (2010) found that fifth graders struggled with number line-based problems that required the application of a concept of linear unit. For example, shown a number line with 8 and 10 marked and labeled, students located 11 at the position for 12 (i.e., they incorrectly considered the distance between 8 and 10 as one unit).
3. Zero is not a natural or counting number and may be unfamiliar to many young children, especially those at risk for academic difficulties (Bofferding, 2014; Butterworth, 1999). For example, Wellman and Miller (1986) found that understanding the concept of 0 often lags behind that of other small numbers. For many young children, then, a left-hand anchor of 0 may be meaningless and consequently unhelpful or even misleading.
4. The placement of the target number above the midpoint of the number line may mislead children and cause them to respond by placing their mark directly under the presented number. Children confused by the unfamiliar continuous representation, the symbol for 0, or both may grasp for any clue as to how they should respond. Such children may conclude that the position of the target number above the number line indicates where they should point on the number line.

In brief, if the goal is to evaluate young children's informal representation of the counting numbers (i.e., the relative magnitudes of numbers representing discrete quantities), the traditional N-P task may be a developmentally inappropriate and confusing task that can lead to underestimating their competence.

Consistent with the conjecture that the conventional N-P task may be difficult for young children to understand, Ramani and Siegler's (2008, 2011; Siegler & Ramani, 2008, 2009) pretest results were neither logarithmic nor linear with a slope appreciably larger than 0. Specifically, the plot was a nearly flat line with an intercept of between 4 and 5 (mean intercept across studies = 4.37). Such findings indicate that regardless of the numeral shown (actual magnitude), children consistently indicated the *middle* of the number line as their estimated magnitude (i.e., many may have used a midpoint response bias).

The conjecture is further supported by the results of Berteletti et al.'s (2010) study. These researchers used a nonconventional N-P task that avoided two of the four previously mentioned concerns with the conventional N-P task: a) The number line was marked ''1'' on the left end and ''10'' on the right, and b) the target number was printed in the upper left corner of the paper. Their preschool participants were required to estimate numbers 2 to 9 except number 5. In two experiments with separate samples, children's percent absolute error (PAE) was reported as 28% and 24% for the youngest age group (aged 3;6–4;5), 24% and 22% for the middle age group (aged 4;6–5;4), and 15% and 20% for the oldest age group (5;5–6;4) for Experiments 1 and 2, respectively. These results are somewhat better than those observed by Ramani and Siegler (2008; Siegler & Ramani, 2008, 2009), who reported baseline PAE values ranging from 27% to 29% for children aged roughly 4;0 to 5;5. That is, children of approximately the same age were more accurate on a number-line task when the left anchor was ''1'' instead of ''0.''

In Berteletti et al.'s (2010) report, the shape of number representation also differed from that revealed by studies that used the conventional N-P task. Specifically, the median estimates of the youngest group were almost equally fit by linear and logarithmic models, whereas older children's estimates were best fit by a linear model, with slopes of 0.95 to 1.27 and 0.94 to 1.04 and y-intercepts of −0.52 to −1.81 and −0.89 to −1.01 for the middle and oldest age groups, respectively, across experiments. These fit statistics, unlike those of Ramani and Siegler (2008, 2011; Siegler & Ramani, 2008, 2009), indicate that Berteletti et al.'s participants, who saw a more meaningful 1-to-10 number line and the target number in the upper left corner, did not show evidence of a midpoint response bias. Although the Berteletti et al. results are consistent with the conjecture that the conventional N-P task is confusing to young children, their research did not directly compare young children's responses on this task with those on the conventional task.

## Rationale for the Current Study

A primary purpose of the present research was to determine if children would be more successful on the number-line task if it were better aligned with their informal discrete view of number. To achieve this aim, young children's performance on the conventional N-P task was directly compared to their performance on a modified N-P task (the conventional N-P task with a few relatively simple modifications to make the task more comprehensible for young children). Another primary purpose was to evaluate the validity of the conventional and the modified N-P tasks. To achieve this aim, children's performance on the modified and conventional tasks was compared to performance on four measures of related number skills.

*Comparison with the conventional N-P task.* The modified N-P task differed from the conventional N-P task in three ways that might better accommodate young children's informal discrete view of the counting numbers. First, the task was introduced in a manner that related the number line to this view and familiar everyday experiences (e.g., the *number* of hops a bunny makes to reach a carrot). The aim of the hopping analogy was to help children translate the number line's continuous representation of distance into discrete and countable reoccurring actions. Second, we designed the number line so that the left reference point was the familiar number ''1'' instead of the less familiar number ''0.'' Third, we removed the target number from above the center of the line and read the number to the child.

*Comparison with theoretically related tasks.* Izard, Pica, Spelke, and Dehaene (2008) identified the successor function (principle) as one prerequisite of a concept of exact natural numbers. This principle entails the understanding that each successive number name in the counting sequence refers to a quantity that is exactly one larger than its predecessor in this sequence. Logically, the successor principle is the conceptual basis for a linear representation of discrete numbers with a constant difference of 1. Put differently, such a representation may be the byproduct of the conceptual insight that is the successor principle. Fluent number-after knowledge is a prerequisite skill for fluent performance on a successor task and by implication an N-P task. However, knowing the count sequence is not a sufficient condition for having a linear representation of number. Knowledge of the successor principle and number-after relations underlies an ability to determine the larger of two adjacent counting numbers (close magnitude comparisons) and fluency with mental arithmetic (Frye et al., 2013). No published study has compared young children's performance on an N-P task to their performance on other theoretically related measures of early numeracy for the specific purpose of evaluating the validity of an N-P task.

## Specific Questions Addressed

The present study was designed to address two main questions regarding the comparison of the modified N-P task to the conventional N-P task (Questions 1 to 2) and to address one main question regarding the comparison of N-P tasks to theoretically/logically related tasks (Question 3).

1. Are young children's estimates as accurate on the conventional N-P task as on the modified N-P task? Specifically, a) are there differences in average PAE of children's estimates on the conventional and modified N-P tasks, and b) are children's estimates stable across time and task?
2. Do young children's estimates have similar shapes on the conventional N-P task as on the modified N-P task? a) Are children's estimates on both the conventional and modified N-P tasks best fit by the same model (e.g., logarithmic, linear, or power)? b) Are task differences in the shape of representations stable over time?
3. Is performance on the N-P tasks related to performance on related measures? Specifically, is a linear representation, as gauged by the conventional N-P task or the modified N-P task, associated with knowledge of the successor principle and number-after relations and fluency with close-number comparisons and mental arithmetic?

