Child Development, November/December 2007, Volume 78, Number 6, Pages 1723 – 1743

Is 27 a Big Number? Correlational and Causal Connections Among Numerical Categorization, Number Line Estimation, and Numerical Magnitude Comparison

Elida V. Laski and Robert S. Siegler
Carnegie Mellon University

This study examined the generality of the logarithmic to linear transition in children's representations of numerical magnitudes and the role of subjective categorization of numbers in the acquisition of more advanced understanding. Experiment 1 (49 girls and 41 boys, ages 5 – 8 years) suggested parallel transitions from kindergarten to second grade in the representations used to perform number line estimation, numerical categorization, and numerical magnitude comparison tasks. Individual differences within each grade in proficiency for the three tasks were strongly related. Experiment 2 (27 girls and 13 boys, ages 5 – 6 years) replicated results from Experiment 1 and demonstrated a causal role of changes in categorization in eliciting changes in number line estimation. Reasons were proposed for the parallel developmental changes across tasks, the consistent individual differences, and the relation between improved categorization of numbers and increasingly linear representations.

Children's understanding of numerical magnitudes is closely related to their general math achievement (Gersten, Jordan, & Flojo, 2005; Siegler & Booth, 2004), estimation skills (Booth & Siegler, 2006), and arithmetic proficiency (Griffin, Case, & Siegler, 1994; Jordan, Hanich, & Kaplan, 2003). Understanding of numerical magnitudes also is a core component of number sense, an ill-defined construct that nonetheless is widely viewed as crucial to success in mathematics and that is a key goal of mathematics instruction (Dowker, 1992, 2003; Dowker, Flood, Griffiths, Harriss, Hook, 1996; Jordan, in press; Kilpatrick, Swafford, & Findell, 2001; National Council of Teachers of Mathematics (NCTM), 1989, 2000, 2006; Sowder, 1992). Although existing data on the relation between mathematical proficiency and understanding of numerical magnitudes are correlational, they are consistent with the view that helping young children develop a better understanding of numerical magnitudes may lead to improved performance on a variety of numerical tasks.

The present study had two main goals. One was to demonstrate the generality of a hypothesized developmental transition from logarithmic to linear representations of numerical magnitudes. This transition had previously been documented on several estimation tasks; the present study was designed to demonstrate its contribution to both developmental changes and individual differences in other aspects of numerical knowledge. The second main goal was to provide causal evidence that changes in young elementary school children's subjective categorization of numbers can lead to improvements in their numerical magnitude representations.

This introductory section includes three parts. First, we describe recent findings regarding age-related changes and individual differences in representations of numerical magnitude. Second, we examine current knowledge regarding subjective categorization of numbers. Finally, we present hypotheses regarding the generality of the transition in numerical representations, the role of categorization of numbers in that transition, and the experiences that might improve the representations, as well as describing how these hypotheses were tested in the present experiments.

Development of Representations of Numerical Magnitude

With age and experience, children progress from logarithmic to linear representations of numerical magnitudes on several types of estimation tasks.

Thanks to the children, teachers, and principals of the Fox Chapel Area School District; to Richard Scheines for help with the path analysis, to Nicole R. Hallinen for assistance coding the data; and to the Department of Education (Grant R305H020060 and Graduate Training Grant R305B040063) and National Institutes of Health (Grant HD 19011) who supported this research.

Correspondence concerning this article should be addressed to Elida V. Laski or Robert S. Siegler, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213. Electronic mail may be sent to evl@andrew.cmu.edu or rs7k@andrew.cmu.edu.

 0009-3920/2007/7806-0007

Number line estimation can be used to illustrate the change. On a typical number line estimation task, children are asked to indicate the position of Arabic numerals on number lines that are empty except for the number 0 at the left end and a larger number (usually 100 or 1,000) at the right end. On each trial, children are asked to estimate the position on the line of a number, with estimates of numbers from previous trials not remaining visible on later ones.

On 0–100 number lines, kindergartners generate estimates that closely fit a logarithmic function, such that the estimates exaggerate differences in the magnitudes of smaller numbers and compress differences in the magnitudes of larger ones. In contrast, second graders produce estimates that fit a linear function, such that their estimates neither exaggerate nor compress differences among numbers throughout the scale (Siegler & Booth, 2004). The logarithmic function appears to reflect an underlying logarithmic representation of numerical magnitudes; the linear function appears to reflect an underlying linear representation of numerical magnitudes (Opfer & Siegler, in press; Siegler & Opfer, 2003). The logarithmic to linear transition pattern holds true for the estimation patterns of individual children of different ages, as well as for group averages. In addition, a parallel progression occurs on 0–1,000 number lines between second and fourth grades, with second graders' estimates fitting a logarithmic function and fourth graders' and older children's fitting a linear function (Booth & Siegler, 2006; Opfer & Siegler, in press).

The logarithmic to linear transition is not unique to number line estimation. Booth and Siegler (2006) obtained similar findings on numerosity estimation and measurement estimation tasks to the ones they obtained on number line estimation. On the numerosity estimation task, children were shown a computer screen displaying a jar with 1 dot and another jar with 1,000 dots, told how many dots were in each jar, and then were asked to hold down a computer mouse until about *N* dots were in a third, initially empty, jar on the screen. The measurement estimation task involved children being shown a very short line said to be one "zip" long and a much longer line said to be 1,000 zips long and then being asked to draw lines about *N* zips long.

On all three estimation tasks, the predominant function of children's estimates for the 0–1,000 range progressed from logarithmic to linear between second and fourth grade. In addition, individual children's degree of linearity on the three estimation tasks was highly correlated, within grade as well as between grades. The linearity of children's estimates on all three tasks also correlated positively with the children's math achievement test scores; this relation again held true within grades as well as between them. Moreover, both age trends and individual differences on computational estimation tasks, which do not yield estimates of linearity, correlate with those on the three estimation tasks that do (Booth & Siegler, 2006; LeFevre, Greenham, & Waheed, 1993; Lemaire & Lecacheur, 2002).

Numerical magnitude comparison tasks of the form "Which is larger, N or M" also have been used to study children's representations of numerical magnitudes. Results of these studies, like results from the estimation tasks, suggest that young children generally rely on logarithmic representations of numerical magnitudes (Berch, Foley, Hill, & Ryan, 1999; Case & Okamoto, 1996; Duncan & McFarland, 1980; Sekuler & Mierkiewicz, 1977; Siegler & Robinson, 1982). This inference is based on the presence of both distance and problem-size effects on speed and accuracy (e.g., Banks, 1977; Dehaene, 1997; Moyer & Landauer, 1967). To be specific, the larger the distance between the numbers, and the smaller the magnitudes of the numbers the faster and more accurate the answer is likely to be.

With age and experience, the numerical representation used to solve magnitude comparison problems changes in ways that eliminate errors and reduce distance and problem-size effects on solution times. These changes are consistent with movement toward a linear representation. However, despite the logarithmic function fitting the data decreasingly well with age and experience, no single linear function fits the (untransformed) data better.

There are several possible interpretations of this set of findings. One possibility is that from preschool to adulthood, people use logarithmic representations on numerical magnitude comparison tasks. This might occur because logarithmic representations maintain ordinal properties of numbers, and maintaining ordinal properties is sufficient to yield perfectly accurate magnitude comparisons if there is no noise in the system. A shortcoming of this account, however, is that it does not explain the observed improvements with age and experience in speed and accuracy of performance on the magnitude comparison task, nor the decreasing fit of the logarithmic function to the data. Another possible interpretation is that the reduced fit of the logarithmic function with age and experience, and the observed improvements in speed and accuracy, reflects a shift to a linear representation. The shortcoming of this interpretation is the lack of direct evidence for use of such a linear representation on the task. Drawing clear interpretations about the underlying representation used on the magnitude comparison task is made yet more difficult by the varied influences on performance on the task, including the spatial numerical association of response codes and the spatial numerical association of response codes (SNARC) and compatibility effects (Dehaene, Bossini, & Giraux, 1993; Nuerk, Kaufmann, Zoppoth, & Willmes, 2004).

The present experimental design allowed two tests of the hypothesis that the logarithmic to linear transition influences magnitude comparison, as well as estimation and categorization. First, a transition from logarithmic to linear representations would imply particularly large changes in performance on those magnitude comparison problems involving the largest numbers. The reason is that large magnitudes are differentiated more clearly in linear than in logarithmic functions. For example, on a linear scale, the distance between 2 and 5 equals that between 89 and 92, but on a log scale, the distance between 2 and 5 is more than 30 times as great. Second, to the extent that children apply a common underlying representation of numerical magnitude to number line estimation, numerical categorization, and numerical magnitude comparison tasks, this underlying representation should influence the accuracy and speed of performance on all three tasks. Thus, if the logarithmic to linear transition is evident on the other two tasks, and individual differences in performance on all three tasks are closely related at all ages, then the logarithmic to linear transition may influence magnitude comparison performance as well.

Development of Numerical Categorization

Even infants and young children divide objects, events, and properties into categories that simplify and help them make sense of the world (Rakison & Oakes, 2003). Might young children also divide numbers into categories, and if so, what role might these categories play in their numerical skills?

Several qualities of numbers could lead children to categorize them. Each number can be viewed as a distinct category indicating a quantity of objects or events (Mix, Huttenlocher, & Levine, 2002). Sets of numbers also can be grouped into categories: small/medium/large, odd/even, integer/noninteger, positive/negative, and so on. Multidimensional scaling studies of adults' ratings of similarity among numbers indicate use of all of these categories (e.g., Shepard, Kilpatric, & Cunningham, 1973).

Of particular interest in the present context is whether young children form subjective categories of numbers based on magnitude and whether they use these subjective categories to simplify the task of acquiring new information about numbers. Previous data suggest that preschoolers do form subjective categories of integers below 10 (Murray & Mayer, 1988; Siegler & Robinson, 1982). For example, Siegler and Robinson (1982) found that when preschoolers were asked to categorize the numbers 1–9 as "small," "medium," or "big," the children assigned the largest numbers to the "big" category and the smallest numbers to the "small" category. In addition, the preschoolers' categorizations predicted their magnitude comparison performance above and beyond the effects of distance and number size. Children who grouped many numbers into a single "big numbers" category compared numerical magnitudes less accurately than children who divided numbers more evenly among the categories.

