Numerical Magnitude Representations Influence Arithmetic Learning

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

Funding for this research was provided by the National Institutes of Health Grant HD 19011 and the Office of Educational Research and Instruction Grant R305H020060. Thanks are also due to the Hopewell Area School District for allowing us to collect data in their schools and to Mary Wolfson for assistance with data collection.

Correspondence should be addressed to Julie L. Booth, Human Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, 15213; phone, 412-268-9418; FAX, 412-268-9433; email, juliebooth@cmu.edu; or Robert S. Siegler, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, 15213; phone, 412-268-2809; FAX, 412-268-2798; email, rs7k@andrew.cmu.edu.
Abstract
We examined whether the quality of first graders’ (mean age 7.2 years) numerical magnitude representations is correlated with, predictive of, and causally related to their arithmetic learning. The children’s pretest numerical magnitude representations were found to be correlated with their pretest arithmetic knowledge and to be predictive of their learning of answers to unfamiliar arithmetic problems. The relation to learning of unfamiliar problems remained after controlling for prior arithmetic knowledge, working memory capacity, and math achievement test scores. Moreover, presenting randomly chosen children with accurate visual representations of the magnitudes of addends and sums improved their learning of the answers to the problems. Thus, representations of numerical magnitude are both correlationally and causally related to arithmetic learning.
Numerical Magnitude Representations Influence Arithmetic Learning

Recent studies have revealed a consistent and robust age-related change in children’s representations of numerical magnitudes. With age and experience, children progress on many tasks from relying on logarithmic representations of numerical magnitudes to relying on linear representations (Booth & Siegler, 2006; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Opfer, Thompson, & DeVries, 2007; Siegler & Booth, 2004; Siegler & Opfer, 2003). The meaning of this finding is illustrated in Figure 1, which displays representative data on number line estimation. When children are presented number lines with endpoints of 0 and 100 and no marks or numbers in between, and then are asked to estimate the positions of specific numbers on the number line, most kindergartners produce estimates that closely fit a logarithmic function. In contrast, most second graders generate a linear pattern of estimates; about half of first graders produce each of the two patterns. As can be seen in the Figure, use of a logarithmic representation entails estimates increasing more rapidly for smaller numbers than for larger ones, whereas use of a linear representation entails estimates increasing at an equal rate throughout the numerical range.

This shift from logarithmic to linear patterns of number line estimates has been found to occur between kindergarten and second grade for 0-100 number lines and between second and fourth grade for 0-1000 ones (Booth & Siegler, 2006; Geary, et al., in press; Laski & Siegler, in press; Siegler & Booth, 2004). The same shift from logarithmic to linear estimation patterns occurs at the same ages on other estimation tasks, in particular measurement estimation (drawing a line of approximately N units, where line lengths of 0 and 1,000 units are present as reference points) and numerosity
estimation (generating approximately N dots on a computer screen, where 1 and 1,000 dots are the reference points) (Booth & Siegler, 2006). A parallel change has been found for kindergartners’, first graders’, and second graders’ categorization of the numbers 0-100 (Laski & Siegler, in press).

The consistency of individual differences in performance on these tasks also attests to the effects of the changes in numerical magnitude representation. Within-grade correlations of individual second and fourth graders’ degree of linearity on number line estimation, measurement estimation, and numerosity estimation tasks ranged from $r = .50$ to $r = .80$ in two experiments (Booth & Siegler, 2006). Within-grade correlations of individual kindergartners’, first graders’, and second graders’ linearity of numerical categorization and linearity of number line estimation ranged from $r = .55$ to $r = .85$ in two other experiments (Laski & Siegler, in press). Within-grade correlations between individual kindergartners’ through fourth graders’ math achievement test scores and their linearity of number line estimation have typically ranged from $r = .50$ to $r = .60$ (Booth & Siegler, 2006; Siegler & Booth, 2004).

The central question of the present study was whether linearity of numerical magnitude representations influences acquisition of new numerical information. This was the first study to address this question. The particular numerical acquisition that was examined was learning answers to unfamiliar arithmetic problems. Though learning arithmetic facts has generally been viewed as a rote memorization process (e.g., Ashcraft, 1992; Campbell & Graham, 1985), we believe that such learning also includes meaningful cognitive activity. In particular, it seems to involve relying not only on verbal associations between problems and answers but also on number sense and an
understanding of how quantities are combined to generate answers with particular
magnitudes (see Jordan et al., 2007, and Baroody, 1985, for similar views). Thus, there
was reason to believe that children’s representations of numerical magnitudes might
influence their arithmetic learning.