## METHOD

### Participants

A total of 45 children were recruited from five private preschool classrooms and one private kindergarten classroom in a midsized Midwestern city. Participants ranged in age from 4;0 to 6;0 ($M_{age} = 5;1$, $SD = 0;5$ months) at the first testing session. Boys comprised 47% of the sample. The children's race was reported as 68.9% Caucasian, 24.4% Asian, 4.4% Black, and 2.2% Hispanic.

### Measures

*Conventional N-P task.* The conventional N-P task (e.g., Siegler & Opfer, 2003) involved presenting a participant with a sheet of paper with a 25-cm line with ''0'' printed below the left end of the line and ''10'' printed below the right end of the line (refer to Frame A of Figure 1). Approximately 2 cm above the midpoint of the line was printed a single digit from 1 to 9. Children were asked to estimate numbers 1 to 9 following the procedures described earlier. Internal consistency of the task, based on children's estimates, was poor for the first administration and good for the second administration (Cronbach's $\alpha = .59$ and .81, respectively).

*Modified N-P task.* The format of the number line on the modified N-P task was identical to the conventional task with one exception—the position of 0 was not marked. Instead, a vertical line positioned 2.5 cm from the left of the number line marked the position of ''1.'' In addition, a bunny (or frog) was printed above the position corresponding to ''0'' on the number line and a

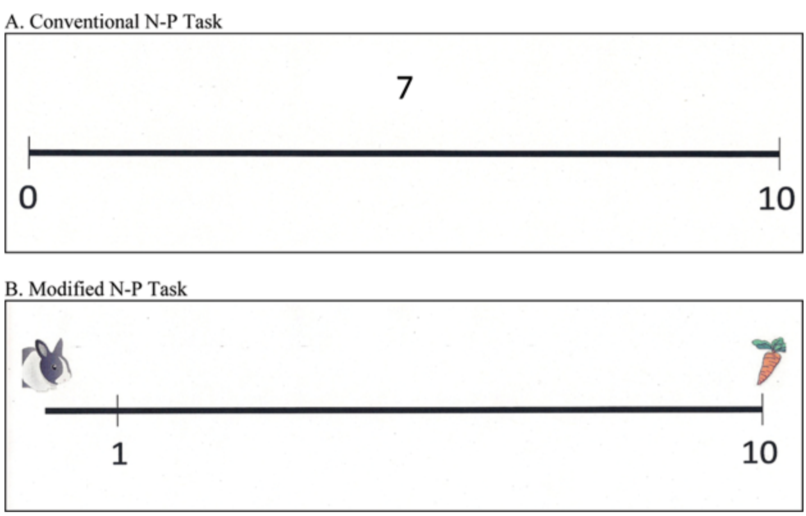


FIGURE 1 A. Representation of the conventional number-to-position (N-P) task (Siegler & Opfer, 2003). B. Representation of the modified N-P task. The modified N-P task differed from the conventional task in three ways: a) a left anchor of ''1,'' not ''0''; b) the target number (e.g., 7) was orally stated, and c) a relatable context was added to the task (the child was asked to estimate where the bunny would land after 7 hops).

carrot (or lily pad) was printed above the vertical mark for ''10'' (refer to Frame B of Figure 1). Half of the children completed the task with the bunny context first and frog context second. The order was reversed for the other half of the sample. For each trial, the examiner said, ''It takes the bunny/frog exactly 10 hops/leaps to get from here (position of bunny/frog) to the carrot/lily pad. If this is where the bunny/frog lands when he hops/leaps one time (point to the vertical line above the ''1'') and this is where the bunny/frog lands when he hops/leaps 10 times (point to the vertical line above 10/below the carrot/lily pad), where will the bunny/frog land when he hops/leaps *n* times? Use this pencil to mark on the line where the bunny/frog would land. Take your best guess.'' Children were asked to estimate numbers 2 to 9. Internal consistency, based on children's estimates, was acceptable for both administrations (Cronbach's $\alpha = .77$ and .78 on the first and second administrations, respectively).

*Number-after knowledge.* Number-after knowledge was assessed via the Lost Dog game. Children were presented with a game mat and a dog figurine. The objective of the game was to help the lost puppy find his way home by answering number-after questions (e.g., ''Tell me what number comes after 3 when we count; 3 and then comes?'') that created a path for the puppy to follow to a dog house. Two trials were presented once (after 3 and 4) and three trials were presented twice (after 5, 6, and 7). One point was awarded for each correct and fast (<3 s) response for a total possible score of 8 points. Internal consistency was acceptable for this task (Cronbach's $\alpha = .82$).

*Close-number comparisons.* Children's skill at determining the larger of two adjacent numbers in the counting sequence was assessed by the Race game. To determine the number of spaces to move around the racetrack, the child was presented a card printed with two numbers and was asked which was greater (e.g., ''Would you rather move 7 spaces or 6 spaces? Which is more, 7 or 6?''). The child then moved that number of spaces. Six number comparisons (2 vs. 3, 5 vs. 4, 6 vs. 7, 4 vs. 3, 5 vs. 6, 8 vs. 7) were presented twice. One point was awarded for correct and fast (<3 s) responses for a possible total of 12 points. Internal consistency was acceptable for this task (Cronbach's $\alpha = .73$).