The verbal statements of young children also suggest that they use subjective magnitude categories to help them solve numerical problems. Consider a protocol taken after a 5-year-old completed a 0–100 number line estimation task:

E: What did you use to help you decide where to put the numbers?
C: My brain.
E: What was your brain telling you?
C: Put it like in the middle or the end.
E: How did you know if it went in the middle or if it went on the end?
C: Because the low numbers go at the other end and the high numbers go at the other end.
E: What goes in the middle?
C: The mediumest numbers.

How might children generate such numerical categories? The answer is unknown, but it seems likely that perceptual and associative information plays a large role in early categorization of numbers. Initially, children may divide numbers into the categories "big" and "little" relying on associated perceptual differences between sets of objects with very different numerosities, such as a set of 10 cookies versus a set of 2 cookies. As children gain experience with number words, they may differentiate them into more nuanced categories, such as "very small," "small," "medium," "big," and "very big." Because differences among set sizes are more perceptually salient when the sets are small (Dehaene, 1997; Fechner, 1987/1869) and because children have more experience with small sets than with larger ones, categories at the low end of the number scale may initially be more differentiated, and therefore have fewer members, than categories of larger numbers.

Increasing experience with larger numbers, and increasing conceptual understanding of the number system, seem likely to lead to increasing differentiation among larger numbers and, within the stated range of numbers, to a more equal distribution of numbers among categories.

Consistent with this analysis, Wynn (1992) found that when young 2-year-olds were asked to give an adult one object, they usually responded correctly, but when asked for any other number, they provided an arbitrary number of objects, with the only constraint that the number of objects exceeded one. Thus, they seemed to divide number words into two categories: "1" and "more than 1." Similarly, multidimensional scaling analyses of 3-year-olds' magnitude comparison errors revealed that they only discriminated between "1" and "2–9" (Siegler & Robinson, 1982). In contrast, parallel analyses of 4-year-olds' performance indicated four categories: "1," "2–3," "4–5," and "6–9." A similar scaling analysis of 6-year-old first graders' magnitude comparison solution times in Sekuler and Mierkiewicz (1977) suggested three categories: "1," "2–4," and "5–9." Note that in all cases, the smaller categories included fewer number words, suggesting that categories at the low end of the scale are more differentiated.

The Present Study

This study included two experiments, each with three purposes. In Experiment 1, one purpose was to examine the generality of the transition with age and experience toward increasingly linear representations of numerical magnitudes, that is, representations in which the magnitudes of numbers are well differentiated in all parts of the scale. This transition had previously been documented on three estimation tasks. Here, we tested whether similar transitions occur on two tasks other than estimation: numerical categorization and numerical magnitude comparison. The results with the estimation tasks, together with the consistent correlations of linearity on the estimation tasks with proficiency on general math achievement tests that sample many numerical skills, suggested that developmental changes on the two new numerical magnitude tasks would parallel those previously found on the estimation tasks.

A second purpose of Experiment 1 was to determine whether consistent individual difference patterns are present across number line estimation, numerical categorization, and numerical magnitude comparison tasks. If a common representation of numerical magnitudes underlies individual children's performance on all three tasks, consistent individual differences should be present across the tasks. These consistent individual difference patterns should be present within each grade, as well as when children from different grades are considered.

A third purpose of Experiment 1 was to test whether the development of categorization of numbers in the 0–100 range from kindergarten to second grade parallels the development of categorization of numbers in the 0–10 range during the preschool years. Piaget's construct of vertical decalage (e. g., Piaget & Inhelder, 1956) suggested that such repetitions of developmental sequences in different age ranges often occur. The previously described research on number line estimation indicated that development repeats itself on at least one task relevant to understanding of numerical magnitudes. The question was whether the same would be true for development of numerical categorization. Repetition of the preschool change pattern during the kindergarten to second grade period seemed likely because just as children obtain fairly extensive experience with numbers up to 10 in the preschool period (through activities such as counting objects), older children obtain fairly extensive experience with numbers up to 100 in the early elementary school period (through activities such as arithmetic).

If development of numerical categorization in the 0–100 range during the early elementary school period parallels development of numerical categorization in the 0–10 range in the preschool period, several outcomes like those observed with preschoolers by Siegler and Robinson (1982) would be expected. Young elementary school students should divide the numbers 0–100 increasingly evenly among categories. Individual children's linearity of categorizations should correlate positively with their accuracy of numerical magnitude comparison. This relation should be present above and beyond the effects of distance and problem size on magnitude comparison.

Experiment 2 also had three purposes: to replicate results from Experiment 1, to demonstrate that feedback on categorization can promote more linear categorization of numbers, and to demonstrate that feedback on categorization promotes greater use of linear representations on other tasks as well. These hypotheses were based on findings that children often approach tasks by attending to the most important components first and refining the details later, as in the adage, "divide and conquer" (Siegler & Robinson, 1982). In the domain of numbers, subjective categorization of numerical magnitudes may provide a useful means for implementing the general divide-and-conquer approach, a hypothesis that is discussed further in Experiment 2.

Experiment 1

Method

Participants

Experiment 1 included 90 children (41 boys and 49 girls): 30 kindergartners (mean age = 6.10 years, *SD* = 0.40), 30 first graders (mean age = 7.19 years, *SD* = 0.55), and 30 second graders (mean age = 8.20 years, *SD* = 0.33). Almost all children in the sample (95%) were Caucasian. Children were recruited from two public schools in the Pittsburgh area; the percentages of children in these schools who were eligible for the free or reduced-fee lunch program were 33% and 17%. Participation in the experiment was voluntary; children received no tangible reward for their involvement. The experimenter was a female Caucasian graduate student.

Materials and Procedure

Children met one-on-one with the experimenter in a single session in the Spring of their school year and performed three tasks: number line estimation, numerical categorization, and numerical magnitude comparison. The order of task presentation was counterbalanced, and all sessions were audio recorded.

Number line estimation. Children were initially given a sheet of paper with a 25 cm horizontal line printed across the middle. The number "0" was below the left end of the line, and the number "100" was below the right end of the line; a small vertical hatch mark was above each number.

The experimenter explained the task by saying "A number line is a line with numbers across it. The numbers on the line go from the smallest number to the largest number, and the numbers go in order, so each number has its very own spot on the number line." Children were then told that they should draw a vertical line where they thought each number belonged on the number line and were given practice marking the positions of 0 and 100 on the line to orient them to the endpoints and to ensure that they understood the task. On these practice trials, the experimenter provided feedback if the child did not place the marks for 0 and 100 at the appropriate ends of the line.

Next, 22 number line test trials were presented. On each trial, the child was presented a sheet of paper identical to the practice page, except that there was a numeral 6 cm above the middle of the line. The 22 numbers sampled the range from 0 to 100. To improve our ability to distinguish between linear and logarithmic estimation patterns, numbers in the first decade were slightly oversampled, with four numbers included between 0 and 10, as opposed to two numbers from each of the remaining decades. The numbers presented were 2, 3, 5, 8, 12, 17, 21, 26, 34, 39, 42, 46, 54, 58, 61, 67, 73, 78, 82, 89, 92, and 97. A different random order of the numbers was generated for each child.

Numerical categorization. To introduce the categorization task, the experimenter said, "I'm going to ask you what you think about some of the numbers between 0 and 100. Some of these numbers are really small, some are small, some are medium, some are big, and some are really big. I'm going to say a number, and you need to tell me if you think the number is a really small number, a small number, a medium number, a big number, or a really big number."

The experimenter then set out five boxes that were labeled, from left to right, "really small," "small," "medium," "big," and "really big," with the labels facing the child. The experimenter read the label on each box from left to right, told children that they could refer to these boxes to help them remember all their choices, and repeated the names on them frequently to help children remember them. Children were first asked to categorize 0 and 100 to orient them to the endpoints and to ensure that they understood the task. On these practice trials, the experimenter provided feedback if a participant did not categorize 0 as "really small" and 100 as "really big". To facilitate comparison of the representations, the same 22 numbers were used as stimuli as on the number line task. Test trials were ordered randomly, with children being asked to indicate on each trial whether *N* was "really small," "small," "medium," "big," or "really big."

Numerical magnitude comparison. Because of the impracticality of presenting all possible pairs of the 22 integers used on the other two tasks, we selected 11 of the 22 numbers—half of the numbers from each decade that were presented on the other two tasks—and asked children to compare the magnitudes of the 55 possible pairs. The numbers chosen were 2, 8, 12, 26, 34, 42, 54, 67, 73, 89, and 97. The pairs of numbers were randomly assigned for each child to one of two blocks: the "which is more" block and the "which is less" block. Each pair of numbers appeared equally often on "which is more" and "which is less" trials. The order of the two blocks was counterbalanced, and the order of pairs within each block varied randomly across children. Children answered all 28 problems from one block (e.g., the "which is more" questions) before answering the 27 problems from the other block. The task instructions were simply, "I'm going to tell you two numbers between 0 and 100, and you tell me which number is more (less).'' Solution times were extracted from the audio recordings using a stopwatch; the times measured the period between when the experimenter finished asking the question and when the student responded.

Results and Discussion

We first examined changes with age on each of the three tasks individually. Then, we examined relations among individual differences on the three tasks. Here and throughout the study, all post hoc comparisons were done with Tukey honestly significant difference (HSD) tests.

Number Line Estimation

To analyze the accuracy of number line estimates, we computed each child's percent absolute error:

$$\left|\frac{\text{Estimate š correct quantity}}{\text{Scale of estimates}}\right|$$

To illustrate, if a child was asked to estimate the location of 61 on a 0 to 100 number line and placed the mark at the point that corresponded to 73, the absolute error would be 12%, $|(73 - 61)/100|$.

Accuracy of number line estimates increased with age; mean percent absolute error decreased from 15% to 13% to 10% between kindergarten and second grade (*SE*s = 0.01), $F(2, 87) = 7.56$, $p < .01$, $\eta^2 = .15$.

To test whether the hypothesized logarithmic to linear shift was present, we examined the fit of the linear and logarithmic functions to the median estimate of each number's magnitude that was generated by children in each grade. Medians rather than means were used to minimize the effect of outliers.

