Developmental and Individual Differences in Pure Numerical Estimation

Julie L. Booth and Robert S. Siegler
Carnegie Mellon University

The authors examined developmental and individual differences in pure numerical estimation, the type of estimation that depends solely on knowledge of numbers. Children between kindergarten and 4th grade were asked to solve 4 types of numerical estimation problems: computational, numerosity, measurement, and number line. In Experiment 1, kindergartners and 1st, 2nd, and 3rd graders were presented problems involving the numbers 0–100; in Experiment 2, 2nd and 4th graders were presented problems involving the numbers 0–1,000. Parallel developmental trends, involving increasing reliance on linear representations of numbers and decreasing reliance on logarithmic ones, emerged across different types of estimation. Consistent individual differences across tasks were also apparent, and all types of estimation skill were positively related to math achievement test scores. Implications for understanding of mathematics learning in general are discussed.

Keywords: individual differences, mathematics, estimation, development

Estimation is a pervasive activity in the lives of both children and adults. Consider a few representative estimation problems: About how long will it take for you to finish your homework? About how many people were at the concert? About how much will each teammate have to pay to buy a $50 present for the coach? Estimation is used to solve these and many other problems because accurate estimates are sufficient for many purposes and because people often lack the knowledge, time, means, or motivation needed to calculate precise values.

Despite the omnipresence and importance of estimation in children's lives, far less is known about the process than about other basic numerical processes, such as counting and arithmetic. A large part of the reason for this limited knowledge about estimation is the diversity of tasks that fall under the heading. Estimating the population of Russia, the product of 175×243, and the speed of a passing car have little in common except for the answer being approximate. This diversity means that tasks that fall under the *estimation* heading will have numerous sources of difficulty and numerous patterns of development. The diversity also means that progress in understanding estimation almost certainly will require well-chosen differentiations among major estimation processes.

One approach to distinguishing among such processes is by classifying estimation tasks according to the knowledge that they demand and then considering the processes needed to estimate accurately on tasks that require particular types of knowledge. Two types of knowledge that seem useful for distinguishing among estimation tasks are knowledge of real-world content and of numbers. Some estimation tasks require knowledge of specific real-world entities, conventional measurement units, or both; an exam-

ple that requires both is estimating the number of miles between London and Paris. Other estimation tasks do not require such real-world knowledge, for example, estimating the number of dots on a page. Similarly, some estimation tasks involve numerical input, output, or both (e.g., mental multiplication requires both). Other tasks do not involve either (e.g., estimating whether there is enough time to cross a street before oncoming traffic arrives).

The process that is of interest in the present study is *pure numerical estimation*. This process can be defined in terms of its goal and the two distinctions made in the previous paragraph: Pure numerical estimation is a process that has a goal of approximating some quantitative value; that uses numbers as inputs, outputs, or both; and that does not require real-world knowledge of the entities for which properties are being estimated or of conventional measurement units. Three examples of pure numerical estimation are approximating the product of 395×112, the location of 26 on a number line, and the number of marbles in a jar. The process seems especially central to estimation because it eliminates non-mathematical knowledge of specific entities and particular measurement units as sources of variability in performance, and because recent research on a pure numerical estimation task revealed an interesting developmental shift in understanding of numerical magnitudes that may influence all types of numerical estimation.

The recently discovered developmental shift involves a change from reliance on logarithmic representations of numerical magnitudes to reliance on linear representations of them. This shift has been found between kindergarten and second grade for estimates of numerical locations on 0–100 number lines (Siegler & Booth, 2004) and between second and sixth grade for estimates of numerical locations on 0–1,000 number lines (Siegler & Opfer, 2003). To be specific, when asked to estimate the locations of numbers on number lines with 0 at one end and 100 at the other, the large majority of kindergartners produced estimates consistent with a logarithmic function, whereas the large majority of second graders produced estimates consistent with a linear function (about half of first graders' estimates were best fit by one function and half by the other) (Siegler & Booth, 2004). Similarly, when asked to

estimate the locations of numbers on number lines with 0 at one end and 1,000 at the other, the large majority of second graders generated logarithmic distributions of estimates, the large majority of sixth graders produced linear distributions, and about half of fourth graders were best fit by each function (Siegler & Opfer, 2003). By second grade, if not earlier, children possess both types of representations, though they apply them in different contexts; thus, almost half of the second graders in Siegler and Opfer's study generated a linearly increasing pattern of estimates on 0–100 number lines and a logarithmically increasing pattern on 0–1,000 number lines. The fact that the shift takes place at different ages for different ranges of numbers makes it likely that the shift is driven by increasing exposure to numbers in each range, rather than to a more general insight into the decimal system or to some maturational change that takes place at a particular age.

The shift from logarithmic to linear representations is important for both theoretical and empirical reasons. At a theoretical level, linear representations reflect the structure of the number system; appropriately representing that structure is fundamental to understanding arithmetic, decimal notation, algebra, and other aspects of mathematics. At an empirical level, the degree to which children generate linear representations is related rather strongly to kindergartners' and first and second graders' math achievement test scores (Siegler & Booth, 2004). In this article, we examine another potential source of importance—the generality of the shift across different types of estimation.

Current Understanding of Developmental and Individual Differences in Estimation

In a recent review of the estimation literature, Dowker (2003) found little evidence for consistent developmental or individual differences across different types of estimation. The studies she reviewed indicated that some types of estimation show substantial improvement with age and experience, whereas other types remain unchanged over long periods. However, the tasks used in the studies varied so widely in the real-world and numerical knowledge required for accurate estimation that any conclusions about either developmental or individual differences were, by necessity, extremely tenuous. The types of knowledge needed to accurately estimate the distance between London and Paris, and the types of instruction that would produce such knowledge, are unrelated to the types of knowledge and instruction needed to accurately estimate the answer to 18×23. Therefore, one purpose of the present research was to determine whether the developmental trend that has been found on number line estimation—from reliance on logarithmic representations to reliance on linear ones—also occurs on other types of pure numerical estimation tasks and, if so, whether the changes occur during the same period of development.

Existing research on individual differences in estimation is similarly inconclusive. As noted above, the few studies that have examined multiple types of estimation have not revealed consistent individual differences across estimation tasks (Dowker, 1998; Hook, 1992; Paull, 1972). However, previous studies have shown that more skillful estimators tend to have better conceptual understanding of mathematics (LeFevre, Greenham, & Naheed, 1993; Petitto, 1990), better counting and arithmetic skills (LeFevre et al., 1993; Newman & Berger, 1984), greater working memory capacity (Case & Sowder, 1990), and higher math achievement test

scores (Siegler & Booth, 2004) than do children who estimate less accurately. Thus, the second main purpose of the present research was to examine the consistency of individual differences in diverse types of pure numerical estimation.

The Present Study

We performed two experiments. In Experiment 1, we presented four pure numerical estimation tasks to kindergartners and first, second, and third graders, using different assessment methods for each task. Children were required to approximate (a) the answers to addition problems (computational estimation), (b) the number of candies in a jar (numerosity estimation), (c) the length of a line in inches (measurement estimation), and (d) the location of a number on a line with numerical anchors at each end (number line estimation). The first three tasks were of interest because they correspond to types of pure numerical estimation that are frequently encountered in the everyday environment; the fourth task (number line estimation) was of interest because of the previously described findings regarding the logarithmic-to-linear shift.

One goal of Experiment 1 was to replicate Siegler and Booth's (2004) findings regarding young elementary schoolchildren's number line estimation, in particular the log-to-linear shift on the 0–100 scale and the relation between linearity of estimates and math achievement test scores. Such replication seemed important because the finding that the large majority of 5- and 6-year-olds represent numerical magnitudes logarithmically was quite counterintuitive, and the result had not been replicated previously. The other goal of Experiment 1 was to determine through less elaborate assessments whether other types of pure numerical estimation also show substantial development in this age range, whether individual differences in them correlate with differences in math achievement test performance, and thus whether more detailed investigation of them was warranted.

In Experiment 2, we assessed the same four types of estimation but this time with rigorous assessment techniques on all tasks. In particular, the tasks used in Experiment 2 were designed to be parallel on all nonessential features and to allow detailed examination of children's numerical representations. The main goals were to determine (a) whether the log-to-linear pattern of development was present on pure numerical estimation tasks other than number line estimation, (b) whether the timing of the log-to-linear shift was similar on the different types of estimation, (c) whether consistent individual differences were present across the four estimation tasks, and (d) whether any consistent individual differences patterns that arose among the estimation tasks were attributable to shared relations to math achievement test scores.