*Successor principle task.* Understanding of the successor principle was assessed by the Shopkeeper task. Testing materials consisted of a Cookie Monster hand puppet donned by the examiner, 2 shopping lists (each consisted of a list of five grocery items on a 4.25-inch (11 cm) × 5.5-inch (14 cm) sheet of paper), 10 brown paper lunch sacks, and 10 sets of small rubber food items (e.g., bananas, carrots). Cookie Monster was introduced as a shopkeeper ''who sometimes gets confused.'' The child was asked to help the examiner make sure Cookie Monster gave the right number of items on the shopping list. The examiner read the first item on the list—''Cookie Monster, I want *six* apples, please''—and then pulled out a paper sack that contained the set of apples. The examiner identified the number of apples in the paper sack (e.g., ''That's *five* apples, Cookie Monster'') and allowed the child to take a quick peek at the apples in the sack. The brief exposure to the contents in the sack ensured that the child would not be able to use counting to figure out the solution. The examiner then asked, ''Did Cookie Monster give me the right number of apples?'' After verifying with the child that Cookie Monster gave the examiner the wrong number of items, the examiner asked, ''I have five apples here, but I need six apples. What can Cookie Monster do so that I have six apples? Can you help Cookie Monster to make the five apples to six apples please?'' Ten trials were presented: four *successor* trials (3→4, 4→5, 5→6, 6→7), four *reverse* successor trials (2→1, 5→4, 6→5, 7→6), and two $n+2$ trials (4→6, 5→7). For successor trials, 1 point was awarded for each fast (<3 s) and correct response (i.e., indicating ‘‘give one more’’) with no evidence of overapplication of the successor principle. Overapplication of the successor principle was indicated if the child responded ‘‘one more’’ on three out of four of the reverse successor items and one out of two $n+2$ items. For reverse successor trials, which gauged the reversibility of the successor principle (and a relatively deep understanding of the principle), 1 point was awarded for each fast (<3 s) and correct response (i.e., indicating ‘‘take one away’’). For $n+2$ trials, 1 point was awarded for any response that indicated that more than one needed to be added (regardless of response time). The maximum score for this task is 10 points. Internal consistency was acceptable for this task (Cronbach’s $\alpha = .91$).

*Mental arithmetic task.* The purpose of this task was to gauge children’s fluency with mentally adding and subtracting small quantities. The examiner put together an initial collection of plastic dinosaurs (ranging from two to seven) out of the view of the child. The examiner then uncovered the collection for 3 s, announced the number of items, and re-covered the collection. For addition trials, the examiner simultaneously placed no items or one item to the left of the covered collection for 3 s, announced the number of items, and then simultaneously slid the dinosaur under the cover. For subtraction trials, the examiner simultaneously removed one dinosaur from the covered collection, announced the number of items removed and left it in view for 3 s, and then removed it from view. The child then was asked, ‘‘How many dinosaurs are hiding now? How much is $n$ dinosaurs and (take away) $n$ dinosaurs?’’ Sixteen trials were presented: seven $n+1$ ($1+1$, $2+1$, $3+1$, $4+1$, $5+1$, $6+1$, $7+1$), five $n-1$ ($2-1$, $4-1$, $5-1$, $6-1$, $7-1$), and four $n+0$ ($4+0$, $5+0$, $6+0$, $7+0$). For mental arithmetic, children were awarded 1 point for each fast (<3 s) and correct response, for a maximum of 16 points. Internal consistency was acceptable for this task (Cronbach’s $\alpha = .76$).

## Procedure

Testing spanned four sessions, with 2 to 7 days between sessions. The mental arithmetic and N-P tasks were completed in four sessions and the number-after, close magnitude comparisons, and successor principle tasks were completed in two sessions. Within test sessions, the order of tasks was randomized. For the N-P tasks, children completed each task twice. Participants were randomly assigned to one of two conditions. In the first condition, children ($n = 22$) completed the conventional task first followed by the modified task. The children in the second condition ($n = 23$) completed the modified task first followed by the conventional task. This ordering was repeated for the third and fourth test sessions. Trials were presented in random order. Children’s performance on the N-P tasks was evaluated in terms of accuracy and linearity.

*Accuracy.* Children’s accuracy was assessed using two methods. First, accuracy was assessed by calculating PAE between the target number and the child’s estimate. PAE was calculated using the following formula (Siegler & Booth, 2004): ([estimate – estimated quantity]/scale of estimates) × 100. For example, if a child was asked to estimate the position of 6 and made a mark at the position corresponding to 6.7, the PAE would be: $([6.7-6.0]/10) \times 100 = 7\%$. The mean of these values across estimates was then calculated for each individual. Because PAE is a summary measure, some valuable information is lost. Therefore, a second method of assessing accuracy was employed to indicate how many times a child correctly estimated the position of a magnitude on the number line. A child's estimate was calculated as being correct if it fell within a $\pm 0.5$-unit band around the target magnitude. For instance, for the target number 5, estimates falling between 4.5 and 5.5 were scored as correct and scores outside that range were scored as incorrect. These scores were then summed to produce a total correct score.

*Shape of representation.* Although it has been well documented that children's numerical estimations are well represented by a linear or logarithmic function (depending on age of the child and range of numbers), it is argued in more recent literature that the proportional judgment account might better account for the shape of children's estimates given the bounded nature of the number-line tasks. Successful completion of a bounded number-line task requires a child to estimate the magnitude of a number relative to one or more reference points (e.g., endpoint). According to the proportion judgment account (see Barth & Paladino, 2011, and Slusser et al., 2013, for an overview), the accuracy of children's magnitude estimates, in part, is dependent upon which, if any, reference points children use to estimate magnitude. Young children, who are unfamiliar with the task demands or the number range, might mark their estimates without regard to a reference point, thereby producing inaccurate estimates that are best fit by a power function (see Frame A of Figure 2). On the other hand, when children mark their estimates relative to an endpoint (i.e., make a judgment about proportion), their estimates might be better fit by a one-cycle power function, wherein estimates to the left of the midpoint are overestimated and estimates to the right are underestimated (Spence, 1990; see Frame B of Figure 2). Finally, children who use an inferred midpoint as well as an endpoint for reference might produce estimates best fit by a two-cycle power function, wherein the pattern of overestimation and underestimation observed in the one-cycle model is repeated to the left and right of the midpoint (see Frame C of Figure 2). As such, in the current study, children's estimates were fit to five models: linear, logarithmic, unbounded power, one-cycle power, and two-cycle power.

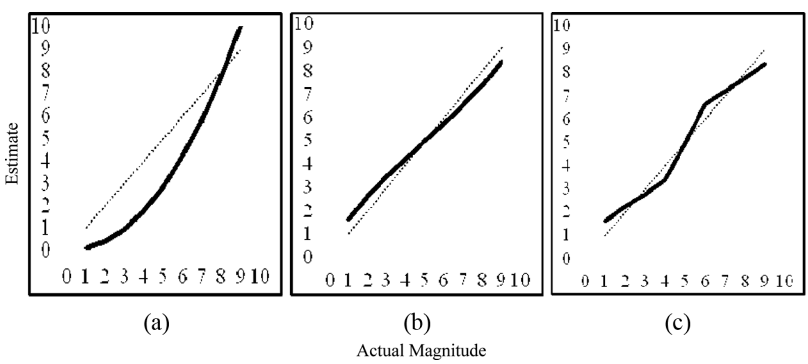


FIGURE 2 A. Children who are unfamiliar with the target number range or who mark estimates without regard to the endpoint might produce estimates best fit by an unbounded power model. B. Children who mark their estimates relative to the endpoint might overestimate values to the left of the midpoint and underestimate values to the right of the midpoint. In these cases, estimates might be best fit by a one-cycle power model (Spence, 1990). C. Children who mark their estimates in relation to two reference points (e.g., endpoint and inferred midpoint) might produce estimates best fit by a two-cycle power model (Hollands, Tanaka, & Dyre, 2002). The dashed line represents $y = x$.