As shown in Figure 1, the variance in median number line estimates that was accounted for by the linear function steadily increased with age (from 91% to 95% to 98%). Conversely, the variance accounted for by the logarithmic function steadily decreased with age (from 95% to 91% to 86%). Paired sample *t* tests were used to compare for each grade the absolute value of the difference between the children's median estimate for each number and the predicted estimate for that number generated by the best fitting linear and logarithmic functions. Consistent with previous findings of increasing linearity of estimates on 0–100 number lines in this age range, the second graders' median estimates were significantly better fit by the linear function than by the logarithmic function ($R^2_{\text{lin}} = .98$ vs. $R^2_{\text{log}} = .86$), $t(21) = 3.42$, $p < .01$, $d = 0.99$. Also as predicted, the fit of the two functions

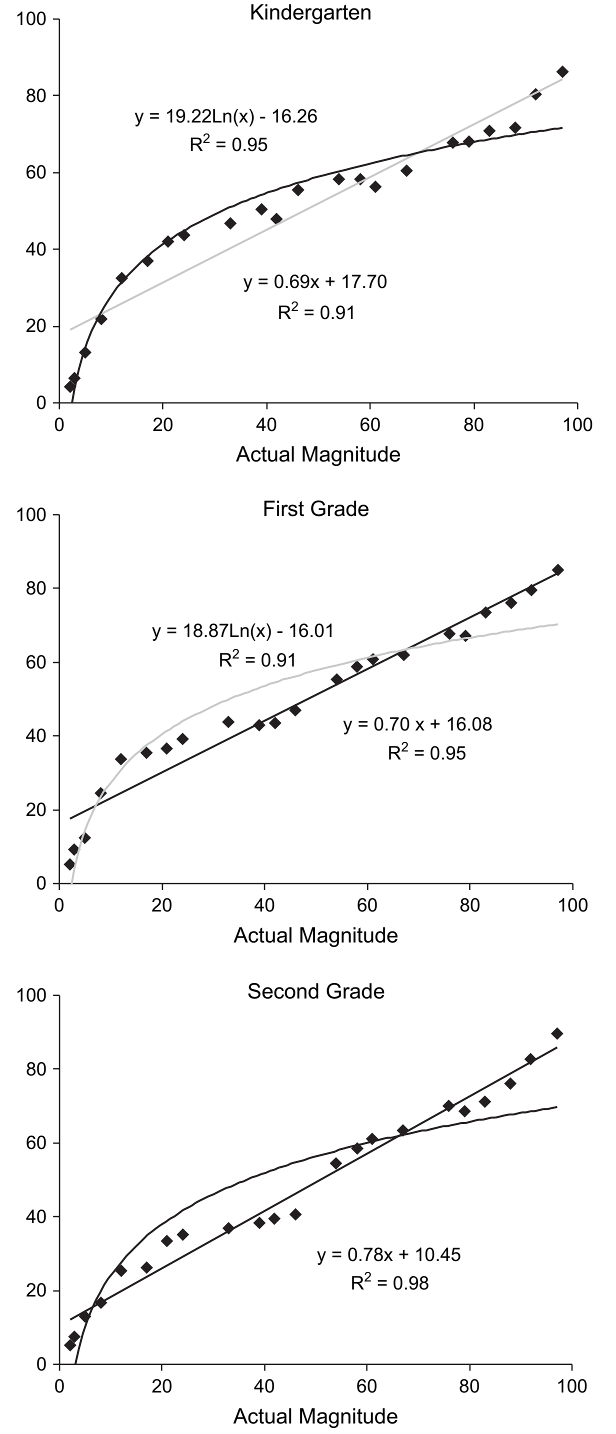

Figure 1. Best fitting function relating the number that was presented to the median number line estimate for that number generated by kindergartners, first graders, and second graders in Experiment 1. Both logarithmic and linear functions are shown in cases where their fits do not differ significantly; only the better fitting function is shown in cases where the fits do differ significantly.

did not differ for first graders (R^2_{log} = .91 vs. R^2_{lin} = .95). For kindergartners, the difference was in the expected direction but was not significant (R^2_{log} = .95 vs. R^2_{lin} = .91).

We next examined the fit of the linear and logarithmic functions to individual children's number line estimates. The percent variance accounted for by the best fitting linear function for each child's estimates increased with age and experience, $F(2, 87) = 8.79$, $p < .0005$, $\eta^2 = .17$. The linear function accounted for an average of 66% of the variance in individual kindergartners' estimates, 74% in first graders', and 88% in second graders'. The percentage of children whose estimates were better fit by the linear function than by the logarithmic one increased from 43% to 53% to 80% between kindergarten and second grade, $\chi^2(2, N = 90) = 8.90$, $p = .01$, $V = .32$.

In addition to estimates becoming more linear with increasing age and experience, the slopes of the best fitting linear function also progressed toward the ideal slope relating estimates to the numbers presented (1.00). Mean slopes of individual children's number line estimates were 0.63, 0.67, and 0.76 for kindergartners, first graders, and second graders, respectively, $F(2, 87) = 5.28$, $p < .05$, $\eta^2 = .11$.

Number Categorization

We predicted that kindergartners would categorize fewer numbers as "very small" or "small" than as "big" or "very big," reflecting a logarithmic distribution in which differences between smaller numbers are exaggerated and differences between larger numbers are compressed. We also predicted that second graders would show this difference to a lesser degree or not at all. To test the hypothesis, we compared the number of numbers that children assigned to the "really small" and "small" categories with the number of numbers they assigned to the "really big" and "big" categories. As predicted, the difference between the number of numbers in the two pairs of categories decreased with age, $F(2, 87) = 4.42$, $p = .02$, $\eta^2 = .09$. Kindergartners assigned considerably more numbers to the two big categories than to the two small ones (*M*s = 9.40 vs. 4.60), $t(29) = 5.24$, $p < .0005$, $d = 1.54$; first graders assigned somewhat more numbers to the two big categories (*M*s = 8.43 vs. 5.40), $t(29) = 3.37$, $p < .01$, $d = 1.07$; and second graders assigned almost the same number to them (*M*s = 8.37 vs. 7.17, *ns*).

To test the fit of the logarithmic and linear representations to all categorizations (including ones in the "medium" category), we assigned each categorization a numerical value ranging from 1 for the "really small" category to 5 for the "really big" category, computed the mean value for each of the 22 numbers that children were asked to categorize, and then computed the absolute deviations of the mean categorization for each number from the best fitting linear and logarithmic functions. The variance in mean categorization values for the 22 numbers that was accounted for by the best fitting linear function increased with grade from 81% to 90% to 95%, and the variance accounted for by the logarithmic function decreased from 97% to 96% to 92%. As shown in Figure 2, kindergartners' mean categorizations of the 22 numbers were better fit by the logarithmic function than by the linear function (R^2_{log} = .97 vs. R^2_{lin} = .81), $t(21) = 3.04$, $p < .01$, $d = 0.86$. As expected, the fit of the two functions did not differ for first graders (R^2_{log} = .96 vs. R^2_{lin} = .90). For second graders, the difference was in the expected direction but was not significant (R^2_{log} = .92 vs. R^2_{lin} = .95).

To test whether these findings regarding group means also fit individual children's performance, we examined the fit of the linear and the logarithmic functions to each child's categorizations. The percentage of children whose categorizations were better fit by the linear function than by the logarithmic one increased from 27% to 63% to 80% between kindergarten and second grade, $\chi^2(2, N = 90) = 18.19$, $p < .0005$, $V = .45$. Similarly, the mean variance in individual children's categorizations accounted for by the best fitting linear function increased from 55% to 72% to 84% from kindergarten to first grade to second grade, $F(2, 87) = 19.28$, $p < .0005$, $\eta^2 = .31$.

Numerical Magnitude Comparison

Neither accuracy nor speed differed significantly between the "which is more" and the "which is less" trials. Therefore, results from the two types of trials for each problem were grouped together.

Accuracy of magnitude comparisons increased with grade. Kindergartners were correct on 89% (*SE* = 2.21) of the comparisons, first graders on 95% (*SE* = 1.39), and second graders on 98% (*SE* = 0.97), $F(2, 87) = 8.26$, $p < .01$, $\eta^2 = .16$. Solution times for correct responses decreased with age, from 2.11 s for kindergartners to 1.43 s for second graders, though the differences were not significant.

To determine if young children's magnitude comparisons reflected a logarithmic representation, we examined the degree to which accuracy and solution times could be predicted from the two variables that have been associated with a logarithmic representation in previous studies of numerical magnitude comparison—distance between the numbers being compared and problem size (here operationalized as

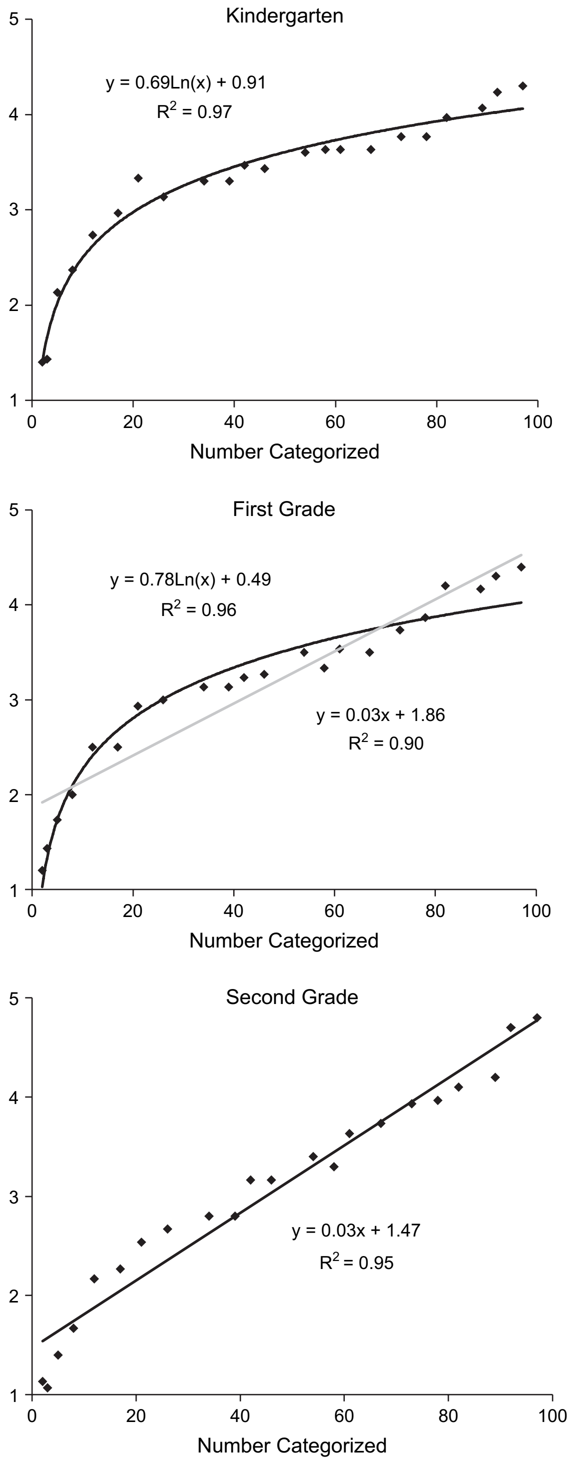

Figure 2. Best fitting function relating the number that was presented to the mean categorization value for that number generated by kindergartners, first graders, and second graders in Experiment 1. Categorizations of "really small" were assigned a value of 1, categorizations of "small" were assigned a value of 2, categorizations of "medium" were assigned a value of 3, and so on. Both logarithmic and linear functions are shown in cases where their fits did not differ significantly; only the better fitting function is shown in cases where the fits did differ significantly.

the sum of the numbers being compared). In a multiple regression analysis of the mean percent correct on each of the 55 problems, the two variables accounted for 46% of the variance for kindergartners, 50% for first graders, and 29% for second graders. A similar pattern was apparent for solution times on correctly solved problems. The two variables accounted for 63% of the variance in kindergartners' times, 52% for first graders', and 31% for second graders'. Thus, the logarithmic pattern was strongly present for kindergartners' numerical magnitude comparisons, at least for the solution time measure, but only weakly present for second graders' comparisons.