Experiment 1

Method

Participants

Ninety children participated: 20 kindergartners (mean age = 5.8 years, SD = 0.4), 25 first graders (mean age = 6.8 years, SD = 0.4), 23 second graders (mean age = 7.9 years, SD = 0.6), and 22 third graders (mean age = 9.1 years, SD = 1.3). Among the participants, 63% were Caucasian, 33% African American, 2% Asian American, and 1% Latino. The experimenter was a Caucasian, female graduate student.

The children were recruited from a public school and a parochial school in a lower to middle income neighborhood; between 60% and 75% of the participants at each grade level came from the public school. Of children who attended the public school, 55% were eligible for free or reduced-cost lunches; at the parochial school, 15% of attendees were eligible for free or reduced-cost lunches. Participation was completely voluntary, and no extrinsic rewards were provided.

Procedure and Materials

Children met one-on-one with the experimenter for a single 20-min session. The session consisted of two phases, with half of the children receiving each phase first. One phase involved a detailed assessment of number line estimation; the other phase involved exploratory assessments of the other three types of estimation.

The number line phase included an initial orienting problem followed by 26 experimental problems. On the orienting problem, children were presented a sheet of paper with a 25-cm line across the middle; the number 0 was printed just below the left end of the line and the number 100 just below the right end. Children were asked to mark where they thought 50 would go on the line. After they did so, they were shown an identical number line with 50 marked in the correct position, told that that was where 50 belonged, and asked if they knew why 50 went there. All children were then told, "Because 50 is half of 100, it goes directly in the middle, halfway between 0 and 100. So 50 is the middle, and it's the only number that goes exactly in the middle."

After the orienting problem, children were presented 26 sheets of paper, each with an identical 25-cm line, and asked to put a single mark on each line to indicate the location of a number. The lines were identical to the one used on the orienting problem except that a number (different on each trial) was printed above the middle of the line. To ensure that we would be able to discriminate between linear and logarithmic estimation patterns, we oversampled the numbers below 30 by including four numbers from each of the first three decades and two numbers from each successive decade. The 26 numbers that were presented were 3, 4, 6, 8, 12, 14, 17, 18, 21, 24, 25, 29, 33, 39, 42, 48, 52, 57, 61, 64, 72, 79, 81, 84, 90, and 96. The order of the sheets was randomized separately for each child.

The other phase of Experiment 1 included brief assessments of measurement, numerosity, and computational estimation. Children were randomly assigned to receive the three tasks in one of the six possible orders.

For the measurement estimation task, the experimenter first showed the children a sheet with a 1-in. (2.54-cm) line in the center of the page, told them that the line was 1 in. long, and asked them to trace the line with their finger and to pay attention to its length. Children were also told that the line would remain present to remind them of the length of an inch. The purpose of presenting the inch marker was to eliminate knowledge of the measurement unit as a source of variability in the children's estimates. After the inch marker was introduced, children were presented two measurement tasks, one requiring production of lengths and the other requiring judgment of lengths. On the production task, children were given blank sheets of paper, reminded of the presence and length of the inch marker, and asked to draw lines of 3, 5, 8, and 10 in. On the judgment task, children were shown four cards, each with a line printed across the center and two possible measures of its length in inches printed below; the task was to choose the measure that was closer to the correct length. The lines were 3, 6, 8, and 9 in., respectively. The two choices were the line's actual length and a number that was 4 or 5 in. from the correct length.

For numerosity estimation, children were shown containers with 22, 34, 46, or 58 Hershey's Kisses and asked to estimate which of two numbers more closely corresponded to the number of candies in the container; the two choices always included the correct answer and a foil that was one half, one and one half, or twice the correct answer. Children were randomly assigned to receive one of three possible sets of numerical choices. Set A involved the choices 11 versus 22 for the container with 22 candies, 34 versus 51 for the container with 34 candies, 46 versus 92 for the container with 46 candies, and 29 versus 58 for the container with 58 candies. Set B involved the choices 22 versus 33, 34 versus 68, 23 versus 46, and 58 versus 87. Set C involved the choices 22 versus 44, 17 versus 34, 46 versus 69, and 58 versus 116. Questions were of the form "Do you think this is 17 or 34 Kisses?"

For computational estimation skill assessment, children were shown 12 cards, each containing an addition problem and three possible answers to the problem. The problems included four 2-digit plus 1-digit items (e.g., 25 + 3), four 2-digit plus 2-digit items (e.g., 35 + 23), two 1-digit + 1-digit + 1-digit items (e.g., 7 + 2 + 9), and two 1-digit + 1-digit + 1-digit + 1-digit items (e.g., 5 + 4 + 2 + 7). All of the sums were between 0 and 100. None of the 2-digit plus 2-digit problems required carrying to obtain the exact number; two of the four 2-digit plus 1-digit problems did, as did all of the problems involving strings of 1-digit numbers. The alternative answers listed for each problem were always consecutive multiples of 10. On each item, the experimenter read the problem aloud and asked the child which of the three choices was closest to the answer (e.g., "Is 34 + 29 closest to 40, 50, or 60?").

There was no time limit for any of the problems. In addition, no feedback was given about specific estimates, though the experimenter frequently offered general praise. Following completion of all tasks, children were told they did a good job, thanked for participating, and returned to their classroom.

We obtained percentile scores from the math achievement tests taken by the children as part of their school requirements. Children in one school took the Stanford Achievement Test, Ninth Edition approximately 2 months prior to the study; children in the other school took the Iowa Test of Basic Skills approximately 4 months after completion of the study.

Results

Development of Number Line Estimation

To measure changes in estimation accuracy, we calculated each child's percent absolute error:

$$\left| \frac{\text{Estimate} - \text{Estimated Quantity}}{\text{Scale of Estimates}} \right|$$

For example, if a child was asked to estimate the location of 15 on a 0–100 number line and placed the mark at the location that corresponded to 35, the percent absolute error would be 20%: $[(35 - 15)/100]$.

An analysis of variance (ANOVA) on each child's mean percent absolute error indicated that accuracy increased with grade, $F(3, 86) = 40.12$, $p < .01$, $\eta^2 = .58$. Kindergartners' mean percent absolute error (24%) was considerably greater than those of first, second, and third graders (12%, 10%, and 9%, respectively). This level of accuracy was closely comparable to that obtained by Siegler and Booth (2004) for the three age groups examined in both studies: kindergartners (percent absolute error = 24% in both studies), first graders (14% vs. 12%), and second graders (10% in both studies).

The next analyses examined whether children showed the hypothesized age-related change from a logarithmic to a linear representation on the number line task. First, the median estimate for each number generated by children in each grade was calculated; then, the differences between that number and the number predicted by the best-fitting logarithmic, linear, and exponential functions were compared. The exponential function fit less well than the other two functions at all grade levels, so it was not included in further analyses of the group data.

As shown in Figure 1, kindergartners' number line estimates were better fit by the logarithmic function ($R^2 = .92$) than by the linear function ($R^2 = .63$), $t(25) = 4.64$, $p < .01$, $d = 0.89$. In contrast, first graders' estimates were better fit by the linear function than by the logarithmic one ($R^2 = .96$ vs. .89), $t(25) = 2.50$, $p < .05$, $d = 0.66$, as were those of second graders ($R^2 = .97$ vs. .88), $t(25) = 4.38$, $p < .01$, $d = 1.12$, and third graders ($R^2 = .98$ vs. .85), $t(25) = 5.66$, $p < .01$, $d = 1.55$.