Model fit was compared using Akaike information criterion, corrected for small sample sizes (AICc; Slusser & Barth, n.d.), and the model with the lowest AICc score was selected as the best-fitting model.

## RESULTS

To facilitate interpretation of the data, we adopted exclusion criteria reported in Slusser et al. (2013). Estimates on a given task were excluded if they were either nonsignificantly or negatively correlated with the target numbers or if 85% or more of the estimates fell within a span of 2.5 cm. Five children met the exclusion criteria on all four N-P tasks and were excluded from the study entirely. Four children met the criteria on three tasks, four met the criteria on two tasks, and five met the criteria on one task. Table 1 provides a breakdown of the number of children meeting exclusion criteria by task and criteria.

### Question 1: Relative Accuracy of N-P Tasks

N-P task differences regarding accuracy and differences in stability of estimates over time are addressed in turn.

*Task differences in accuracy.* Mean PAE values for each task are reported in Table 2. The results of a mixed analysis of variance (ANOVA) with task (conventional N-P task PAE or modified N-P task PAE) as the within-subjects variable and task order (conventional N-P task first or modified N-P task first) as the between-subjects variable indicated a significant main effect of task, $F(1, 29) = 16.80$, $p < .001$, but not task order, $F(1, 29) = 0.07$, $p = .792$. Differences in PAE between the first administrations of the conventional and modified N-P tasks, $t(30) = 4.16$, $p < .001$, and between the second administration of each task, $t(27) = 4.69$, $p < .001$, favored the modified task. Moreover, as illustrated in Figure 3, participants who received the conventional task at the first time point had appreciably higher PAE values (i.e., were less accurate) than were those who tested on the modified task. At the second time point, the mean PAE of those who received the modified task second dropped to nearly the same level as that of their peers at the first time point, and the PAE of those who received the conventional task second soared to nearly the same level as that of their peers at first time point. This same cycle repeated itself for the third and fourth time points.

TABLE 1
Number of Cases Excluded From Analyses by Task and Exclusion Criteria

| | *Conventional N-P Task* | | *Modified N-P Task* | |
|---|---|---|---|---|
| *Exclusion Criteria* | *Time 1* | *Time 2* | *Time 1* | *Time 2* |
| Not Correlated | 6 | 6 | 7 | 12 |
| Negatively Correlated | 5 | 3 | 4 | 0 |
| Within 2.5 cm | 0 | 1 | 0 | 1 |

TABLE 2
Mean Percent Absolute Error of Estimates by Task Across Conditions

| *N-P Task* | *n* | *PAE* | *SD* | $R^2_{lin}$ | $Slope_{lin}$ |
|---|---|---|---|---|---|
| Conventional | | | | | |
| First Administration | 34 | 17.49 | 6.87 | .78 | .77 |
| Second Administration | 35 | 17.75 | 8.75 | .77 | .77 |
| Modified | | | | | |
| First Administration | 34 | 11.89 | 5.16 | .79 | .85 |
| Second Administration | 32 | 12.12 | 5.74 | .80 | .84 |

*Note*. PAE = percent absolute error. $R^2$ and slope values reported for the linear model. First administration refers to the first time a participant received a task (whether at Time Point 1 or Time Point 2); second administration refers to the second time a participant received a task (whether at Time Point 3 or Time Point 4).

A breakdown of percentage of children producing correct estimates (within $\pm 0.5$ units of the target number) on both administrations of each task by condition is provided in Table 3. Note that participants in both conditions did appreciably better placing numbers near the endpoints of the number line, the numbers 2 and 9 particularly, on the modified N-P task. Overall, children correctly estimated more points on the number line when completing the modified task than when completing the conventional task, $t(30) = 3.19$, $p < .01$.

Examining the strategies used by children to make their estimations may provide additional insight into the patterns of the results described in the previous paragraphs. Evidence of response biases was observed in 18% of estimates across tasks and time points (including estimates of children excluded from the analyses). Response biases were defined as patterns of responding that were observed for at least 75% of responses per task (85% in the case of marking responses at both endpoints of the number line). The most frequently observed response bias (61% of cases) consisted of children marking the majority of their estimates at one or both

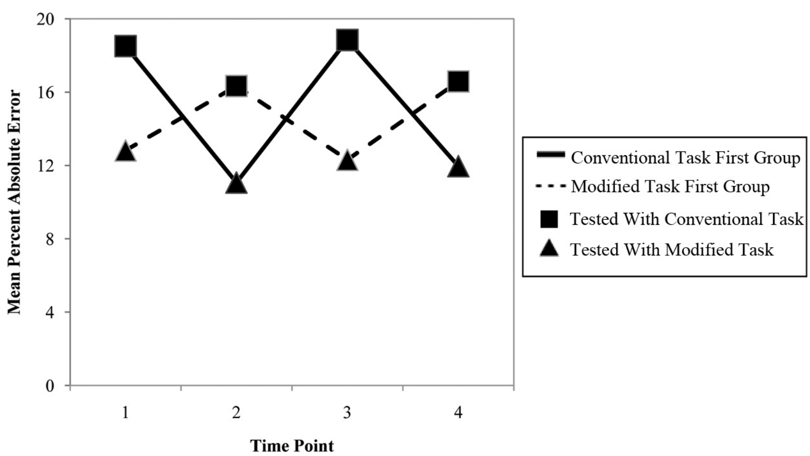


FIGURE 3 Mean percent absolute error values of children's estimates on the conventional and modified N-P task by condition.