Although there was no single, straightforward measure of linearity of magnitude comparisons that predicted accuracy and solution times, the decrease with age in the fit of the logarithmic function to the magnitude comparison data, together with the increasing accuracy of the magnitude comparisons, suggested that children might be moving toward a linear representation on this task, as on number line estimation and categorization. As noted earlier, such a change in representations would imply particularly large improvements on comparisons involving large numbers. There was substantial evidence for such changes. One type of evidence involved the slopes of the best fitting function relating problem size (as measured by the sum of the numbers being compared) to solution times and accuracy. The relation between problem size and solution times decreased significantly with age. The mean of the slopes relating problem size to each child's solution times decreased from 0.007 ($SE = 0.001$) for kindergartners to 0.004 ($SE = 0.001$) for first graders to 0.001 ($SE = 0.0004$) for second graders, $F(2, 85) = 10.18$, $p < .0005$, $\eta^2 = .19$. The group-level slope relating problem size and percent correct also decreased with age, from 0.21 for kindergartners to 0.19 for first graders to 0.14 for second graders. Because of the binary nature of the accuracy data for individual children, no comparable statistical test could be performed on them.

Analyzing changes with age in the difference between individual children's percentage of errors on the smallest and largest quarter of problems allowed a statistical test of whether the accuracy data showed the same pattern as the solution time data. The difference between percentage of errors on the largest and smallest 14 problems decreased from kindergarten (16%, $SE = 3.43$) to first grade (8%, $SE = 3.15$) to second grade (3%, $SE = 1.63$), $F(2, 87) = 5.83$, $p < .01$, $\eta^2 = .12$. These data, like the above-described data on slopes of solution times, suggested that the same type of transition from logarithmic to linear representations that was present on the number line estimation and categorization tasks might also influence magnitude comparisons.

Previous studies of multidigit numerical magnitude comparison have shown that speed and accuracy are generally greater when the relative magnitudes of the units digits and the numbers as a whole are compatible than when they are incompatible (Nuerk, Weger, & Willmes, 2001). To illustrate, 26 versus 73 is an incompatible problem because the smaller number (26) has the greater units digit (6). In contrast, 23 versus 76 is a compatible problem because the larger number also has the larger units digit. The effect is theoretically important because it shows that overall magnitude is not the only factor influencing magnitude comparisons.

To test whether such compatibility effects influenced magnitude comparisons in the present experiment, we matched each of the 16 incompatible problems with the compatible problem whose distance between the two numbers was the closest possible. This matching resulted in distances for the incompatible and compatible problems of 28.9 and 29.0, respectively. The average problem sizes (sums of the numbers) also were similar, 87 for the incompatible problems and 80 for the compatible ones.

In keeping with past findings, compatibility influenced kindergartners' and first graders' performance. Kindergartners correctly answered 90% of compatible and 83% of incompatible problems, $t(29) = 2.96$, $p < .01$, $d = 0.43$. First graders correctly answered 97% of compatible and 93% of incompatible problems, $t(29) = 2.24$, $p = .03$, $d = 0.40$. Second graders were equally accurate on both types of problems (98% vs. 97%).

Relations Among Tasks

To determine the degree to which consistent individual differences in quality of performance were present on the three tasks, we computed correlations among (a) the variance in each child's number line estimates that was accounted for by the best fitting linear function, (b) the variance in each child's number categorizations that was accounted for by the best fitting linear function, and (c) each child's percent correct on the magnitude comparison task.

Individual differences on the three tasks proved to be strongly correlated: $r(88) = .83$, $p < .0005$ for performance on the number line and numerical categorization tasks, $r(88) = .73$, $p < .0005$ for performance on the categorization and magnitude comparison tasks, and $r(88) = .71$ $p < .0005$, for performance on the number line and magnitude comparison tasks.

These strong and consistent relations in individual children's performance on the three tasks were present within grades as well as between them. As shown in Table 1, kindergartners' correlations for the three pairs of tasks ranged from $r = .63$ to $r = .82$, first graders' correlations ranged from $r = .56$ to $r = .82$, and second graders' correlations ranged from $r = .80$ to $r = .83$ (all *df*s = 28; all *p*s < .0005). These strong correlations were consistent with the view that performance on all three tasks reflected the same underlying representation.

We also tested the hypothesis that subjective categorizations of numbers would contribute to magnitude comparison performance above and beyond the contribution of the distance and problem-size effects. To test this hypothesis, we first computed a *categorization model score* for each problem. This was done by assuming that when a child assigned the two numbers in a magnitude comparison problem to different, correctly ordered categories (e.g., by saying that 26 was a small number and 46 was a medium number), the magnitude comparison answer would be correct; that when a child assigned the two numbers to different, incorrectly ordered categories (e.g., by saying that 46 was a small number and 26 a medium number), the magnitude comparison answer would be incorrect; and that when the child assigned the two numbers to the same category, the answer had a 50% probability of being correct. The mean of the children's scores for each problem was the categorization model score for that problem.

These categorization model scores were used within hierarchical linear regression analyses of children's numerical magnitude comparisons. Distance between the numbers and the sum of the numbers for each problem were entered first; then, the categorization model score for the problems was entered to determine whether categorizations accounted for additional variance.

For kindergartners, the distance and problem-size effects together accounted for 46% of the

Table 1
Within-Grade Correlations (rs) of Linearity of Number Line Estimation, Linearity of Categorization, and Accuracy of Magnitude Comparison (all dfs = 28, all ps < .0005)

	Tasks		
Age	Number line and categorization	Number line and magnitude comparison	Categorization and magnitude comparison
Kindergarten	.82	.69	.63
First grade	.82	.56	.70
Second grade	.80	.83	.82

variance in percent correct on each problem; considering the categorization model scores increased this only to 48% (*ns*) However, for first and second graders, the categorization model scores accounted for significant variance beyond the distance and problem-size effects. For first graders, the explained variance increased from 50% to 65% ($p <$ .0005); for second graders, it increased from 29% to 38% ($p = .01$).

To summarize, the results of Experiment 1 were consistent with the three hypotheses that motivated the experiment. First, the developmental sequence toward decreasing use of logarithmic representations and increasing use of linear representations was present on the numerical categorization task as well as on the number line task, thus demonstrating that the transition extends beyond the estimation tasks on which it had been demonstrated previously. On the magnitude comparison task, the changes with age and experience toward decreasing fit of the logarithmic function, increasing accuracy, reductions in problem-size effects, and strong correlations with linearity on the other two tasks suggested that a similar logarithmic to linear transition was present on that task too.

Second, individual differences in proficiency on the three tasks were highly correlated. The relations among linearity of number line estimation, linearity of categorization, and percent correct magnitude comparisons were strongly present within each grade as well as between them. The relations between the number line estimation and the two nonestimation tasks in the present experiment were as strong as those among the three estimation tasks in Booth and Siegler (2006). This again highlights the generality of the transition.

Third, the development of categorization of numbers in the 0–100 range from kindergarten to second grade paralleled the development of categorization of numbers in the 0–10 range during the preschool years. Among the parallels were changes with age and experience toward increasingly even division of numbers among categories and toward greater differentiation of numbers at the high end of the range. Other parallels were that linearity of categorization and accuracy of magnitude comparison were positively correlated and that categorization predicted magnitude comparison accuracy above and beyond the distance and problem-size effects. These last two findings suggested that providing experiences that led children to categorize numbers in a more linear pattern might also produce more linear patterns of performance on other numerical tasks. This hypothesis was tested in Experiment 2.

Experiment 2

Experiment 2 was designed to test whether providing feedback on categorizations would lead not only to improved categorization but also to improved performance on other numerical tasks as well. During a training period, kindergartners were presented experience with categorization that was intended to promote division of the numbers 0–100 into five equal size categories and thereby to promote a linear representation of the numbers. The same three tasks as in Experiment 1—numerical categorization, number line estimation, and numerical magnitude comparison—were used as pretest and posttest measures.

The experiment allowed tests of three main hypotheses. The first hypothesis was that the kindergartners' pretest performance would replicate the performance of age peers in Experiment 1. The second hypothesis was that providing feedback that encouraged children to sort numbers into equal size categories would generate more linear categorizations than providing identical sorting experience but no feedback. The third hypothesis was that the categorization feedback would generalize beyond the categorization task and lead to more successful performance on the number line and magnitude comparison tasks as well. The logic was that if subjective categorizations of numbers influence representations of numerical magnitude, then increasing the linearity of the representations in the categorization context should also increase the linearity of performance on the other two tasks.

Underlying the experiment was the belief that "divide and conquer" is a highly general learning strategy that both children and adults use in many situations. Examples of the divide-and-conquer approach are omnipresent. Explorers map out the most important landmarks—coastlines, mountains, and rivers—before filling in the details of the terrain. Writers generate outlines before drafting the specific sentences in articles. Artists sketch the major features of paintings before filling in the details.

Children use the same divide-and-conquer approach from young ages onward. Toddlers identify the most important words and rely on them alone in telegraphic speech (Braine, 1971). Kindergartners solve balance scale, shadows projection, probability, and other problems by identifying one crucial variable and consistently relying on it, even though other variables are also influential (Siegler, 1981). Beginning readers focus on the first letter or the first and last letters to identify words (Ehri & Wilce, 1985).