To ensure that these results were not due to averaging across children, we compared the variance accounted for by the best-fitting linear, logarithmic, and exponential functions for individual children's estimates. Because the number of children who were best fit by the exponential function was minimal (2 children), only children who were best fit by the linear or the logarithmic function were examined further. The type of function that fit the most

children varied with age, $\chi^2(3, N = 90) = 25.61$, $p < .01$, $V = .38$. The logarithmic function provided the best fit for 80% of kindergartners and the linear function for 15%. In contrast, the linear function provided the best fit for the majority of second and third graders (74% and 86%, respectively), whereas the logarithmic function provided the best fit for only 26% of second graders and 14% of third graders. First graders were equally likely to be best fit by the linear and logarithmic functions (56% and 40% of children, respectively). Paired-sample t tests indicated that the fit of the linear (lin) function to individual children's estimates was worse than the fit of the logarithmic (log) function for kindergartners (mean $R^2_{lin} = .36$, $SE = .06$, vs. mean $R^2_{log} = .49$, $SE = .05$), $t(19) = 4.17$, $p < .01$, $d = 0.51$; equal to the fit of the logarithmic function for first graders (mean $R^2_{lin} = .77$, $SE = .04$, vs. mean $R^2_{log} = .75$, $SE = .03$), $t(24) = 0.92$, ns, $d = 0.12$; better than the
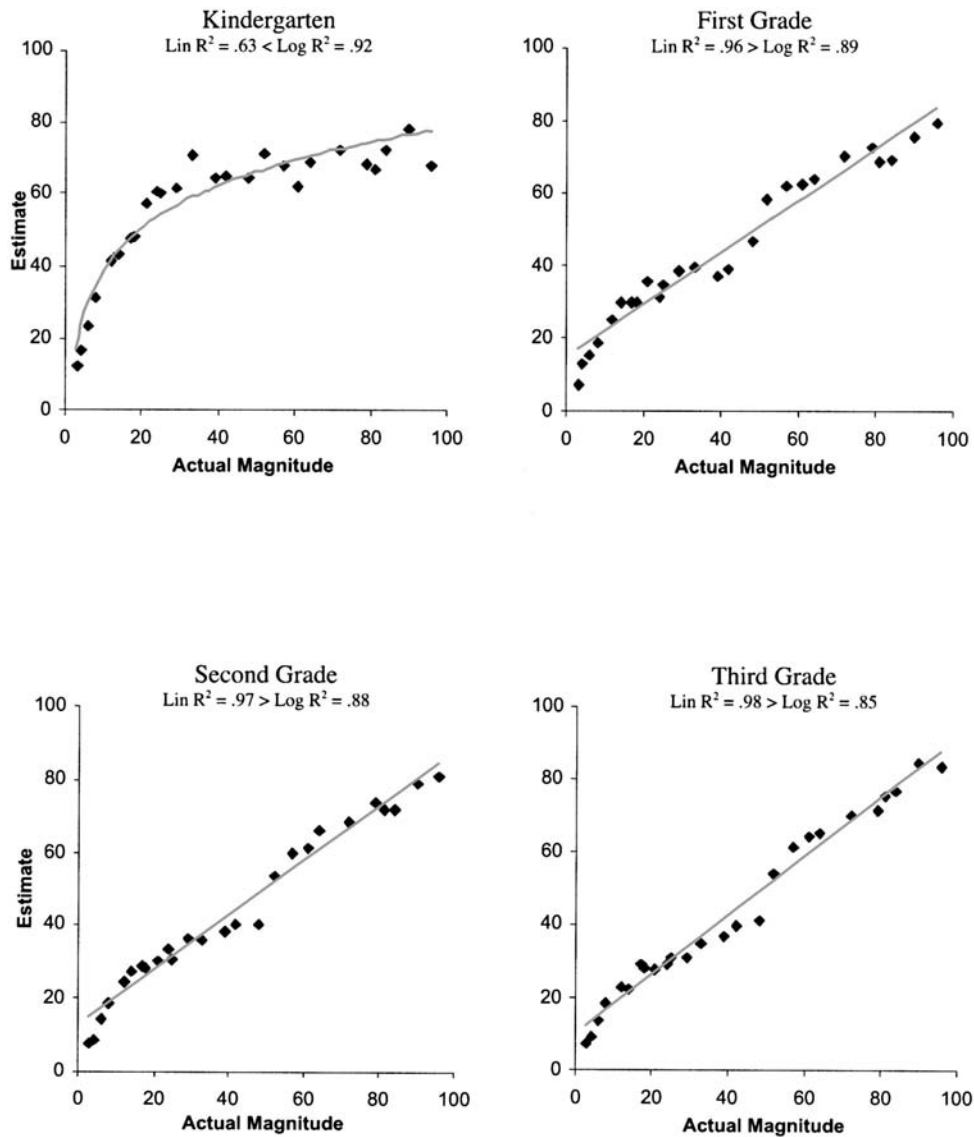

Figure 1. Experiment 1: Best-fitting equations for median number line estimates for kindergartners and first, second, and third graders. Lin = linear; Log = logarithmic.

fit of the logarithmic function for second graders (mean R^2_{lin} = .87, SE = .02, vs. mean R^2_{log} = .80, SE = .02), $t(22)$ = 3.62, $p <$ .01, d = 0.73; and better than the fit of the logarithmic function for third graders (mean R^2_{lin} = .87, SE = .03, vs. mean R^2_{log} = .78, SE = .02), $t(21)$ = 3.81, $p <$ .01, d = 0.81. A one-way ANOVA indicated that the fit of the linear function to individual children's estimates increased with grade, $F(3, 86)$ = 37.33, $p <$ .01, η^2 = .57, with the fit of the linear function being better for first graders than for kindergartners and better for second and third graders than for first graders.

Development of Measurement, Numerosity, and Computational Estimation

We next conducted one-way ANOVAs on changes with age and grade on the other three estimation tasks. For measurement estimation, performance was evaluated separately for the two tasks. Skill at the length production task was measured by the percent absolute error on the line drawings. Such error decreased with grade, $F(3, 86)$ = 3.77, $p <$ .05, η^2 = .12. The percent absolute errors for the lines of kindergartners (43%, SE = .05), first graders (37%, SE = .04), and second graders (33%, SE = .04) were greater than those of third graders (23%, SE = .03). Proficiency at the length judgment task, as measured by the percentage of problems on which the students chose the better estimate, did not change with age.

Performance on the numerosity and computational estimation tasks was measured by the percentage of problems on which the students selected the more accurate estimate. For numerosity estimation, kindergartners were less accurate (53% correct, SE = .06) than were first graders (68%, SE = .03), second graders (76%, SE = .04), or third graders (73%, SE = .05), $F(3, 86)$ = 5.23, $p <$ .01, η^2 = .15 (chance was 50% for this task). For computational estimation, kindergartners were less accurate than first graders (36%, SE = .03, vs. 52%, SE = .04, correct), and both were less accurate than second and third graders (75%, SE = .03, and 79%, SE = .03, correct, respectively), $F(3, 86)$ = 31.76, $p <$ .01, η^2 = .53 (chance was 33% correct).

Estimation and Math Achievement

To examine relations between estimation performance and math achievement, we correlated children's estimation scores with their national percentile rankings on the math section of their achievement tests. Two third-grade students did not take the achievement tests and therefore were not included in this analysis.

Children's R^2_{lin} on the number line estimation task correlated positively with their math achievement test scores at all four grade levels: kindergarten, $r(18)$ = .51, $p <$ .05; first grade, $r(23)$ = .44, $p <$ .05; second grade, $r(21)$ = .54, $p <$ .01; and third grade, $r(18)$ = .45, $p <$ .05. Within each grade, children who generated more linear patterns of estimates also had higher achievement test scores. However, no significant relations were found between mean absolute error of number line estimation and achievement test scores, although trends ($p <$ .10) in the expected direction were found for first and second graders.

Individual differences on the computational estimation task were also related to individual differences in achievement test scores. Percentage correct on computational estimation items cor-

related positively with achievement test scores for kindergartners, $r(18)$ = .56, $p <$ .01; second graders, $r(21)$ = .81, $p <$ .01; and third graders, $r(18)$ = .60, $p <$ .01. Numerosity judgment scores correlated positively with math achievement scores for second graders, $r(21)$ = .53, $p <$ .01, and third graders, $r(18)$ = .47, $p <$ .01, though not for kindergartners or first graders. No significant relations with achievement test scores were found for either measurement estimation task at any grade level.

Discussion

The results of the experiment can be easily summarized. Kindergartners' to third graders' performance on the number line task replicated the previously observed shift from reliance on a logarithmic representation of numerical magnitudes to reliance on a linear representation of them. Estimation accuracy also improved substantially during this period; second and third graders' percent absolute error was less than half that of kindergartners. At all four grade levels, individual differences in linearity of number line estimates were positively related to individual differences in overall math achievement. Performance on the other three less-formally assessed estimation tasks also improved substantially in this age period, and individual differences in them were related to individual differences in math achievement test scores at several grades on the computational and numerosity estimation tasks.