TABLE 3
Mean Total Scores for N-P Tasks and Percent Correct for Numerals by Task by Condition

| Condition | | *First time point* | *Second time point* | *Third time point* | *Fourth time point* |
|---|---|---|---|---|---|
| | | *Conventional Task (n = 18)* | *Modified Task (n = 18)* | *Conventional Task (n = 18)* | *Modified Task (n = 18)* |
| Conventional N-P task first | Mean total score (SD) | 1.61 (1.42) | 2.39 (1.24) | 1.83 (1.72) | 2.61 (1.72) |
| | Number Estimated | % Correct | % Correct | % Correct | % Correct |
| | 1 | 17 | — | 22 | — |
| | 2 | 11 | 67 | 6 | 56 |
| | 3 | 6 | 22 | 17 | 33 |
| | 4 | 11 | 17 | 6 | 28 |
| | 5 | 0 | 28 | 17 | 11 |
| | 6 | 17 | 17 | 17 | 22 |
| | 7 | 11 | 22 | 28 | 11 |
| | 8 | 33 | 17 | 22 | 39 |
| | 9 | 56 | 50 | 50 | 61 |
| | | *Modified task (n = 16)* | *Conventional task (n = 16)* | *Modified task (n = 14)* | *Conventional task (n = 17)* |
| Modified N-P Task First | Mean Total Score (SD) | 2.63 (1.89) | 1.62 (1.20) | 2.71 (1.44) | 1.65 (1.50) |
| | Number Estimated | % Correct | % Correct | % Correct | % Correct |
| | 1 | — | 25 | — | 24 |
| | 2 | 63 | 6 | 79 | 18 |
| | 3 | 31 | 6 | 21 | 12 |
| | 4 | 19 | 6 | 36 | 6 |
| | 5 | 13 | 25 | 7 | 18 |
| | 6 | 19 | 13 | 14 | 6 |
| | 7 | 19 | 6 | 7 | 12 |
| | 8 | 31 | 19 | 43 | 24 |
| | 9 | 69 | 56 | 64 | 47 |

extreme ends of the number line. This occurred equally on both task types. More than half of those cases involved marking estimates exclusively on the left end of the number line. In 20% of cases with evidence of a response bias, estimates were marked within a narrow band at the midpoint of the number line. All except one instance of this type of response bias was observed on the conventional task. Twenty percent of response-bias cases consisted of estimates marked within narrow bands at locations other than midpoints or endpoints. Although not systematically recorded and analyzed, examiners reported that children frequently used pulsing motions to count out the placement of their estimates. This strategy did not necessarily lead to accurate estimates as most of the children focused on number sequence, not unit size. This strategy also produced a truncated range of estimates. Some children used this strategy for numbers up to 6 and then worked backward from 10 for numbers 7 to 9.

*Task differences in stability over time.* Results of mixed repeated-measures ANOVAs with task order (conventional N-P task first or modified N-P task first) as a between-subjects variable indicated no significant main effects of time on PAE for the conventional,

TABLE 4
Estimate Correlations Across Administrations for Each Task Controlling for Age in Months

| | *Task* | |
|---|---|---|
| *Estimate* | *Conventional* ($n = 28$) | *Modified* ($n = 17$) |
| 1 | .40* | — |
| 2 | .08 | .59* |
| 3 | .31† | .28 |
| 4 | .37* | .42* |
| 5 | .31† | .50* |
| 6 | .25 | .36† |
| 7 | .67* | .28 |
| 8 | .51* | .45* |
| 9 | .53* | .55* |

*Note.* † $p < .10$. * $p < .05$.

$F(1, 29) = 0.76$, $p = .392$, and modified, $F(1, 28) = 0.14$, $p = .708$, N-P tasks or of task order. Correlations between children's individual estimates across administrations controlling for age in months are presented in Table 4. Estimates for the same numbers on the same tasks exhibited moderate stability over time with median correlations of .37 and .44 on the conventional and modified tasks, respectively. Across tasks, estimates for the first administrations of conventional and modified tasks were relatively stable ($r$s = .38–.65, $p$s < .05) with exceptions for numbers 3 ($r = .15$, *ns*) and 6 ($r = -.06$, *ns*).

## Question 2: Shape of Representation Prompted by N-P Tasks

N-P task differences regarding shape and differences in stability over time are addressed in turn.

*Task differences in shape.* Analysis of median estimates indicated that the unbounded power model best fit the data for each task at each time point (see Figure 4). Further analysis of individual estimates on the conventional N-P task revealed that 71% of the 34 valid sets of estimates were best fit by the unbounded power model, 17% by the one-cycle model, 6% by the two-cycle model, and 3% each for the linear and logarithmic models. The unbounded power model was the best-fitting model for 55% of the valid sets of estimates on the modified N-P task, whereas the one-cycle and two-cycle models best explained 21% of the sets of estimates each and the linear model best fit 3% of estimates.

*Task differences in stability over time.* Consistency in model fit was examined for changes over time and task. Overall, 58% of children's sets of estimates were best fit by the same model on the first and second administrations of the conventional N-P task. The same was true for 60% of cases on the modified N-P task. Comparing the first administration of each task type, 61% of cases were fit by the same model. Of the 27 cases with valid estimates at all four time points, 33% were best fit by the same model across all time points.