In the domain of numbers, subjective categorization of numerical magnitudes may provide a useful means for implementing the general divide-and-conquer approach. For example, a kindergartner or first grader who is learning single-digit addition might view the numbers 11 – 19 as ''medium size'' because they are larger than the familiar numbers 1 – 10 but smaller than the many minimally understood numbers in the 20s and beyond. It seems likely that over time, experience with numbers provides implicit feedback about how numbers can be categorized and that this feedback leads to changes in categorization. This belief led to the hypothesis that explicit feedback on categorizations could promote improved categorization. If children divide and conquer new numerical ranges by creating a few categories of numbers within them, and if explicit feedback improves the categorization scheme by creating a more equal (linear) distribution of numbers within it, then the more linear categorization of numbers may produce more linear, and more accurate, performance on other tasks in the same numerical range.

Method

Participants

Experiment 2 included 40 kindergartners (13 boys and 27 girls) of mean age 5.93 years ($SD = 0.36$). All of the children were recruited from two schools in the same school district as in Experiment 1. The populations of the schools were similar to those in Experiment 1 in both racial composition and economic status; 93% of the children were Caucasian and 12% were eligible for the free or reduced-fee lunch program. Participation in the experiment was voluntary, and children received no tangible reward for their involvement. The experimenter was the same female, Caucasian, graduate student as in Experiment 1.

Procedure

Overview. The experiment included six sessions. In Sessions 1 and 6, children completed a pretest and a posttest, which consisted of the same three tasks, administered in the same way, as in Experiment 1.

In Sessions 2 – 5, children gained experience with three categorization training tasks: the midpoint categorization, variable number categorization, and triad tasks. The midpoint categorization and triad tasks were presented in that order in Sessions 2 and 3; the triad and variable number categorization tasks were presented in that order in Sessions 4 and 5. In all four training sessions, children in the feedback condition received feedback regarding the correct answer after each response, whereas children in the no-feedback condition were only provided general encouragement.

Stimuli. To generate stimuli for the three training tasks, the numbers between 0 and 100 were assigned to five equal size categories. The ''really small'' category included the numbers 1 – 20, the ''small'' category 21 – 40, the ''medium'' category 41 – 60, the ''big'' category 61 – 80, and the ''really big'' category 81 – 100. The midpoints of these categories—10, 30, 50, 70, and 90—were used as the stimuli for the midpoint categorization task. On the other two tasks, 30 other numbers were used, 6 numbers from each of the five categories. These numbers included the 22 from Experiment 1 and also 27, 38, 45, 57, 64, 68, 83, and 91.

Midpoints categorization task. On the midpoints categorization task, children were presented 30 cards, each showing a single number and asked to assign each number to the appropriate category. Each category midpoint—10, 30, 50, 70, and 90—was shown on 6 of the 30 cards.

Five baskets were arranged in front of the child from left to right. Each basket had a category label (e.g., ''really small'') and a picture of the appropriate size bear. To the left of the leftmost basket was a card with the numeral ''0.'' To the right of the rightmost basket was a card with the numeral ''100.'' In front of each basket was a stack of pennies corresponding to the category midpoint (e.g, a stack of 10 pennies in front of the ''really small'' box) and the corresponding numeral (e.g., ''10''). Children were given the following instructions:

> Mama Bear has five baby bears—one that is really small, one that is small, one that is medium, one that is big, and one that is really big. She gives each of them between 0 and 100 pennies when they help her clean the house. To be fair, she always gives really small bear 10 pennies—a really small number of pennies. She gives small bear 30 pennies—a small number of pennies. She gives medium bear 50 pennies—a medium number of pennies. She gives big bear 70 pennies—a big number of pennies. And she gives really big bear 90 pennies—a really big number of pennies. In this game, your job is to help Mama Bear give each bear his pennies. I'll show you a number and you give it to just the right bear.

After these instructions, but before children began to categorize the numbers, the stacks of pennies and all the numeral cards, except for the 0 and 100 cards, were removed.

The five midpoints were presented in one of two semirandom orders (random except for the stipulation that there be no consecutive presentations of the same number). All children received the numbers in one predetermined order during Session 2 and in a different predetermined order during Session 3.

The two experimental groups differed only in whether feedback was provided on each trial. Following correct responses, children in the feedback condition were told, for example, ''That's right; 70 is big.'' Following incorrect responses, children in the feedback condition were told, for example, ''Oops! 70 is big. Let's give it to big bear.'' Children in the no-feedback condition were presented identical problems and periodically given general encouragement, but they were never provided information about the quality of their choices.

Triad task. Children were presented 30 triad problems—six trials for each of the five categories. Each triad problem included three numbers: a standard, an incorrect choice, and a correct choice. The standard was always the category midpoint. The correct choice was another number from the same category as the midpoint. The incorrect choice was a number from a different category, and therefore, more distant from the midpoint than the correct choice was. The triads were constructed, so that on the six trials for each category, the incorrect choice came at least once from each of the other four categories.

At the beginning of the triad task, children were told that they were going to help Mama Bear figure out some other numbers that are ''really small,'' ''small,'' ''medium,'' ''big,'' or ''really big'' because sometimes she likes to give each bear a few more or a few less pennies than usual. The experimenter then stated that in this game, there was a rule: ''Numbers are really small if they are close to 10 when you count; they're small if they are close to 30 when you count; they're medium if they are close to 50 when you count; they're big if they are close to 70 when you count; and they're really big if they are close to 90 when you count.''

The triads were presented one at a time on a computer screen. On each trial, the experimenter identified the numerals and the category. For example, the experimenter said, ''This is 30; 30 is a small number. Which of these two numbers, 26 or 46, also is a small number?'' The triads were presented in one of four semirandom orders (random except for the stipulation that on consecutive trials, the incorrect choice could not come from the same category). All children were presented the triads in the same order during each training session.

After correct choices on the triad task, children in the feedback group were told, for example, ''That's right! 46 goes with 50 because they are both medium numbers.'' After incorrect responses, they were told, for example, ''Oops! Those two numbers don't go together. Fifty is a medium number and 27 is a small number; 46 is closer to 50 when you count, so 46 and 50 go together.'' Children in the no-feedback group were given identical instructions and problems and also were given general encouragement but no specific feedback.

Variable numbers categorization task. At the outset of this task, children were asked to sort 30 cards—six different numbers from each of the five categories. The numbers were presented in one of two semirandom orders (random except for the stipulation that consecutive numbers could not be from the same category). All children received the 30 numbers in one predetermined order during Session 4 and in a different predetermined order during Session 5.

The materials, procedure, and feedback were the same as in the midpoints categorization task. The instructions were slightly modified, so that the categorization rule was more inclusive: ''Remember, really small bear always gets a really small number of pennies—a number close to 10 when you count. Small bear gets a small number of pennies—a number close to 30. Medium bear gets a medium number of pennies—a number close to 50. Big bear gets a big number of pennies—a number close to 70. Really big bear gets a really big number of pennies—a number close to 90.''

Results

We first focus on the pretest data to determine whether they replicated the results of Experiment 1. Then, we test whether feedback during training led to a more even categorization of the numbers on the posttest. After this, we examine whether changes in categorization generalized on the posttest to the number line and magnitude comparison tasks. Finally, we perform a path analysis to better understand the process that led to pretest–posttest changes.

Pretest Performance

Kindergartners' pretest performance replicated that of peers in Experiment 1 on all three tasks.

Number categorization. As in Experiment 1, the mean values of the numbers that children assigned to each category indicated that they understood the task: 3.27, 10.01, 37.28, 61.50, and 72.56 for the five categories, ordered from ''really small'' to ''really big.''

Also as in Experiment 1, kindergartners placed fewer numbers in the smaller categories than in the larger ones, reflecting a logarithmic distribution. A paired sample *t* test indicated that children assigned fewer numbers to the "really small" and "small" categories ($M = 4.35$, $SE = 0.31$) than to the "big" and "really big" categories ($M = 10.80$, $SE = 0.79$), $t(39) = 6.46$, $p < .0005$, $d = 1.70$.

Other analyses supported the view that as in Experiment 1, kindergartners' pretest categorizations were distributed logarithmically. The kindergartners' mean categorizations of each number, summed across individuals, were better fit by the logarithmic function than by the linear one ($R^2_{\log} = .96$ vs. $R^2_{\text{lin}} = .73$), $t(21) = 3.90$, $p < .01$, $d = 1.06$). The variance in mean categorizations accounted for by the logarithmic function was almost identical to that in Experiment 1 (96% vs. 97%). Similarly, the logarithmic function provided a better fit to the categorizations of 80% of individual children identical to the percentage in Experiment 1.

Number line estimation. The percent absolute error of children's pretest number line estimates in Experiment 2 (19%) resembled that of peers in Experiment 1 (15%). Also as in Experiment 1, the logarithmic function provided a better fit to the kindergartners' median number line estimates than did the linear function ($R^2_{\log} = .96$ vs. $R^2_{\text{lin}} = .78$), $t(21) = 3.60$, $p < .01$, $d = 1.03$. The variance in median number line estimates accounted for by the logarithmic function was almost identical to that in Experiment 1 (96% vs. 95%). The best fitting linear function accounted for a mean of 57% of the variance in the estimates of individual children compared to 66% in Experiment 1. The slopes were also similar across experiments. The mean slope of individual kindergartners' estimates was 0.63 in Experiment 1, and 0.57 in Experiment 2. The logarithmic function provided a better fit to the number line estimates of 83% of the kindergartners. This was considerably more than the 57% of kindergartners for whom this was true in Experiment 1, but highly similar to the 80% and 81% of kindergartners for whom it was true in the two experiments of Siegler and Booth (2004).

Numerical magnitude comparison. The kindergartners correctly answered 85% of problems quite similar to the 89% in Experiment 1. Distance between the numbers being compared and the sum of the numbers being compared together accounted for 59% of the variance in the kindergartners' response times (vs. 63% in Experiment 1) and 58% of the variance in their accuracy (vs. 46% in Experiment 1). Also as in Experiment 1, the kindergartners were more accurate on compatible than on incompatible problems (89% vs. 78% correct), $t(39) = 5.38$, $p < .0005$, $d = 0.76$.

Relations among tasks. As in Experiment 1, proficiency on the three tasks was closely related. The linearity of each child's number categorizations correlated $r(38) = .65$, $p < .0005$ with the linearity of the child's number line estimates. The fit of the linear function to a child's numerical categorizations correlated $r(38) = .59$, $p < .0005$ with the child's percent correct on the magnitude comparison task. The linearity of number line estimates correlated $r(38) = .57$, $p < .0005$ with percent correct magnitude comparisons. These relations were somewhat weaker than those of kindergartners in Experiment 1 (*r*s = .82, .69, and .63, respectively), but the relations were quite strong in both cases.