The results from Experiment 1 also raised two major questions. One was whether the logarithmic-to-linear shift is present not only on number line estimation but also on other types of pure numerical estimation. The measurement, numerosity, and computational estimation tasks in Experiment 1 did not yield any measure of linearity and therefore did not allow this question to be addressed.

The other main question raised by the results of Experiment 1 was how strongly different types of pure numerical estimation are related. The tasks used in Experiment 1 were inadequate for answering questions about the strength of relations among different types of estimation because of their limited number of items (several tasks had only four items) and extremely variable response formats (on different tasks, children responded by drawing lines, judging line lengths, answering two-choice multiple-choice items, answering three-choice multiple-choice items, and placing hatch marks on lines). Nonetheless, a number of relations among the four types of estimation were evident. Individual differences in number line and computational estimation were positively related at all four ages (rs = .38–.66, two ps < .10 and two ps < .01). It was probably not coincidental that these were the two tasks with a reasonable number of items (26 and 12, respectively). Relations were also present for some of the other pairs of tasks at some grades; for example, accuracy of computational and numerosity estimation was significantly related for kindergartners and second graders (rs = .50 and .54, respectively, ps < .05), though not for first or third graders. The methodological limitations noted above made these results on consistency of individual differences in estimation far from conclusive, but the results were sufficiently promising to justify a more rigorous assessment of individual, as well as developmental, differences in pure numerical estimation.

Experiment 2

Experiment 2 was designed to examine four specific issues: (a) whether the developmental progression from logarithmic to linear

representations that was observed on number line estimation also is present on other types of pure numerical estimation, (b) whether individual differences in different types of pure numerical estimation are interrelated, (c) whether the relation of linearity of estimation to math achievement test scores that has been demonstrated through second grade is also present beyond that point, and (d) whether any consistent individual differences in estimation that emerge are simply reflections of a common relation to overall math achievement. To address these issues, we presented second and fourth graders with detailed assessments of the four types of estimation examined in Experiment 1. The problems involved numbers between 0 and 1,000, rather than 0 and 100, because Siegler and Opfer (2003) reported a logarithmic-to-linear shift on 0–1,000 number lines in this age range.

We predicted that the previously observed developmental trend toward increasingly linear estimation patterns would extend to pure numerical estimation problems other than those involving number lines. The reasoning was that all types of pure numerical estimation draw on the same representations of numerical magnitudes and that the increasing familiarity with numbers in the 0–1,000 range that was hypothesized to lead to the logarithmic-to-linear shift in number line estimation in Siegler and Opfer (2003) would produce the same effect on other types of pure numerical estimation.

Our second prediction was that individual differences in linearity of estimates on the different tasks would be positively correlated. The reasoning again was that all types of pure numerical estimation are based on a common representation of numerical magnitudes and that this common core would be clearly evident when irrelevant sources of variation, such as those in number of items and response formats, were eliminated.

The third prediction was that individual differences in accuracy and linearity of estimation on the four tasks would be related to children's math achievement test scores. This prediction was based on the belief that children in this age range, like younger children, vary in the representation of numerical magnitude that they use with numbers in this range and that reliance on linear representations both reflects and promotes better math learning in general.

Finally, the fourth prediction was that individual differences in performance on the four estimation tasks would be related to each other, above and beyond their common relation to math achievement. This prediction was based on our belief that performance on the estimation tasks is a quite direct reflection of a common numerical representation, whereas math achievement test performance reflects that representation but also reflects other factors, such as memorization of arithmetic facts and mastery of mathematical procedures.

Method

Participants

Fifty-eight children took part: 30 second graders (mean age = 7.8 years, $SD = 0.4$) and 28 fourth graders (mean age = 9.9 years, $SD = 0.4$). Among the participants, 96% were Caucasian, 2% African American, and 2% Indian American. The children were recruited from a public school in which 19% of children were eligible for free or reduced-cost lunches; the school was in a predominantly middle-class area. Participation in the study was completely voluntary, and children were not given extrinsic rewards for their participation. The experimenter was a Caucasian, female graduate student.

Teachers within the school taught estimation skills more frequently than peers in other schools in which we have conducted studies of estimation. On a questionnaire about teaching practices, four of the five teachers indicated that they taught computational and measurement estimation once a week or more and that they taught numerosity estimation about once a month; the other teacher indicated that she taught numerosity and number line estimation about once a week and computational and measurement estimation about once a month.

Procedure and Materials

Children met one-on-one with the experimenter for a single 20-min session. The experimenter presented the four estimation tasks in random order.

The number line procedure was identical to that in Experiment 1, except that the scale was 0–1,000 instead of 0–100 and the orienting task was replaced with two practice trials; one required marking 1,000 on a number line, and the other involved marking 0 on the line. The 22 numbers that were presented in the experimental phase again slightly overrepresented the low end of the scale; they included 4 numbers between 0 and 100 and 2 numbers from each successive hundred. The numbers were 3, 7, 19, 52, 103, 158, 240, 297, 346, 391, 438, 475, 502, 586, 613, 690, 721, 760, 835, 874, 907, and 962. Order of presentation of the numbers was randomized separately for each child.

The measurement task involved presentation of 22 sheets of paper, each with two preprinted lines near the top. One (very short) line measured 0.034 cm and was labeled *1 zip*; the other (much longer) line measured 34 cm and was labeled *1,000 zips*. The desired length (in number of zips) for the line that the child was to produce on a given trial was printed in the middle of the page, and the bottom of the sheet was left blank for drawing the line. The same numbers used for the number line task were used to indicate the desired length in zips; the numbers were again presented in random order. Children were first shown 1 of the measurement estimation sheets, told the length of the two lines at the top of the page, and asked to trace each line with a finger. They were also told that these two lines would be on each page to remind them of the lengths of 1 zip and 1,000 zips. Then, children were shown a special viewer made of clear Plexiglas. The viewer held the sheets of paper snugly and contained a window 1,000 zips (34 cm) long that could be used to draw straight lines. After practice drawing lengths of 1 zip and 1,000 zips along the edge of the viewer, children were asked to draw lines of the 22 lengths listed above, 1 per page.

For the numerosity estimation task, stimuli were presented with a computer program that generated dots in a box on a monitor. When full, the box held 1,000 dots in 40 columns and 25 rows. Each child was first shown an empty box and a full box, told that the empty one had 0 dots and the full one 1,000 dots, and informed that the two boxes would always be present to remind them of what 0 and 1,000 dots looked like. The children were then shown how to hold the "Increase" button and the "Decrease" button to make dots appear in or disappear from a third box; the dots appeared or disappeared in random order from the third box when the relevant button was held down. Children were also told that when the box contained the number of dots they wanted on a trial, they should click the "Finish" button. When they did, the computer recorded the number of dots that the box contained and reset the number of dots to 0. After 2 practice trials making pictures with 1,000 and 0 dots, children were presented with 22 trials, with numbers matched to those presented in the number line and measurement estimation tasks. As in the other tasks, the numbers were presented in random order.

Because computational estimation differs from the other types of pure numerical estimation in several inherent ways—the presence of two numerical inputs rather than one, the relevance and likely use of previously learned arithmetic facts, and the relevance and likely use of previously learned rounding rules—the procedure used to examine it differed in several ways from those used on the other three tasks. The experimenter

showed children a series of 20 cards, each containing an arithmetic problem. Ten of the cards (5 addition, 5 subtraction) presented problems with one 3-digit and one 2-digit number (e.g., $377 + 82$, $443 − 38$). The other 10 cards (5 addition, 5 subtraction) presented problems with two 3-digit numbers (e.g., $227 + 195$, $639 − 344$). Answers to addition problems ranged from 222 to 997; answers to subtraction problems ranged from 87 to 816. The experimenter read each problem aloud, and children verbally stated their answers.

Achievement scores from the math sections of the Terra Nova Test that the children had taken approximately 2 months prior to the study as part of their regular school activities were obtained from school records.

Results

We first examined changes with age, then consistencies of individual differences across the four types of estimation, and then the relation of estimation performance to math achievement test scores.

Age-Related Changes

We examined two measures of age-related changes on each task: changes in the best-fitting function and changes in accuracy. As in Experiment 1, the exponential function fit the children's median estimates less well than did the other two functions, so it was not included in further analyses.