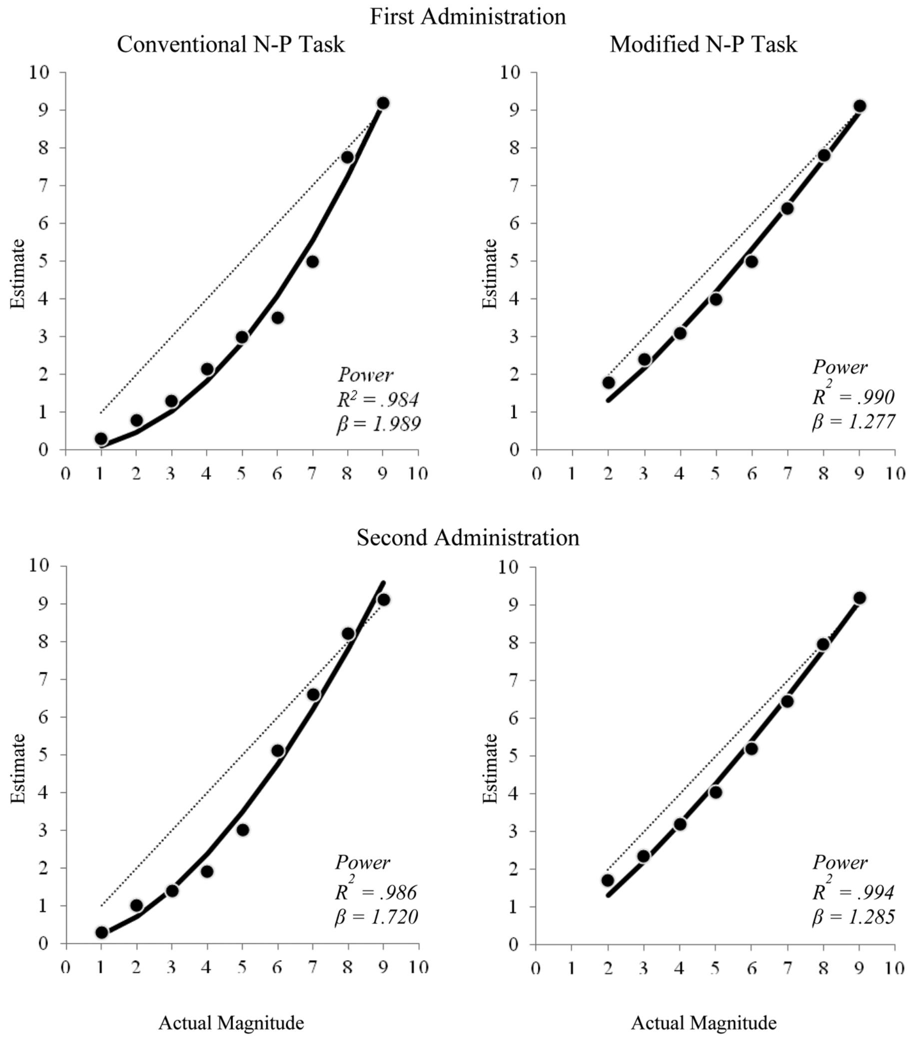


FIGURE 4 Children's median estimates on the first and second administrations of each N-P task. The solid line represents the best-fitting model. The dashed line represents y = x.

## Question 3: Relationship to Related Number Skills

Although the N-P tasks were generally challenging, the other related math tasks proved to be relatively easy for the majority of children (number-after knowledge, $M = 6.70$, $SD = 1.64$, max = 8; close-number comparison, $M = 11.53$, $SD = 0.75$, max = 12; successor principle task, $M = 8.10$, $SD = 2.68$, max = 10; mental arithmetic task, $M = 13.78$, $SD = 1.87$, max = 16). Relationships among performance on the related tasks and overall accuracy (PAE) and linearity ($R^2$s and slopes from the linear model) on the modified and conventional N-P tasks were examined by calculating partial correlation coefficients controlling for age in months

TABLE 5
Correlations Among N-P Tasks and Related Number Skills

| | *Conventional N-P Task* | | *Modified N-P Task* | |
|---|---|---|---|---|
| | *PAE* | $R^2_{lin}$ | *PAE* | $R^2_{lin}$ |
| Number-after knowledge | .01 | .01 | −.04 | .08 |
| Close-number comparison | −.14 | −.19 | .03 | .00 |
| Successor understanding | .02 | −.02 | −.17 | .15 |
| Mental arithmetic | −.23 | −.06 | −.07 | .01 |

*Note*. PAE = percent absolute error. $R^2$ s reported for the linear model.

(see Table 5). Significant relationships were observed between the children's fluency with mental arithmetic and linear slope on the conventional ($r = .42$, $p = .02$) and modified ($r = .38$, $p = .03$) tasks only. Using Steiger's (1980) $z$ test, the difference between correlations was not statistically significant, $z = .39$, $t(28) = 0.39$, *ns*.

## DISCUSSION

Performance on the modified N-P task differed substantially from that on a conventional version in that children produced more accurate estimates on the modified task. Despite differences in estimation accuracy, both tasks demonstrated a significant relationship between linear slope values and performance on a mental arithmetic task. No relationships were observed among the N-P tasks and measures of other related number skills (i.e., number-after knowledge, close-number magnitude comparison, or successor principle). The results suggest that the modified N-P task better fits children's informal discrete view of number. However, the results are mixed regarding the validity of using either task for gauging magnitude judgments about numbers with a cardinal meaning.

### Relative Accuracy of N-P Tasks

The participants were significantly more accurate, as measured by PAE and number of correct estimates, on the modified N-P task than on the conventional N-P task between each time point and within conditions across time points. This is particularly striking as individual children served as their own control. Although the accuracy of children's estimates varied as a function of task type, children's estimates on each task were relatively stable during a period of 1 to 2 weeks. Overall, whether participants received the conventional task in the first and third sessions or the second and fourth sessions, they were more accurate on each of two administrations of the modified version than on each of the two administrations of the conventional task. This indicates that the conventional N-P task remained relatively unclear to children even after experiencing greater success on the modified N-P task.

### Shape of Representation Prompted by N-P Tasks

Previous research with the conventional task (Geary et al. 2007, 2008; Opfer & Siegler, 2007; Praet & Desoete, 2014; Siegler & Opfer, 2003) or a nonconventional task (Berteletti et al., 2010) has suggested that children's representation of number starts out logarithmically (i.e., the estimated difference between 7 and 8 is *larger* than the difference between 2 and 3) and then shifts to a linear representation (i.e., the estimated difference between 7 and 8 is the *same* as the difference between 2 and 3). Indeed, Berteletti et al. (2010) found that 5- and 6-year-olds had generally shifted from a logarithmic to a linear representation of the first 10 counting numbers. However, Barth and Paladino (2011) and Slusser et al. (2013) proposed a proportion judgment account of children's representations of numerical magnitude estimations on number-line tasks that takes into account the bounded nature of the tasks. To be successful, a child must make a judgment about proportion using available, and possibly inferred, reference points. A power function, therefore, might be better suited to capture nuances in estimates than a linear or logarithmic model (Barth, Slusser, Cohen, & Paladino, 2011).

In the present study, median and individual number-line estimates of the numbers 1 to 9 on the conventional- and 2 to 9 on the modified- N-P task were better fit by an unbounded power model than by any other model tested. This was particularly true for the conventional task, where the unbounded power model was preferred for more than two thirds of children's sets of estimates (compared with a little more than half of estimate sets on the modified task). Although no other published study has examined the fit of power models to estimates of numbers 1 to 10, the results of this study appear to be inconsistent with work conducted with a larger range of numbers and slightly older children. Slusser et al. (2013) found that for 5- and 6-year-olds, the unbounded power model was preferred for an unfamiliar range of numbers (0–100) but that the linear model was preferred for a familiar number range (0–20). If the proportion judgment account is true, the results of this study might indicate that fewer children used reference points to mark their estimates on the conventional task than on the modified task. In other words, the modified N-P task may have done a better job at eliciting proportion judgments (42% of cases were fit by a one-cycle or two-cycle power model) due to its use of a context familiar to children—a hopping bunny (or frog) to aid children in breaking down the number line's continuous representation of distance into discrete and countable reoccurring actions. On the other hand, the conventional task was approached more often as an open-ended estimation task.