Pretest – Posttest Changes

Number categorization. Consistent with the hypothesis that the feedback would influence children's categorizations, the percent variance in mean categorization for each number that was accounted for by the linear function increased in the feedback condition from 70% on the pretest to 96% on the posttest, $t(21) = 2.81$, $p = .01$, $d = 0.63$ (Figure 3). Linearity also increased in the no-feedback group from 73% on the pretest to 88% on the posttest, $t(21) = 2.40$, $p < .05$, $d = 0.42$. This latter finding indicated that performing the categorization training tasks, even in the absence of feedback, influenced performance on the posttest categorization task, though the effect was not as great as when feedback was provided.

Analysis of individual children's performance provided converging evidence. Among children in the feedback group, the percent variance in individual children's categorizations that was accounted for by the linear function increased from a mean of 49% on the pretest to a mean of 75% on the posttest, $t(19) = 5.19$, $p < .0005$, $d = 1.09$. Among children in the no-feedback group, the change was less dramatic; the best fitting linear function accounted for a mean of 52% of the variance in individuals' categorizations on the pretest, versus 61% of the variance on the posttest, $t(19) = 2.57$, $p < .05$, $d = 0.42$. An analysis of covariance (ANCOVA) of the mean variance in individual children's categorizations that was accounted for by the linear function at posttest, holding constant differences in pretest scores, revealed a main effect of feedback, $F(1, 37) = 7.35$, $p = .01$, $\eta^2 = .17$. This indicated that the linear function provided a better fit to the posttest categorizations of children who had received feedback.

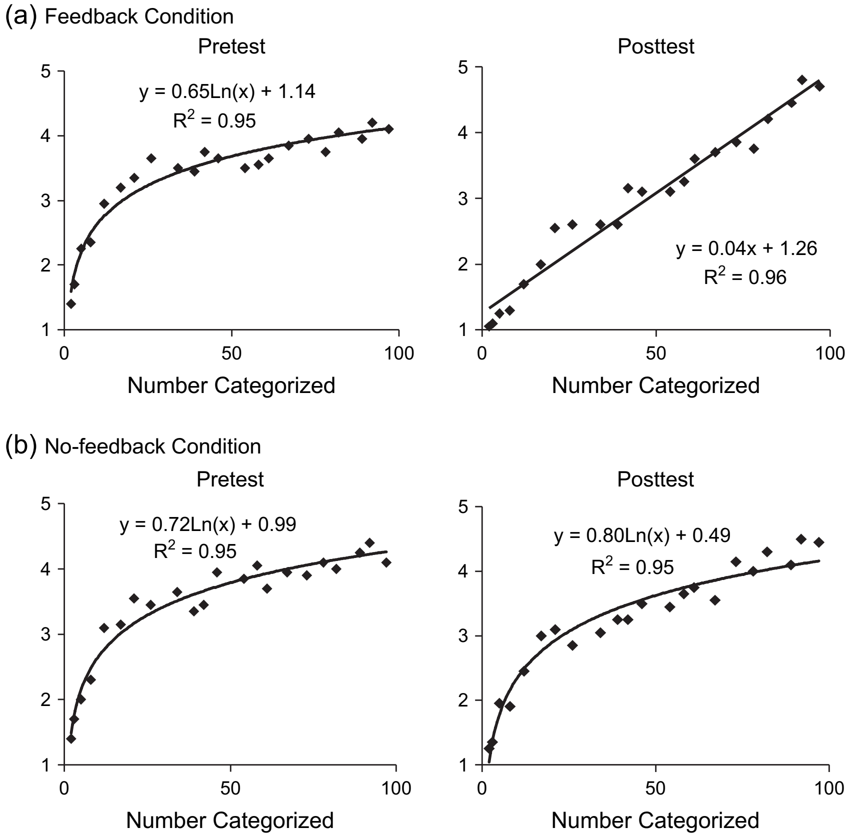

Figure 3. Best fitting function relating the number that was presented to the mean categorization value for that number generated by children in the feedback condition (top) and the no-feedback condition (bottom) on the pretest and posttest of Experiment 2.

Number line estimation. Although children were given no experience with the number line task during training, the number line estimates of children who received feedback on their categorizations became dramatically more linear from pretest to posttest (Figure 4). Among children who received feedback on their categorizations, the percent variance in median number line estimates accounted for by the linear function increased from 77% on the pretest to 96% on the posttest, $t(21) = 5.60$, $p < .0005$, $d = 1.15$. Among children who were not given feedback, the percent variance accounted for by the linear function showed a smaller increase from 70% to 79% (*ns*).

Analysis of changes in the linearity of individual children's number line estimates provided converging evidence. Among children who received feedback, the percent variance in number line estimates accounted for by the linear function increased from a mean of 56% on the pretest to a mean of 76% on the posttest, $t(19) = 4.22$, $p < .0005$, $d = 4.00$. In contrast, among children in the no-feedback group, the mean variance in individual children's estimates accounted for by the linear function decreased slightly from 60% on the pretest to 58% on the posttest. An ANCOVA that statistically controlled for pretest scores indicated that the increase in linearity was greater among children who received feedback, $F(1, 37) = 14.88$, $p < .0005$, $\eta^2 = .29$.

The categorization feedback also led to increases in the slopes of the kindergartners' number line estimates. Among children who received feedback on their categorizations of numbers, the mean slope of the best fitting linear function for the number line estimates increased from a mean of 0.55 ($SE = 0.03$) on the pretest to a mean of 0.71 ($SE = 0.04$) on the posttest, $t(19) = 3.86$, $p < .01$, $d = 0.93$. In contrast, the mean slope of number line estimates among children who did not receive feedback on their categorizations was 0.58 ($SE = 0.03$) on both pretest and posttest. An ANCOVA that controlled for differences in pretest slopes indicated that the pretest–posttest increases in slopes were greater for children who received feedback, $F(1, 37) = 12.52$, $p < .01$, $\eta^2 = .25$.

Numerical magnitude comparison. There was no change from pretest to posttest in the accuracy of

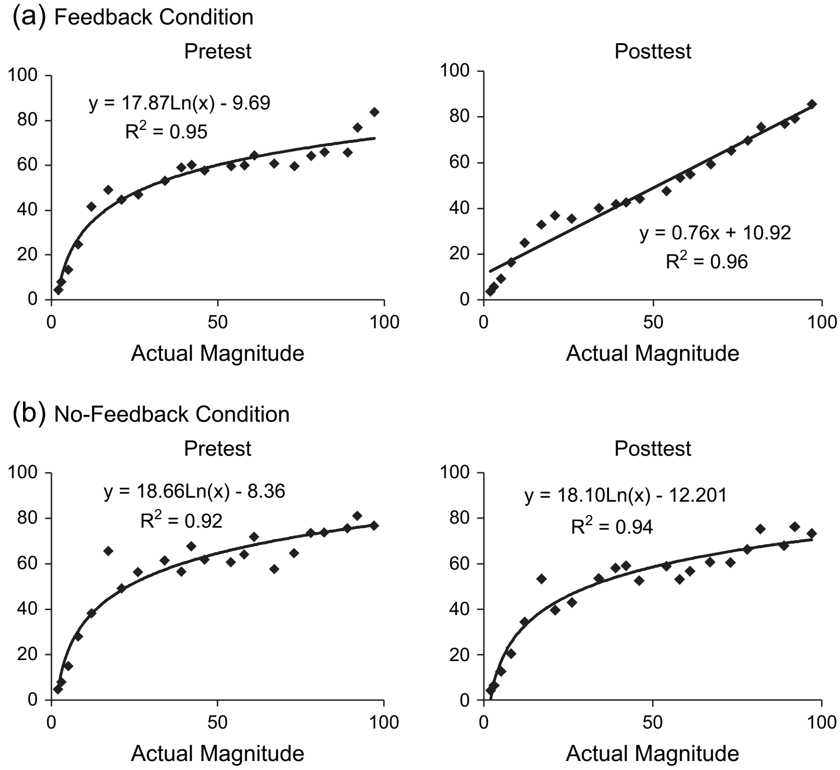

Figure 4. Best fitting function relating the number that was presented to the median number line estimate for that number generated by children in the feedback condition (top) and the no-feedback condition (bottom) on the pretest and posttest of Experiment 2.

magnitude comparisons regardless of whether children received feedback. Children who received feedback on their number categorizations answered an average of 87% of magnitude comparison problems correctly on the pretest and 88% on the posttest. Children who did not receive feedback answered an average of 84% of magnitude comparison problems correctly on the pretest and 85% on the posttest. Analysis of individual children's response times on magnitude comparison problems answered correctly also revealed no differences between feedback and no-feedback conditions on either the pretest or the posttest.

Path of Learning

Path analysis was used to examine in a single framework the relations among pretest scores on the three tasks, posttest scores on the three tasks, and provision of feedback in the learning phase. The final path model that was used to estimate the relations among variables, along with the path coefficients estimated from the raw data, is shown in Figure 5. We arrived at this path model by using TETRAD IV to search among all alternative path models consistent with our background and theoretical knowledge (for a description of the search algorithms used, see Spirtes, Glymour, & Scheines, 2000). The knowledge used to constrain the model search was (a) pretest performance and feedback condition could not cause each other, (b) posttest performance on any measure could not cause pretest performance on any measure, and (c) for each task, the pretest score was a direct cause of the posttest score. We did not bias the model specification further and thus allowed the data to inform us about the causes of posttest performance.

The best fitting model found by TETRAD had a p value of .49, which indicated that it fit the data quite well (in path models, the p value is a goodness-of-fit measure that varies between 0 and 1 in a way similar to R^2 (Bollen, 1989). All paths in Figure 5 were statistically significant at the $p = .05$ level, except for the path between pretest and posttest linearity of number line estimation.

The path model contains three important pieces of information. First, controlling for pretest performance and other factors, children who received feedback on their categorizations during training increased the linearity of their posttest categorizations by an average R^2 of .11 more than children who did not receive such feedback (bottom left path).