Number line estimation. On the number line task, the logarithmic and linear functions fit second graders' median estimates for each number equally well ($R^2_{lin} = .91$ and $R^2_{log} = .88$), $t(21) = 1.07$, *ns*, $d = 0.38$ (see Figure 2). In contrast, fourth graders' median estimates were much better fit by the linear function than by the logarithmic one ($R^2_{lin} = .98$ and $R^2_{log} = .71$), $t(21) = 6.50$, $p < .01$, $d = 1.90$. From another perspective, the variance in the group medians accounted for by the linear function increased between second and fourth grade (from 91% to 98%), and the variance accounted for by the logarithmic function decreased (from 88% to 71%).

Analyses of the number line estimates of individual children yielded similar results. The type of function that fit the most children varied with age, $\chi^2(1, N = 58) = 15.23$, $p < .01$, $V = .51$.

The logarithmic function provided the better fit for 60% of second graders and the linear function for 40%. In contrast, the linear function provided a better fit than the logarithmic function for 89% of fourth graders, whereas the reverse was true for only 11%. Seen from another perspective, the percentage of children for whom the linear function provided the better fit increased from 40% of second graders to 89% of fourth graders, and the percentage for whom the logarithmic function provided the better fit decreased from 60% to 11%. Paired-sample *t* tests indicated that for individual second graders, the fit of the linear and logarithmic functions was equal (mean $R^2_{log} = .68$, $SE = .04$, vs. $R^2_{lin} = .66$, $SE = .03$), $t(29) = 0.65$, *ns*, $d = 0.11$, and that for individual fourth graders, the linear function provided a better fit (mean $R^2_{log} = .66$, $SE = .02$, vs. mean $R^2_{lin} = .85$, $SE = .03$), $t(27) = 6.19$, $p < .01$, $d = 1.38$. A one-way ANOVA indicated that the fit of the linear function to individual children's number line estimates increased with grade, $F(1, 56) = 17.80$, $p < .01$, $\eta^2 = .24$.

Percent absolute error of number line estimates decreased with grade. Second graders' estimates were farther from the correct answer than were the estimates of fourth graders (percent absolute error = 17%, $SE = .01$, vs. 10%, $SE = .01$), $F(1, 56) = 24.61$, $p < .01$, $\eta^2 = .31$. Thus, the accuracy of estimation in the current study was somewhat better than that obtained by Siegler and Opfer (2003) for second graders (percent absolute error = 19%) and fourth graders (12%), though the improvement between second and fourth grade was identical (7% in both cases).

Measurement estimation. A similar pattern of change emerged in analyses of the measurement estimation data. Second graders' median estimates on the measurement estimation task were equally well fit by the logarithmic function ($R^2 = .91$) and the linear function ($R^2 = .85$), $t(21) = 0.36$, *ns*, $d = 0.12$ (see Figure 3). In contrast, fourth graders' measurement estimates were better fit by the linear function than by the logarithmic one ($R^2 = .98$ and .74), $t(21) = 5.89$, $p < .01$, $d = 1.63$. Thus, the variance accounted for by the linear function increased with age (from 85% to 98%), and the variance accounted for by the logarithmic function decreased with age (from 91% to 74%).
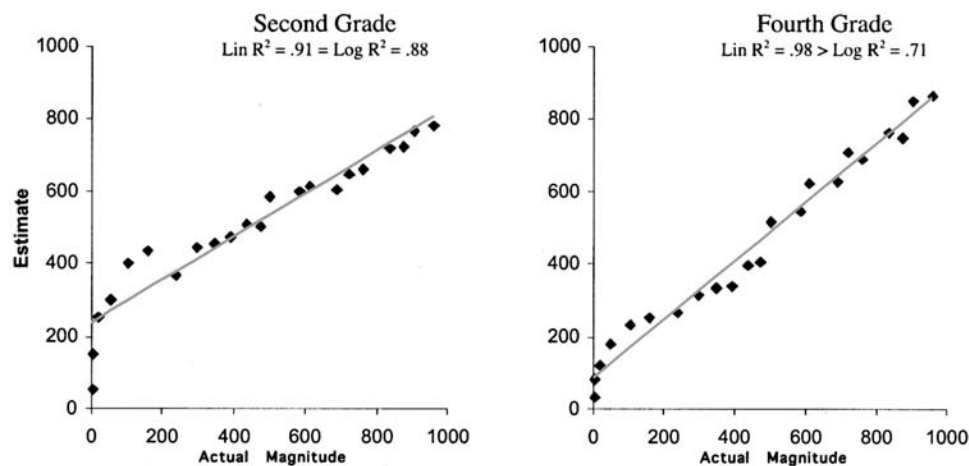

Figure 2. Experiment 2: Best-fitting equations for median number line estimates for second and fourth graders. Lin = linear; Log = logarithmic.
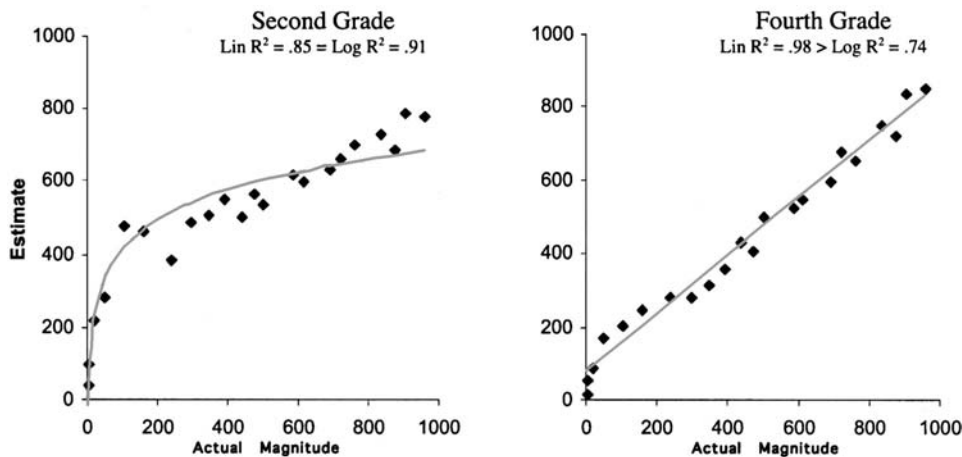
Figure 3. Experiment 2: Best-fitting equations for median measurement estimates for second and fourth graders. Lin = linear; Log = logarithmic.

A similar pattern emerged in analyses of individual children's measurement estimates. The type of function that fit the estimates of the greatest number of children varied with age, $\chi^2(1, N = 57) = 13.01$, $p < .01$, $V = .48$. Individual second graders' measurement estimates were more likely to be best fit by the logarithmic function than by the linear function (70% vs. 30%). In contrast, fourth graders' estimates were more likely to be best fit by the linear function than by the logarithmic function (78% vs. 22%). Thus, the percentage of children best fit by the linear function increased from 30% to 78%, and the percentage best fit by the logarithmic function decreased from 70% to 22%. Individual second graders' estimates were fit better by the logarithmic function than by the linear one (mean $R^2_{\log} = .67$, $SE = .03$, vs. mean $R^2_{\lin} = .59$, $SE = .03$), $t(29) = 2.32$, $p < .05$, $d = 0.43$. In contrast, individual fourth graders' estimates were better fit by the linear than by the logarithmic function (mean $R^2_{\lin} = .83$, $SE = .03$, vs. mean $R^2_{\log} = .68$, $SE = .02$), $t(27) = 5.22$, $p < .01$, $d = 1.09$. The

mean fit of the linear function to individual children's estimates also increased with grade, $F(1, 56) = 26.99$, $p < .01$, $\eta^2 = .33$.

Estimation accuracy, like the linearity of estimates, increased with age. Percent absolute error decreased from 19% ($SE = .01$) among second graders to 12% ($SE = .01$) among fourth graders, $F(1, 56) = 17.13$, $p < .01$, $\eta^2 = .23$.