Behavioral observations noted during testing are consistent with the notion that many children treated the task as open-ended. First, children were often observed using pulsing motions to count out the placement of their estimates. This strategy did not necessarily lead to accurate estimates as most of the children focused on number sequence, not unit size; that is, the pulses were not consistent in length or were small in comparison to the length of the number line. Children tended to use this strategy for the first six numbers in the counting sequence, but not the remaining numbers. These same children often moved their pencil to the right anchor point and worked backward from 10 to place their estimates of numbers 7, 8, and 9. This explains why estimates of 9 were among the most accurate. Second, 11% of all sets of estimates indicated evidence of a response bias wherein children marked their estimates at the extreme ends of the number line, particularly the left end.

## Relationship to Related Measures of Linear Representation

In this study, we examined the relationship between young children's performance on an N-P task and their performance on measures of related number skills. Theoretically, children who perform well on tasks assessing number-after knowledge, close magnitude comparisons, an understanding of the successor principle, and mental arithmetic should also exhibit a more linear representation of number on a number-line task. These tasks assess a discrete quantity (collection-based) view of number—knowledge that either is logically the basis of a linear representation of the counting sequence or is related to it. No association between accuracy (PAE or number correct) on the N-P tasks and performance on related math skills was observed. To place an estimate accurately on a bounded number line, such as the ones used in this study, one must, at the very least, mark an estimate relative to available reference points. The results of the statistical analyses and behavioral observations indicated that the majority of children did not take reference points into consideration when marking their estimates. In other words, children could have produced linear estimates with a slope approaching 1 and still be inaccurate. This suggests two things. First, number-line tasks gauge a more advanced measurement (continuous-quantity) view of number. Our attempt to make the N-P task counting-based was only partially successful. Children may have been more successful on the modified N-P task because the introduction of a hopping bunny (or leaping frog) better enabled some of them to connect the task demands to a discrete-quantity view of number; however, it did not help them to partition the number line into equal units. Second, accuracy may not be a very useful metric when gauging young children's number-line magnitude representations.

Linearity ($R^2_{lin}$ and slope) of number-line estimates on either task was not associated with performance on related number skill tasks with one exception: Linear slope values on both tasks were associated with performance on the mental arithmetic task. In contrast to our results, Ramani and Siegler (2008) found that linearity (i.e., $R^2_{lin}$) of number-line estimates was moderately correlated with performance on a magnitude comparison task for children enrolled in Head Start. In a similar study with a sample of middle-class preschoolers, Ramani and Siegler (2011) observed small, yet significant correlations between linearity ($R^2_{lin}$ and slope) and a magnitude comparison task. The task used in the current study differed from the one used in Ramani and Siegler's (2011) study in that it required children to make judgments about adjacent numbers. However, even though children were successful on the task overall, their ability to make fine distinctions between quantities did not aid their performance on the number-line task.

## Limitations and Future Directions

A few limitations of the present study should be considered when interpreting the results of this study. Due to sample characteristics, the results may be limited in their generalizability. The children participating in this study were sampled from private, university-affiliated preschool and kindergarten classrooms, suggesting our sample may come from more educated or advantaged households. However, if the N-P tasks were confusing to our sample, they are not likely to be less confusing to samples of less-advantaged children. Additionally, the majority of children performed very well on the four related math skills measures. Little variability in scores, however, may have impeded our ability to better gauge the relationship among these tasks and accuracy or linearity of number-line estimates. Finally, our sample size was small and prohibited the examination of age-related differences in accuracy and shape of representation.

Three modifications were made to the conventional N-P task: The left anchor on the number line was ''1'' rather than ''0,'' the target number was read to the child instead of printed over the center of the number line, and a hopping analogy was introduced in the task's directions. These relatively simple changes were associated with children being able to connect the task demands to their informal discrete-quantity view of the counting numbers and thus enabled them to perform better on this task than on the conventional N-P task. Unclear, though, is whether an anchor of 1, a noncentered target, a hopping analogy, or all three account for task differences in accuracy and responses biases. The role of each of these factors should be examined systematically in future studies to determine which factor or combination of factors is critical to understanding the N-P task.

Given that 40% of the children in the original sample were unable to provide sets of estimates that could be analyzed during one or more time points, that model fit was not consistent within task over time, and that performance on the modified and conventional N-P tasks was not related to three out of four measures of theoretically related number skills, it appears that number-line tasks have limited validity. Future research on young children's magnitude estimations of cardinal numbers might better use a different task to assess number representation. The present results also suggest using longitudinal analyses of individual performance to document changes over time and explore the ability of N-P tasks to predict future mathematics achievement.

## Conclusions

Given the importance of numerical magnitude estimation (National Council of Teachers of Mathematics, 2006; Siegler & Booth, 2004), it is essential to identify reliable and valid measures of this key aspect of number sense with young children to accurately discern the course of early development or evaluate the effectiveness of early intervention. The essential first step in assessing number knowledge is explicitly identifying what meaning of number is the target of interest. A cardinal number meaning refers to the total of distinct (countable) items in a collection (discrete quantities), whereas a measurement meaning refers to the amount of a nondiscrete quantity that cannot be counted or quantified without applying a measurement process (continuous quantity; Fuson, 1988). The ordered numbers of the counting sequence can represent the relative magnitudes of either number meaning—discrete quantities or continuous quantities. For example, the statement ''eight is more than seven'' can represent the fact that a collection of eight blocks has (one) more block than a collection of seven blocks *or* that 8 inches is (1 inch) longer than 7 inches. Previous conclusions about the results of the conventional N-P task did not take into consideration the distinction between cardinal and measurement meanings of number. Specifically, the conventional N-P task involves a bounded number line, which is a model of the measurement meaning of number. Thus, children's magnitude estimates with this task may reflect their measurement, not cardinal, representation of the counting sequence.