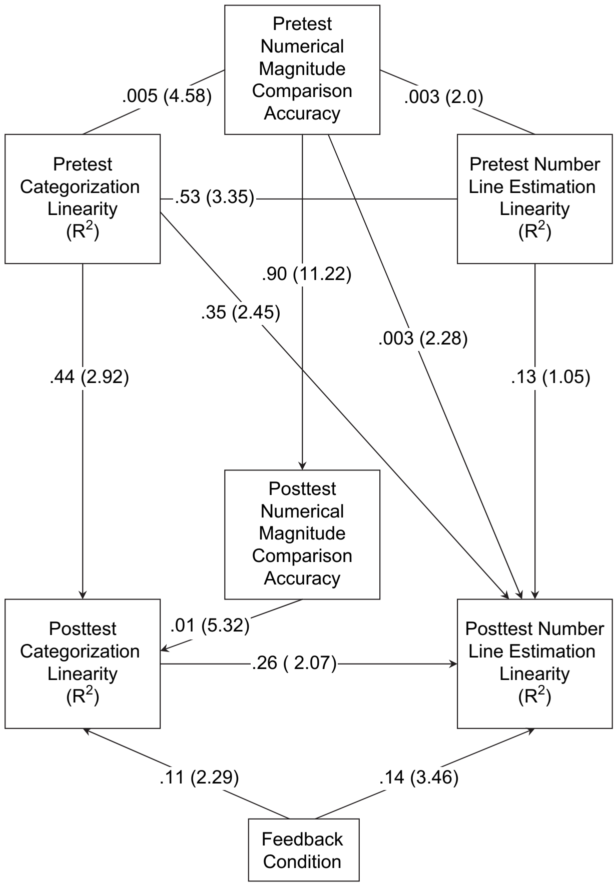

Figure 5. Path analysis of relations among feedback condition, pretest performance, and posttest performance on the three tasks in Experiment 2. Path coefficients (with *t* values in parentheses) are indicated on each path.

Second, there were two paths from feedback condition to the linearity of number line estimates at posttest. The direct path (feedback condition → posttest estimation linearity) indicates that children who received feedback on their categorizations increased the linearity of their posttest estimations by an average R^2 of .14 more than children who did not receive feedback (path at bottom right). The indirect path (feedback condition → posttest categorization linearity → posttest estimation linearity) indicates that feedback on categorizations during training increased the R^2_{lin} for posttest estimation by an average of an additional .03 through this route (i.e., .11 × .26). Thus, the overall effect of feedback during training was to increase the R^2_{lin} for posttest estimation by an average of .17, a gain of 29% (calculated as .17/the mean R^2_{lin} of the posttest estimates of children in the no-feedback group). Third, posttest accuracy on numerical magnitude comparison problems was only influenced by pretest accuracy on those problems; it was not influenced by feedback condition.

Thus, results from the path analysis confirmed the finding from other analyses that categorization feedback during the learning phase influenced posttest categorization and posttest number line estimation. Results of the path analysis also confirmed the finding from other analyses that categorization feedback did not influence posttest magnitude comparisons. Finally, the path analysis added the unique information that the influence of categorization feedback on number line estimation was realized through both a direct and an indirect route.

General Discussion

The results of this study indicated that the developmental transition from logarithmic to linear representations of numerical magnitude extends beyond the estimation tasks on which it had previously been documented; that consistent individual differences are present across estimation, categorization, and magnitude comparison; and that providing experiences that increase the linearity of children's categorizations of numbers also leads to greater linearity on other numerical tasks. In this concluding section, we examine implications of these and other findings for understanding of development, individual differences, and instruction.

Developmental Changes

Previous research indicated that between kindergarten and second grade, children progress from logarithmic to linear representations of numerical magnitude on 0 – 100 number line estimation problems (Siegler & Booth, 2004). Previous research also indicated that between second and fourth grade, children progress from logarithmic to linear representations of numerical magnitudes on three different 0 – 1000 estimation tasks: number line, measurement, and numerosity (Booth & Siegler, 2006; Siegler & Opfer, 2003). The present study replicated Siegler and Booth's (2004) findings on 0 – 100 number line estimation and demonstrated that the logarithmic to linear transition extends to at least one task other than estimation, numerical categorization.

Evidence of this transition on the categorization task was of special interest because it ruled out an alternative interpretation of all of the prior findings regarding the progression. On all three of the previously studied estimation tasks, linearity was necessary for high levels of accuracy. Thus, it was possible to view the transition not as one from a logarithmic representation to a linear one but rather as one from an incorrect to a correct one (with the incorrect answers happening to fall into a logarithmic pattern). In contrast, on the categorization task, there was no objectively correct answer. Nonetheless, both group and individual data showed the logarithmic to linear progression that had been observed earlier on the estimation tasks where there were objectively correct answers.

Whether a similar logarithmic to linear transition occurred on the numerical magnitude comparison task was less certain. Arguing against such a transition was the lack of a single, straightforward measure that directly pointed to a linear representation being used on the task. In contrast, this may have been more a problem of identifying an appropriate measure than of children not showing the same transition. As indicated by the compatibility effects in both Experiments 1 and 2, numerical magnitude comparison performance is influenced by numerous factors other than the representation of numerical magnitudes and appears to be a less direct reflection of the logarithmic to linear transition than the other two tasks.

Nonetheless, several findings suggested that the logarithmic to linear transition influenced performance on the magnitude comparison task. One type of evidence consistent with this conclusion was that the fit of the logarithmic function to children's accuracy and solution times was quite strong in kindergarten but decreased substantially by second grade. A second type of evidence arguing in the same direction was the consistently strong relation between the individual children's linearity on the other two tasks and the accuracy of their magnitude comparisons. Over the two experiments, the eight within grade correlations between each child's magnitude comparison accuracy and the child's linearity on the number line and categorization tasks ranged from $r = .56$ to $r = .83$. A third type of evidence for a logarithmic to linear progression influencing performance on the magnitude comparison task was that linearity of categorization predicted speed and accuracy of magnitude comparison, above and beyond the prediction of distance and problem size. A fourth type of evidence was the decreasing influence of problem size on both errors and solution times on magnitude comparisons that occurred between kindergarten and second grade; problem size is related to discriminability of the numbers being compared in logarithmic representations but not in linear ones. Thus, it seems likely that the same logarithmic to linear transition that was clearly evident on the number line estimation and categorization tasks influenced magnitude comparisons as well.

The present findings also had a more general implication for understanding development, one that can be summarized as, "Development repeats itself." Classic theorists such as Piaget and Inhelder (1956), Vygotsky (1934/1962), and Werner (1957) proposed that developmental changes at different ages show extensive parallels. Studies of number line estimation have provided support for this principle. Booth and Siegler (2006) observed the same types of changes between second and fourth grade in the 0–1,000 range as Siegler and Booth (2004) had observed between kindergarten and second grade in the 0–100 range. Similar parallels have been observed between learning of single-digit addition from kindergarten to second grade and learning of single-digit multiplication between third and fifth grade (Geary, 1994, 2006; Siegler, 1987); between development of crawling and subsequent development of walking (Adolph, 1997); and between early development of grammar and later development of story telling and map making (Karmiloff-Smith, 1992).

The present study adds development of numerical categorization to this set of examples of development repeating itself. Numerous phenomena that were evident in previous studies of development of categorization of the numbers 1–9 during the preschool period also were evident in the present study of development of categorization of the numbers 1–100 during the early elementary school period. In both cases, children at first differentiate clearly among numbers at the low end of the scale but not among numbers at the high end of the scale; later, they differentiate clearly among numbers throughout the scale. In both cases, children group numbers into categories such as "small numbers" and "big numbers." In both cases, the linearity of the categorizations is strongly related to accuracy of magnitude comparison, even after the influence of problem size and distance is statistically controlled. Moreover, in both cases, providing experiences that improve categorizations of the numbers improves performance on other numerical tasks. It seems likely that searching for additional examples of development repeating itself will lead to identification of many more such parallels.

Individual Differences

Prior research indicated the presence of consistent individual differences in use of the linear representation in second and fourth graders' performance on three estimation tasks (Booth &

Siegler, 2006). The present research indicated that this consistency is not limited to estimation or to children in second and fourth grades. Within-grade as well as between-grade correlations among kindergartners', first graders', and second graders' number line estimation, numerical categorization, and numerical magnitude comparison performance were as strong as the previously obtained correlations among the estimation tasks. To be specific, the within-grade correlations of the linearity of estimation on the three tasks in Booth and Siegler (2006) ranged from r = .54 to r = .65 for second graders and from r = .60 to r = .84 for fourth graders. The corresponding within-grade correlations in the present experiment among linearity of number line estimation, linearity of categorization, and accuracy of magnitude comparison ranged from r = .57 to r = .82 for kindergartners, r = .56 to r = .82 for first graders, and r = .80 to r = .83 for second graders. Thus, the consistency of individual differences on these three tasks, only one of which involved estimation, was at least as strong as the consistency of individual differences on the three estimation tasks.

The strength of these relations among experimental tasks helps explain the quite strong relations between linearity of numerical representations and overall math achievement test scores that are present from kindergarten through at least fourth grade (Booth & Siegler, 2006; Siegler & Booth, 2004). These relations of performance on the experimental tasks to math achievement test scores are not as strong as the relations among the experimental tasks but they are still substantial. Across four experiments that included kindergartners and first, second, third, and fourth graders, the correlations have averaged around r = .50 (Booth & Siegler, 2006; Siegler & Booth, 2004). The present findings indicated that linearity of numerical magnitude representations influence a wide range of tasks, not just ones involving estimation. It seems likely that the tasks that are influenced by the linearity of numerical magnitude representations include ones that are sampled on achievement tests, thus producing the relation. Further research, however, is needed to determine the direction of the relation and to determine the influence of other factors that might contribute to both math achievement and linearity of magnitude representations.

Individual differences in the linearity of numerical representations clearly are not the only contributor to individual differences in math achievement test scores. Variation among children in working memory functioning (Geary, Hoard, Nugent, & Byrd-Craven, 2007; Hecht, Close, & Santisi, 2003), arithmetic fact retrieval (Geary & Hoard, 2005; Jordan et al. 2003), strategy execution and counting (Jordan, 2007), and other capabilities are also influential. Still, both the strength of the relation of linearity of numerical representations to overall math achievement test scores and the variety of measures of linearity that show the relation are striking.

One interpretation of these findings is that linearity of numerical representations provides an operational definition of the concept of number sense. As mentioned earlier, number sense is an ill-defined concept that nonetheless is viewed by both educators (NCTM, 2000) and researchers (Jordan, 2007; Sowder, 1992) as central to individual differences in numerical proficiency. One interpretation of number sense is that it is the ability to discriminate among numerical magnitudes throughout a range of numbers and use the discriminations to constrain and judge the plausibility of outcomes of mathematical operations. Within this interpretation, use of linear representations of numbers is essential to number sense because it allows differentiation among numerical magnitudes throughout the range. In contrast, reliance on logarithmic representations results in numbers at the high end of the range being lumped together as "all those big numbers." Another implication of this perspective is that people may have good number sense within one range of numbers but not within other ranges. For example, they might have good number sense for values in the hundreds but not for values in the millions or billions.