Numerosity estimation. Changes with age in numerosity estimation closely paralleled the changes in number line and measurement estimation. Second graders' numerosity estimates were equally well fit by the logarithmic function ($R^2 = .90$) and the linear function ($R^2 = .85$), $t(21) = 0.60$, *ns*, $d = 0.19$ (see Figure 4). In contrast, fourth graders' numerosity estimates were better fit by the linear function than by the logarithmic one ($R^2 = .96$ and .78), $t(21) = 3.60$, $p < .01$, $d = 1.02$. Seen from another perspective, the fit of the linear function to the median numerosity estimate for each number increased from .85 for second graders to .96 for fourth graders, whereas the fit of the
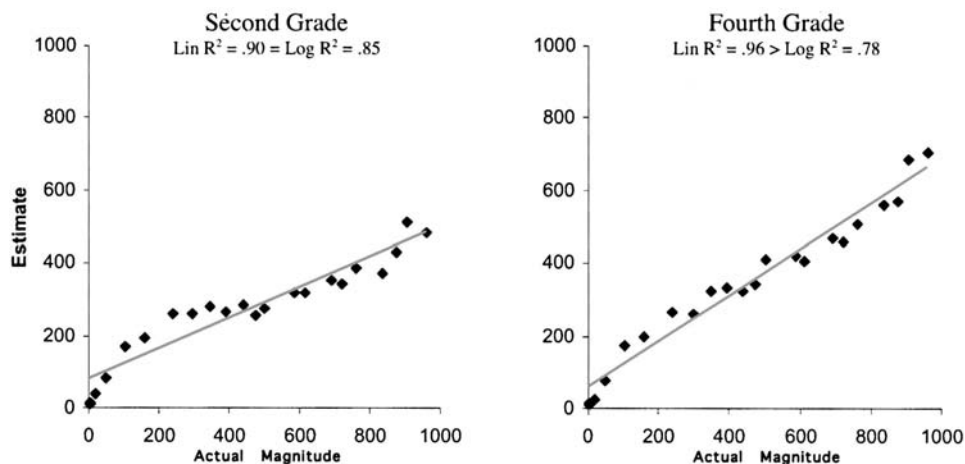

Figure 4. Experiment 2: Best-fitting equations for median numerosity estimates for second and fourth graders. Lin = linear; Log = logarithmic.

logarithmic function decreased from .90 for second graders to .78 for fourth graders.

The data on individual children's estimates showed a similar pattern. The number of children for whom the logarithmic and linear functions provided the best fit varied with age, $\chi^2(1, N = 56) = 9.22$, $p < .01$, $V = .41$. For second graders, the two functions provided the best fit for roughly equal numbers of children (the logarithmic function fit best for 57% and the linear function for 43%), whereas for fourth graders, the linear function provided the best fit for 82% and the logarithmic function for 18%. Put another way, the percentage of children for whom the linear function provided the best fit increased from 43% to 82%, whereas the percentage of children for whom the logarithmic function provided the best fit decreased from 57% to 18%. In addition, the fit of the linear and logarithmic functions to individual second graders' estimates was equal (mean $R^2_{\log} = .59$, $SE = .02$, vs. mean $R^2_{\lin} = .57$, $SE = .03$), $t(29) = 1.02$, ns, $d = 0.12$, and the linear function provided a better fit to the estimates of individual fourth graders (mean $R^2_{\lin} = .77$, $SE = .03$, vs. mean $R^2_{\log} = .66$, $SE = .02$), $t(27) = 4.15$, $p < .01$, $d = 0.82$. The mean fit of the linear function to individual children's estimates also increased, $F(1, 56) = 20.56$, $p < .01$, $\eta^2 = .27$.

Accuracy, like linearity, increased with age. Percent absolute error of numerosity estimates decreased from 22% ($SE = .01$) among second graders to 15% ($SE = .01$) among fourth graders, $F(1, 56) = 16.71$, $p < .01$, $\eta^2 = .23$.

Computational estimation. The computational estimation task did not allow calculation of linearity scores. However, percent absolute error showed the same 7% decrease as it did on the other three tasks (and in Siegler & Opfer, 2003, in the same age range). In this case, mean percentage of absolute error decreased from 11% ($SE = .02$) among second graders to 4% ($SE = .00$) among fourth graders, $F(1, 56) = 13.30$, $p < .01$, $\eta^2 = .19$.

Relations Among Estimation Tasks

To examine relations among different types of estimation, we first correlated individual children's percent absolute error on the four estimation tasks. As shown in Table 1, individual differences in the accuracy of the three highly parallel tasks—number line,

measurement, and numerosity estimation—were consistently related (five of the six correlations among these three types of estimation were significant). Individual differences in second graders' computational estimation accuracy were also related to individual differences on two of the other three tasks, but no relations between individual differences in fourth graders' computational estimation and individual differences on the other three tasks were significant at the .05 level. This lack of relations between fourth graders' performance on the computational estimation task and their performance on the other three tasks may have been due to ceiling effects; mean percent absolute error for fourth graders' computational estimates was 4%, with a standard deviation of only 2%.

We next examined the consistency of individual differences in children's linearity of estimates ($R^2_{\lin}$) on the three tasks for which linearity could be computed (number line, measurement, and numerosity). As shown in Table 2, for second graders, all six correlations were significant, and for fourth graders, five of the six correlations were significant. The correlations tended to be quite substantial; 8 of the 12 were between .50 and .84. Also as shown in Table 2, five of the six correlations between children's percent absolute error on the computational estimation task and their linearity on the other three tasks were significant. In light of the ceiling effects noted in the previous paragraph, the relation of individual differences in the accuracy of fourth graders' computational estimation to individual differences in their linearity on the other three estimation tasks was surprising.

If the linearity of children's representations of numerical magnitude is the key to these relations among accuracy of estimation on different tasks, then partialing out each child's linearity on a task should reduce or eliminate the correlations between the child's accuracy on that task and on the other two estimation tasks for which linearity could be computed. This prediction proved accurate. Partialing out linearity on one of the two tasks resulted in only 2 of the 12 possible correlations among percent absolute error remaining significant. The 2 correlations that remained significant after partialing out linearity involved fourth graders: the correlation between accuracy on the number line and numerosity tasks,

Table 1

Experiment 2: Correlations Among Individual Children's Percent Absolute Error on Four Estimation Tasks

Task	Measurement	Numerosity	Computational
	Second graders		
Number line	.62**	.45*	.53**
Measurement		.28	.57**
Numerosity			.30
	Fourth graders		
Number line	.76**	.41*	.35
Measurement		.38*	.33
Numerosity			.31

Note. $df = 28$ for second graders and 26 for fourth graders.
* $p < .05$. ** $p < .01$.

Table 2

Experiment 2: Correlations Among Individual Children's Linearity on Number Line, Measurement, and Numerosity Estimation Tasks and Percent Absolute Error on the Computational Estimation Task

Task	Measurement	Numerosity	Computational
	Second graders		
Number line	.65**	.55*	−.52**
Measurement		.54**	−.42*
Numerosity			−.46*
	Fourth graders		
Number line	.84**	.70**	−.29
Measurement		.60**	−.50**
Numerosity			−.39*

Note. $df = 28$ for second graders and 26 for fourth graders.
* $p < .05$. ** $p < .01$.

$r(25) = .44$, $p < .05$, and the correlation between accuracy on the number line and measurement tasks, $r(25) = .55$, $p < .01$.

This effect was not symmetrical; 7 of the 12 correlations between each child's linearity on pairs of tasks remained significant after partialing out the child's percent absolute error on one of the tasks. For second graders, after partialing out each child's percent absolute error on the numerosity task, significant correlations remained between the child's linearity on the numerosity and number line tasks, $r(27) = .39$, $p < .05$, and between the child's linearity on the numerosity and measurement tasks, $r(27) = .44$, $p < .05$. Similarly, after partialing out second graders' percent absolute error on the measurement estimation task, a significant correlation remained between linearity on the measurement and number line tasks, $r(27) = .37$, $p < .05$. For fourth graders, after partialing out percent absolute error on the measurement estimation task, significant correlations remained between R^2_{lin} on that task and the number line task, $r(25) = .71$, $p < .01$, and between R^2_{lin} on that task and the numerosity task, $r(25) = .46$, $p < .05$. Similarly, after partialing out fourth graders' percent absolute error on the numerosity task, significant correlations remained between R^2_{lin} on that task and the number line task, $r(25) = .73$, $p < .01$, and on that task and the measurement task, $r(25) = .56$, $p < .01$. Thus, the consistency of individual differences on these estimation tasks seemed largely attributable to differences in the linearity of children's representations of numerical magnitudes.