The results of the present study indicate that the conventional, or even the modified, N-P task underestimates the linearity of young children's representation of cardinal numbers. Performance on the conventional N-P task was significantly inferior to that on an N-P task modified to make it more consistent with a cardinal view of number and to eliminate potentially confusing features such as using 0 to mark an endpoint and placing the number above the middle of the number line. The significant gap in performance was evident even within subjects and after a child was exposed to the more comprehensible modified N-P task once or twice. Moreover, performance on both N-P tasks was not associated with three out of four theoretically related tasks. The task demands of conventional N-P tasks go beyond those for comparing the magnitude of numbers that represent discrete quantities (e.g., recalling the magnitudes of the target number and the endpoint[s] and judging the former in relation to the latter) and entail relatively advanced competencies (e.g., judging the proportion of the target magnitude in relation to the magnitude of the endpoint[s] and then mapping the estimated proportion onto a spatial representation of a measurement meaning of number; Slusser et al., 2013). In short, although bounded N-P tasks may be useful for assessing more developmentally advanced children's representation of ordinal numbers, such tasks are not valid and suitable for assessing young children's linearity of the cardinal numbers 1 to 10.

The results of the present study are consistent with other research indicating that young children initially associate the counting sequence with discrete quantities and a cardinal meaning of number and then with continuous quantities and an ordinal meaning of number also (see e.g., Cohen & Sarnecka, in press; Fuson, 1988; Lehrer, 2003; Thompson & Preston, 2004). Accurate magnitude judgments of cardinal numbers as operationally defined by a linear representation with a slope of 1 may be a byproduct of constructing the successor principle or at least depend on its construction, which most participants in the present had already done. Similar accuracy with magnitude judgments of numbers as measures may have to await the construction of a concept of linear unit. These conjectures need to be tested empirically.

## NOTE

Portions of the data were presented at the Fall 2011 Conference of the Society for Research on Educational Effectiveness.

## FUNDING

This research was supported, in part, by grants from the Institute of Education Sciences and the U.S. Department of Education (R305A080479 and R305B100017). Any opinions, findings, conclusions, or recommendations expressed are solely those of the authors and do not necessarily reflect the position, policy, or endorsement of the Institute of Education Sciences.

## REFERENCES

Baroody, A. J., Lai, M.-L., & Mix, K. S. (2006). The development of young children's number and operation sense and its implications for early childhood education. In B. Spodek & O. Saracho (Eds.), *Handbook of research on the education of young children* (pp. 187–221). Mahwah, NJ: Lawrence Erlbaum.

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, *14*, 125–135. doi: 10.1111/j.467-7687.2010.00962.x

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

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

Bofferding, L. (2014). Negative integer understanding: Characterizing first graders' mental models. *Journal of Research in Mathematics Education*, *45*, 194–245. doi: 10.5951/jresematheduc.45.2.0194

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *41*, 189–201. doi: 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.

Butterworth, B. (1999). *The mathematical brain*. London, England: Macmillan.

Case, R., & Okamoto, Y. (1996). The role of conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development*, *61*(1/2, Serial No. 246).

Cohen, D. J., & Sarnecka, B. W. (in press). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology*.

Ernest, P. (1985). The number line as a teaching aid. *Educational Studies in Mathematics*, *16*, 411–424.

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Fuson, K. C. (1988). *Children's counting and concepts of number*. New York, NY: Springer-Verlag.

Fuson, K. C. (2009). Avoiding misinterpretations of Piaget and Vygotsky: Mathematical teaching without learning, learning without teaching, or helpful learning-path teaching. *Cognitive Development*, *24*, 343–361. doi: 10.1016/j.cogdev.2009.09.009

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, *78*, 1343–1359. doi: 0009-3920/2007/7804-0019

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. doi: 10.1080/87565640801982361

Hollands, J. G., Tanaka, T., & Dyre, B. (2002). Understanding bias in proportion production. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 563–574. doi: 10.1037//0096-1523.28.3.563

Huntley-Fenner, G. (2001). Why count stuff? Young preschoolers do not use number for measurement in continuous dimensions. *Developmental Science*, *4*, 456–462. doi: 10.1111/1467-7687.00187

Izard, V., Pica, P., Spelke, E., & Dehaene, S. (2008). Exact equality and successor function: Two key concepts on the path towards understanding exact numbers. *Philosophical Psychology*, *21*, 491–505. doi: 10.1080/09515080802285354

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

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.

Lehr, R., Jenkins, M., & Osana, H. (1998). Longitudinal study on children's reasoning about space and geometry. In R. Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 137–167). Mahwah, NJ: Erlbaum.

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.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics: A quest for coherence*. Reston, VA: Author.

National Research Council. (2001). *Adding it up: Helping children learn mathematics* (J. Kilpatrick, J. Swafford, & B. Findell, Eds.). Washington, DC: National Academy Press.

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

Praet, M., & Desoete, A. (2014). Number line estimation from kindergarten to Grade 2: A longitudinal study. *Learning and Instruction*, *33*, 19–28. doi: 10.1016/j.learninstruc.2014.02.003

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*, 375–394.

Ramani, G. B., & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *Journal of Applied Developmental Psychology*, *32*, 146–159. doi: 10.1016/j.appdev.2011.02.005

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge.

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*, 662–674. doi: 10.1016/j.cognition.2008.05.007

Saxe, G. B., Earnest, D., Sitabkhan, Y., Haldar, L. C., Lewis, K. E., & Zheng, Y. (2010). Supporting generative thinking about integers on number lines in elementary mathematics. *Cognition & Instruction*, *28*, 433–474.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*, 428–444.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237–243.

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, *11*, 655–661. doi: 10.1111/j.1467-7687.2008.00714.x

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*, 545–560. doi: 10.1037/a0014239

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.

Slusser, E. B., & Barth, H. (n.d.). *An Excel worksheet for proportion-judgment analyses of number-line data*. Unpublished manuscript, Wesleyan University, Middletown, CT. Retrieved from http://works.bepress.com/emily_slusser/4/

Slusser, E. B., MacDonald, K., & Barth, H. C. (2011, March). *Converging evidence for a proportion-judgment account of numerical estimation: Results from two types of number-line tasks*. Poster session presented at the meeting of the Society for Research in Child Development, Montreal, QC, Canada.

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

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 683–692.

Steffe, L. P., von Glasersfeld, E., Richards, J., & Cobb, P. (1983). *Children's counting types: Philosophy, theory, and application*. New York, NY: Praeger Scientific.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251.

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

Wellman, H. M., & Miller, K. F. (1986). Thinking about nothing: Development of concepts of zero. *British Journal of Developmental Psychology*, *4*, 31–42.

Whyte, J. C., & Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Developmental Psychology*, *44*, 588–596. doi: 10.1037/0012-1649.44.2.588