Consistent with this view, the improvements with age in Experiment 1 and with categorization experience in Experiment 2 reflected in large part an increased differentiation at the high end of the numerical range (Figures 1–4). Even preschoolers from middle-income backgrounds seem to have a good sense of numerical magnitudes in the 0–10 range (Siegler & Ramani, 2006). In contrast, judging from the constant confusion in both print and electronic media of millions, billions, and trillions, it seems likely that few people of any age have good number sense in those ranges. The challenge is to help more people gain a good sense of numerical magnitudes across a wider range of numbers. Providing experiences that encourage use of linear representations over wider numerical ranges seems a promising means for pursuing this goal.

Instructional Implications

Results from the present experiments provided both correlational and causal evidence for the general view that "divide and conquer" is an effective means of promoting learning. It also provided evidence for the more specific view that one instance of the divide-and-conquer approach—subjective categorization of numbers—provides a useful means for learning about unfamiliar numerical ranges. In Experiment 1, first and second graders' subjective categorizations of numbers predicted the accuracy of their magnitude comparisons involving those numbers above and beyond distance and problem-size effects. In Experiment 2, explicit feedback on categorization that was intended to promote a more effective "divide-and-conquer" approach led to more linear subjective categorizations of numbers and also to improved number line estimation.

The likely reason why divide and conquer is such a pervasive approach is that it reduces the cognitive demands of complex tasks and thus makes good performance possible well before perfect performance can be attained. Put another way, it provides an entrance into complex content domains that beginning learners might otherwise find overwhelming. For example, in the present set of magnitude comparison problems, dividing the 0 – 100 range into the five equal size categories that were presented, and invariably applying the simple categorization model described previously, would yield 94% correct answers. Similarly, using the categorizations to constrain number line estimates, for example, by mapping the categories onto corresponding parts of the number line and realizing that a very big number must be placed further toward the high end than a big number, would promote linear and accurate estimation.

Another way of thinking about why categorization helps children learn about numerical magnitudes is that it promotes simultaneous attention to absolute and relative qualities of numbers. For example, accurate number line estimation requires consideration of the relation between the magnitude of a particular number and the magnitudes of other numbers, as well as attention to the absolute quantity represented by the number. Young children tend to focus on one or the other but not both. For example, when told that they would be categorizing numbers between 0 and 20, 0 and 50, or 0 and 100, kindergartners categorized 18 as a big number in all three contexts. In contrast, second graders categorized 18 as a big number in the 0 – 20 context but as a small number in the 0 – 100 context (Siegler & Laski, 2006). As this example illustrates, learning numerical relations may be particularly difficult because the relations change with context; a big number in one context is a small number in another. Categorization has been found to help children consider relations as well as absolute properties in domains other than number (Gentner & Markman, 1997; Namy & Gentner, 2002) and may well do the same with numbers.

An instructional implication of these findings is that it may often be preferable to convey approximately correct representations and strategies rather than trying to teach optimal ones directly. In situations where young children do not encounter instruction, they often generate partially correct understandings before they generate entirely correct ones. Children who have generated these partial understandings are far more likely to benefit from feedback and other forms of instruction than children who lack such partial understandings. For example, children who know that weight influences the motion of balance scales, but who do not know that distance is also influential, are more likely to benefit from feedback about the role of distance than are children who do not understand the role of weight (Siegler, 1976, 1981; Siegler & Chen, 1998). Similarly, children whose gestures already express attention to relevant variables on mathematical equality and conservation problems are more likely to learn to solve the problems correctly than children who do not show such partial understanding in gesture (Alibali, 1999; Church & Goldin-Meadow, 1986; Goldin-Meadow, Alibali, & Church, 1993). Moreover, when children are taught a correct solution strategy, those who do not show partial understanding through gesture often generalize more narrowly than do those who show the partial understanding (Goldin-Meadow, 2001). Thus, promoting thinking that is in the right direction may often allow children to divide and eventually conquer the problems.

References

Adolph, K. E. (1997). Learning in the development of infant locomotion. *Monographs of the Society for Research in Child Development*, *62* (Serial No. 251).

Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology*, *35*, 127 – 145.

Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. *Psychology of Learning and Motivation*, *11*, 101 – 159.

Berch, D. B., Foley, E. J., Hill, R. J., & Ryan, P. M. (1999). Extracting parity and magnitude from Arabic numerals: Developmental changes in number processing and mental representation. *Journal of Experimental Child Psychology*, *74*, 286 – 308.

Bollen, K. A. (1989). *Structural equations with latent variables*. Oxford, UK: John Wiley & Sons.

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

Braine, M. D. S. (1971). The acquisition of language in infant and child. In C. E. Reed (Ed.), *The learning of language* (pp. 7 – 87). New York: Appleton-Century-Crofts.

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

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*, 43 – 71.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*, 371 – 396.

Dowker, A. (1992). Computational estimation strategies of professional mathematicians. *Journal for Research in Mathematics Education*, *23*, 45 – 55.

Dowker, A. (2003). Young children's estimates for addition: The zone of partial knowledge and understanding. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 243 – 265). Mahwah, NJ: Erlbaum.

Dowker, A., Flood, A., Griffiths, H., Harriss, L., & Hook, L. (1996). Estimation strategies of four groups. *Mathematical Cognition*, *2*, 113 – 135.

Duncan, E. M., & McFarland, C. E. (1980). Isolating the effect of symbolic distance and semantic congruity in comparative judgments: An additive-factors analysis. *Memory and Cognition*, *8*, 612 – 622.

Ehri, L. C., & Wilce, L. S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly*, *20*, 163 – 179.

Fechner, G. T. (1987/1869). My own viewpoint on mental measurement (1869). *Psychological Research/Psychologische Forschung*, *49*, 213 – 219.

Geary, D. C. (1994). *Children's mathematical development: Research and practical implications*. Washington, DC: American Psychological Association.

Geary, D. C. (2006). Development of mathematical understanding. In W. Damon & R. M. Lerner (Series Eds.) & D. Kuhn and R. S. Siegler (Vol. Eds.), *Handbook of child psychology. Vol. 2. Cognition, perception, and language* (6th ed., pp. 777 – 810). Hoboken, NJ: Wiley.

Geary, D. C., & Hoard, M. K. (2005). Learning disabilities in arithmetic and mathematics: Theoretical and empirical perspectives. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 253 – 267). New York: Psychology Press.

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2007). Strategy use, long-term memory, and working memory capacity. In D. B. Berch & M. M. M. Mazzocco (Eds.), *Mathematical learning disabilities: Research, theory, and practice* (pp. 83 – 105). Baltimore: Paul H. Brookes Publishing.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*, 45 – 56.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, *38*, 293 – 304.

Goldin-Meadow, S. (2001). Giving the mind a hand: The role of gesture in cognitive change. In J. L. McClelland & R. S. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives* (pp. 5 – 31). Mahwah, NJ: Erlbaum.

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, *100*, 279 – 297.

Griffin, S. A., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 25 – 49). Cambridge, MA: MIT Press.

Hecht, S. A., Close, L., & Santisi, M. (2003). Sources of individual differences in fraction skills. *Journal of Experimental Child Psychology*, *86*, 277 – 302.

Jordan, N. (2007). Do words count? Connections between mathematics and reading difficulties. In D. B. Berch & M. M. M. Mazzocco (Eds.), *Mathematical learning disabilities: Research, theory, and practice* (pp. 107 – 120). Baltimore: Paul H. Brookes Publishing.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development*, *74*, 834 – 850.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.

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

LeFevre, J., Greenham, S. L., & Waheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. *Cognition and Instruction*, *11*, 95 – 132.

Lemaire, P., & Lecacheur, M. (2002). Children's strategies in computational estimation. *Journal of Experimental Child Psychology*, *82*, 281 – 304.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Quantitative development in infancy and early childhood*. New York: Oxford Press.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519 – 1520.

Murray, P. L., & Mayer, R. E. (1988). Preschool children's judgments of number magnitude. *Journal of Educational Psychology*, *80*, 206 – 209.

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, *131*, 5 – 15.

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

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics: Higher standards for our students . . . Higher standards for ourselves*. Washington, DC: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (2006). Curriculum focal points for prekindergarten through grade 8 mathematics. Washington, DC: National Council of Teachers of Mathematics, Retrieved October 3, 2006, from http://www.nctm.org/focalpoints/downloads.asp.

Nuerk, H. -C., Kaufmann, L., Zoppoth, S., & Willmes, K. (2004). On the development of the mental number line: More, less, or never holistic with increasing age? *Developmental Psychology*, *40*, 1199 – 1211.

Nuerk, H. -C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting tens and units back into different bins. *Cognition*, *82*, B25 – B33.

Opfer, J., & Siegler, R. S. (in press). Representational change and children's numerical estimation. *Cognitive Psychology*.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London: Routledge and Kegan Paul.

Rakison, D. H., & Oakes, L. M. (2003). *Early category and concept development: Making sense of the blooming, buzzing, confusion*. New York: Oxford University Press.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development*, *48*, 630 – 633.

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1973). The internal representation of numbers. *Cognitive Psychology*, *6*, 82 – 138.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, *8*, 481 – 520.

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Society for Research in Child Development Monographs*, *46* (Whole No. 189).

Siegler, R. S. (1987). Some general conclusions about children's strategy choice procedures. The neo-Piagetian theories of cognitive development: Toward an integration [Special issue]. *International Journal of Psychology*, *22*, 729 – 749.

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

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology*, *36*, 273 – 310.

Siegler, R. S., & Laski, E. V. (2006). Young children's logarithmic representations of numerical magnitude: What's the source? Manuscript in preparation.

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

Siegler, R. S., & Ramani, G. B. (2006). Early development of estimation skills. *APS Observer*, *19*, 34/44.

Siegler, R. S., & Robinson, M. (1982). The development of numerical understandings. *Advances in Child Development and Behavior*, *16*, 241 – 312.

Sowder, J. T. (1992). Estimation and number sense. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 371 – 389). New York: Macmillan.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge, MA: The MIT Press.

Vygotsky, C. S. (1934/1962). *Thought and language*. New York: Wiley.

Werner, H. (1957). The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development: An issue in the study of human behavior* (pp. 125 – 148). Minneapolis: University of Minnesota Press.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*, 220 – 251.