Estimation and Math Achievement

To examine the relation of estimation accuracy to math achievement, we correlated individual children's percent absolute error on each of the estimation tasks to that child's percentile ranking on the achievement test. One of the second graders did not take the achievement test, so her data were excluded from this analysis.

Accuracy and linearity of estimation were consistently related to math achievement test scores at both grade levels. As shown at the top of Table 3, second graders' math achievement test scores were related to the linearity of their estimates for all three types of estimation for which linearity could be computed; the same was true for fourth graders for two of the three tasks, and a trend in the predicted direction was present on the third task. As shown at the bottom of Table 3, second graders' percent absolute error was related to their achievement test scores on all four types of estimation; for fourth graders, the same relation was present on three of the four types of estimation.

Were Relations Among Estimation Tasks Attributable to Shared Relations to Math Achievement?

We next tested whether relations among different types of estimation were attributable to their common relation to overall math achievement. With math achievement test scores controlled for, partial correlations were computed for the linearity of estimates on the three pairs of tasks on which this measure could be computed. For second graders, correlations of the linearity of estimates on two of the three pairs of tasks remained significant after math achievement was partialed out: the correlations between the linearity of number line and measurement estimation, $r(26) = .48$, $p < .01$, and between the linearity of number line and numerosity estimation, $r(26) = .39$, $p < .05$. A trend toward significance was also present on the third pair of tasks, measurement and numerosity estimation, $r(26) = .34$, $p < .10$. Among fourth graders, the relations among individual children's linearity of estimates for all three pairs of tasks remained significant after the children's math achievement scores were partialed out: number line and measurement estimation, $r(25) = .78$, $p < .01$; number line and numerosity estimation, $r(25) = .65$, $p < .01$; and measurement and numerosity estimation, $r(25) = .51$, $p < .01$.

The correlations of percent absolute error on pairs of estimation tasks decreased somewhat more when achievement test scores were partialed from the correlations. For second graders, only the correlation between percent absolute error on the number line and measurement estimation tasks remained significant, $r(26) = .39$, $p < .05$, though trends toward significance were also present for the correlations between performance on the computational and number line estimation tasks and between performance on the computational and measurement estimation tasks, $rs(26) = .33$, $ps < .10$. For fourth graders, the partial correlation between percent absolute error on the number line and measurement estimation tasks remained significant, $r(25) = .70$, $p < .01$, as did the correlation between number line and numerosity estimation, $r(25) = .38$, $p < .05$; a trend toward significance was also present for the correlation between percent absolute error on the measurement and numerosity estimation tasks, $r(25) = .33$, $p < .10$.

Discussion

Numerous researchers and organizations of mathematics educators have concluded that children have poor estimation skills (Case & Sowder, 1990; Dowker, 2003; Geary, 1994; Hiebert & Wearne,

Table 3

Experiment 2: Correlations Between Individual Children's Math Achievement Scores and Their Linearity and Accuracy on Four Estimation Tasks

Variable	Number line	Measurement	Numerosity	Computational
Linearity				
Second grade	.53**	.62**	.48**	
Fourth grade	.47*	.54**	.35	
Percent absolute error				
Second grade	−.57**	−.67**	−.45*	−.53**
Fourth grade	−.45*	−.45*	−.16	−.49**

Note. Linearity could not be calculated on the computational estimation task. $df = 27$ for second graders and 26 for fourth graders.
* $p < .05$. ** $p < .01$.

1986; Joram, Subrahmanyam, & Gelman, 1998; National Council of Teachers of Mathematics, 2000). The explanations that have been advanced—inadequate number sense, mindless symbol manipulation, and poor understanding of principles—seem generally in the right ballpark but are vague and are more restatements of the data than explanations of them. The present study was based on a more specific explanation: that a large part of the reason for elementary schoolchildren's poor estimation, as well as for developmental and individual differences in their estimation, is reliance on logarithmic rather than linear representations of numerical magnitudes. Results of both experiments were consistent with this explanation. In this concluding section, we discuss the implications of the findings for understanding children's poor estimation skills as well as for understanding developmental and individual differences in those skills.

Development of Pure Numerical Estimation

The previously observed developmental progression away from reliance on logarithmic representations and toward reliance on linear ones was replicated and extended. The present findings replicated Siegler and Booth's (2004) and Siegler and Opfer's (2003) observations that between kindergarten and fourth grade, children increasingly often generate linearly increasing estimates of numerical magnitudes and decreasingly often generate logarithmically increasing estimation patterns. In the present study, this developmental trend was found to apply not only to number line estimation, for which it had previously been found, but also to measurement and numerosity estimation.

The magnitude of the developmental shift was remarkably similar across the present and previous studies. First, consider the change from kindergarten to second grade on 0–100 number lines. In Experiment 1 of the present study, there was a 54% decrease, from 80% of kindergartners to 26% of second graders, in the percentage of children whose number line estimates were best fit by the logarithmic function. In the two experiments by Siegler and Booth (2004), the average decrease between kindergarten and second grade was 48%—from 80% to 32%. Similarly, the percent absolute error in Experiment 1 of the present study decreased from 24% among kindergartners to 10% for second graders. In the two experiments by Siegler and Booth (2004), the percent absolute error decreased from an average of 25% among kindergartners to 12% among second graders.

Now consider the improvement from second to fourth grade on 0–1,000 number lines. In Experiment 2 of the present study, the percentage of children whose number line estimates were best fit by the logarithmic function decreased from second to fourth grade by 49%—from 60% to 11%. Siegler and Opfer (2003) found that the decrease over the same period was 47%—from 91% to 44%. Similarly, in Experiment 2 of the present study, the percent absolute error decreased 7% from second to fourth grade, from 17% to 10%, just as it did in the same age range in Siegler and Opfer's study, in which the decrease was from 19% to 12%. The absolute level of number line performance varied considerably from experiment to experiment, but the magnitude of improvement was strikingly similar.

Equally striking was the comparability of improvement across different types of pure numerical estimation. In Experiment 2 of the present study, the percent absolute error decreased by exactly 7% on all four tasks: from 17% to 10% on number line estimation, from 19% to 12% on measurement estimation, from 22% to 15% on numerosity estimation, and from 11% to 4% on computational estimation. Similarly, the percentage of variance accounted for by the linear function increased 13% on measurement estimation (from 85% to 98%), 11% on numerosity estimation (from 85% to 96%), and 7% on number line estimation (from 91% to 98%). The similarity of the developmental changes in accuracy and linearity suggests a common source of development across pure numerical estimation tasks.

Several other results from Experiment 2 provide additional evidence that inappropriate representation of numerical magnitudes is a general problem across different types of pure numerical estimation. First, the average fit of the linear function to individual children's estimates was similar across the three tasks for which linearity could be computed. This was true for second graders, for whom R^2_{lin} accounted for 66%, 59%, and 57% of the variance in number line, measurement, and numerosity estimation, respectively. It was also true for fourth graders, for whom R^2_{lin} accounted for 85%, 83%, and 77% of variance, respectively. In addition, at both grade levels, the fit of the linear representation was highest for number line estimates, followed by measurement estimates and then numerosity estimates. This pattern seemed to reflect the degree to which numerical magnitude representations were the primary determinant of children's responses. Thus, it was not surprising that R^2_{lin} was somewhat lower on the numerosity task than on the other tasks, because the numerosity task, unlike the others, required consideration of two spatial dimensions. Number of spatial dimensions that children need to consider has been found to influence the difficulty of estimation tasks (Siegel, Goldsmith, & Madson, 1982).

Might the developmental parallels across the different types of estimation in Experiment 2 have been due to the tasks used to measure them differing only in trivial ways? Close examination of the tasks argued against this possibility. As noted in the previous paragraph, the numerosity estimation task required consideration of both the width and the height of the box of dots, the number line task required placement of a hatch mark on a line, and the measurement estimation task required drawing a line of a specified length. Performance on all three tasks also correlated substantially with performance on computational estimation problems, which involved a numerical rather than a spatial response and which involved two numerical inputs rather than one. Thus, the developmental parallels among the four types of pure numerical estimation reflected more than task similarity.

Individual Differences

The data on individual differences also supported the conclusions that performance on different pure numerical estimation tasks has a common core and that the common core is the linearity of representations of numerical magnitude. Some consistent individual differences across the four estimation tasks were present even in Experiment 1, in which the tasks varied on numerous irrelevant dimensions and in which two of the tasks had only 4 items. The consistencies were stronger in Experiment 2, in which all four tasks had at least 20 items and in which three of the four tasks used parallel formats. There, all six correlations among individual second and fourth graders' performance on the three

tasks for which linearity could be computed were significant and substantial, ranging from .54 to .84. In addition, five of the six correlations of individual children's linearity on these three tasks with their percent absolute error on computational estimation, the task on which linearity could not be computed, also were significant.

These relations among individual differences in linearity of estimates could not be explained in terms of general math ability. Estimation performance on all three tasks for which linearity could be computed was related to math achievement test scores, but partialing out the achievement test scores left five of the six correlations significant (and a trend toward significance was present on the sixth correlation). In most cases, the partial correlations were quite substantial; for example, for fourth graders, they ranged from .51 to .78. In addition, all three correlations of linearity of individual children's estimates on the pairs of tasks for which linearity could be computed were higher than any of the three correlations between linearity of estimates on those tasks and math achievement test scores.

In contrast, partialing out math achievement test scores did eliminate the correlations between computational estimation performance and performance on the other three estimation tasks. Computational estimation skill during elementary school seems likely to depend largely on ability to memorize arithmetic facts and to use specific procedures for solving the problems (e.g., rounding). Accurate execution of the typical rounding process used by elementary school students (substituting 0s for all but the leftmost digit and then performing the operation) requires sufficient computational skill to perform the operation. Consistent with this analysis, skill at computational estimation is empirically related to skill at exact computation (Dowker, 2003; LeFevre et al., 1993). Computational skill also is tested extensively on achievement tests, which probably accounts for the large effects of partialing out the achievement test scores from correlations involving computational estimation.

Comparing correlations of the linearity of estimates on different tasks with correlations of the accuracy of estimates on different tasks also supported the interpretation that degree of linearity was a key determinant of the individual differences in estimation, at least on the three tasks for which linearity could be computed. In all six correlations among the three tasks for which linearity could be computed, the correlations between individual children's degree of linearity on pairs of tasks were greater than the correlations between their accuracy (percent absolute error) on the same pairs of tasks. Moreover, partialing out linearity rendered nonsignificant almost all correlations between accuracy on pairs of tasks, but partialing out accuracy of estimates did not have a comparable effect on the relations of linearity on pairs of tasks.

Other factors not measured in the present study almost certainly contribute to individual differences in children's estimation. For example, working memory functioning may influence children's ability to simultaneously consider the low and high ends of the scale and their relation to the number being estimated and thus may reduce the accuracy of estimates. In addition, future studies should obtain a measure of reading achievement, which would be informative for evaluating whether the estimation tasks are assessing a pure numerical competency or more general aptitude levels. Nonetheless, the present findings about individual differences, like those about developmental differences, support the view that reli-

ance on nonoptimal representations of numerical magnitudes is a major source of children's inaccurate estimation.

Questions and Implications

One class of questions raised by the present findings concerns the breadth of influence of logarithmic representations of numerical magnitudes. The present findings indicate that kindergartners frequently rely on logarithmic representations when generating estimates in the 0–100 range and that second graders frequently do so when operating in the 0–1,000 range. Would older children and adults perform similarly in the 0–1,000,000 or 0–1,000,000,000 range or on scales with odd endpoints such as 0–432,751, especially when under time pressure?

The overlapping waves perspective, which underlies the present research, suggests that varied representations of numerical magnitude persist over time and that when operating in unfamiliar numerical ranges, adults and older children may rely on the intuitive and widely applicable logarithmic representation. It may be the case that people never automatically use linear representations of numerical magnitude for all types of numbers. Instead, when operating in unfamiliar numerical ranges, even adults may need to override the impulse to use the logarithmic representation. Viewed from an evolutionary perspective, logarithmic representations may serve as the default option for all types of quantitative computations. Rats and pigeons, as well as human infants and adults, have been found to apply logarithmic representations to a wide variety of quantitative dimensions (Banks & Hill, 1974; Dehaene, 1997; Holyoak & Mah, 1982). Such representations have survival value, because differences at the low end of quantitative dimensions frequently matter more than do equal size differences among larger values. For example, the difference between 1 and 2 pieces of food would be more important for a hungry animal than would the difference between 41 and 42 pieces. Thus, adults as well as elementary schoolchildren may rely on logarithmic representations in unfamiliar numerical ranges. The frequent confusion of millions, billions, and trillions in news magazines and political discussions lends plausibility to this prediction.

A second class of questions involves whether children rely on logarithmic representations on numerical tasks other than estimation and, if so, what consequences this reliance has for learning of these tasks. For example, reliance on a logarithmic representation of numerical magnitudes might facilitate learning of multiplication facts with small products but hinder learning of facts with large products. The reason is that the logarithmic representation exaggerates the psychological distance between small numbers but understates the psychological distance between large ones. Greater psychological distance seems likely to make answers more distinctive and hence more memorable. The present findings, as well as those of Siegler and Booth (2004), indicate that reliance on a logarithmic representation is associated with low math achievement; this association makes it unlikely that the absolute level of performance would be higher for any range of arithmetic problems for children who adopted a logarithmic representation. Factors other than numerical representations, such as working memory capacity and speed of processing, also influence learning of arithmetic facts (Geary, Hoard, Byrd-Craven, & DeSoto, 2004). However, the present perspective does predict that the greater the sum or product, the greater the difference will be between the learning

of children who rely on logarithmic representations and that of peers who rely on linear representations.

A third class of questions involves practical applications of the present findings. One question concerns whether using number lines to display the linear magnitudes associated with each operand when arithmetic problems are presented and then displaying the magnitude of the answer when feedback is given will improve children's learning. Learning answers to arithmetic problems is not a rote process; children learn not only the verbal label of the correct answer but also its approximate magnitude. Evidence for this conclusion comes from both verification and production tasks. On verification tasks, children are quicker to reject errors that are far from the correct answer to an arithmetic problem than errors that are close to it (Ashcraft, 1987). On production tasks, a disproportionate percentage of children's errors on arithmetic problems are close misses (Siegler, 1988). Reliance on a logarithmic representation would increase the difficulty of learning the magnitudes of answers, especially in high ranges in which the logarithmic representation minimizes the psychological distance between numbers. Thus, presenting spatial displays of the linear magnitudes associated with operands and answers during learning might well hasten the learning of correct answers and also enable children to generate plausible rather than implausible errors. More generally, identifying representations of numerical magnitude as a source of children's difficulty in estimation, and perhaps in learning math more generally, opens a new dimension for mathematics educators to consider.

References

Ashcraft, M. H. (1987). Children's knowledge of simple arithmetic: A developmental model and simulation. In J. Bisanz, C. J. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology* (pp. 302–338). New York: Springer-Verlag.

Banks, W. P., & Hill, D. K. (1974). The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology Monograph, 102,* 353–376.

Case, R., & Sowder, J. T. (1990). The development of computational estimation: A neo-Piagetian analysis. *Cognition and Instruction, 7,* 79–104.

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

Dowker, A. D. (1998). Individual differences in normal arithmetic development. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 275–302). Hove, England: Psychology Press.

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. 243–266). Mahwah, NJ: Erlbaum.

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

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology, 88,* 121–151.

Hiebert, J., & Wearne, D. (1986). Procedures over concepts: The acquisition of decimal number knowledge. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 199–223). Hillsdale, NJ: Erlbaum.

Holyoak, K. J., & Mah, W. A. (1982). Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology, 14,* 328–352.

Hook, L. (1992). *Estimation strategies of psychology students.* Unpublished manuscript, University of Oxford, Oxford, England.

Joram, E., Subrahmanyam, K., & Gelman, R. (1998). Measurement estimation: Learning to map the route from number to quantity and back. *Review of Educational Research, 68,* 413–449.

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

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

Newman, R. S., & Berger, C. F. (1984). Children's numerical estimation: Flexibility in the use of counting. *Journal of Educational Psychology, 76,* 55–64.

Paull, D. R. (1972). *The ability to estimate in mathematics.* Unpublished doctoral dissertation, Columbia University.

Petitto, A. L. (1990). Development of numberline and measurement concepts. *Cognition and Instruction, 7,* 55–78.

Siegel, A. W., Goldsmith, L. T., & Madson, C. R. (1982). Skill in estimation problems of extent and numerosity. *Journal for Research in Mathematics Education, 13,* 211–232.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General, 117,* 258–275.

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

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