Studies of the effects of preschool mathematics curricula provided another reason
to hypothesize that numerical magnitude representations influence learning of answers to
arithmetic problems. Several preschool programs, in particular the Right Start and
Number Worlds curricula (Griffin, 2004; Griffin, Case, & Siegler, 1994), are designed to
inculcate a mental number line representation, akin to the linear representation of
numerical magnitudes hypothesized in the present research. Griffin, Case, and their
colleagues found that having low-income preschoolers participate in a wide range of
numerical activities that were intended to stimulate such a representation -- numerical
board games, computer games, monetary activities, number-related projects, number-
related songs, and so on -- improved the children’s mathematical performance in first
grade.

It seems likely that the extensive experience with numbers provided by Project
Right Start and Number Worlds contributes to well-differentiated knowledge of a wide
range of numerical magnitudes (i.e., knowledge in which students discriminate clearly
among the magnitudes of large as well as small numbers, rather than just among the
magnitudes of small numbers). Well-differentiated knowledge of both large and small
numbers seems likely to constrain the range of potential answers that receive more than
minimal activation when arithmetic problems are presented. Connectionist models posit
that multiple soft constraints limit potential responses to any given problem by
suppressing activations of unlikely responses and boosting activation of more likely ones (e.g., McClelland, Rumelhart, & Hinton, 1986). This leads to most errors being close misses, as well as to increasingly accurate performance, both of which characterize learning of arithmetic (Geary, 2006).

The implications of this analysis for arithmetic learning may be clarified by imagining a first grader being asked whether 24+18 might equal 82. A child who used a linear representation of the magnitudes of numbers between 0 and 100 would quickly answer “no,” because the problem would elicit little if any activation of such a large answer. In contrast, for a child with a logarithmic representation of numerical magnitudes between 0 and 100, 82 would not be such an implausible answer, because the logarithmic representations of 42 and 82 are not hugely different (3.74 and 4.41). Perhaps for that reason, children who generated logarithmic patterns of number line estimates were more likely to categorize 42 and 82 as being within the same numerical category than were children who generated linear estimation patterns (Laski & Siegler, in press). Thus, linear representations of numerical magnitudes should increase the likelihood that errors on relatively large arithmetic problems will be close misses, and should also increase the likelihood of retrieving correct answers.

Errors and solution time patterns on verification tasks are consistent with this analysis and with the interpretation that representations of numerical magnitudes influence arithmetic performance above and beyond knowledge of the correct answers. Both children and adults more quickly reject errors that are far from the correct sum than errors whose magnitudes are closer to the correct value (Ashcraft, Fierman, & Bartolotta, 1984; Campbell & Graham, 1985). For this and the previously described reasons, we
hypothesized that the linearity of representations of numerical magnitudes would influence learning of unfamiliar arithmetic problems.

The Current Study

The present study had three main purposes. One was to examine whether first graders’ representations of the magnitudes of numbers between 0 and 100 are related to their existing arithmetic knowledge and to their ability to acquire new knowledge of arithmetic problems in this numerical range. For the reasons described in the previous section, both relations seemed plausible. First graders seemed the ideal age group for testing this hypothesis, because at least within the U. S., this is the grade in which there are the most substantial individual differences in representations of numerical magnitudes between 0 and 100 (Laski & Siegler, in press; Siegler & Booth, 2004).

A second purpose of the study was to examine the relation to arithmetic learning of several other variables that seemed likely to influence such learning: overall mathematical knowledge (as indexed by scores on standardized achievement tests); prior knowledge of addition; and working memory for numbers. Although many studies of arithmetic knowledge have been conducted, we were unable to locate any that systematically examined relations of prior numerical knowledge to acquisition of new arithmetic knowledge under experimentally controlled conditions that guaranteed equal exposure to the problems. The present study was designed to provide information on this basic question.

In addition to being of interest in itself, examination of these predictive relations also made possible a rigorous test of the hypothesis that numerical magnitude representations influence arithmetic learning above and beyond general mathematical
proficiency. If the linearity of numerical magnitude representations per se is influential, then measures of the linearity of the magnitude representations should predict learning of unfamiliar arithmetic problems above and beyond the influences of standardized math achievement test scores, prior knowledge of addition, and working memory for numbers.

The third major purpose of the study was to test whether provision of visual representations of numerical magnitudes exerts a causal influence on learning of answers to addition problems. External representations, including pictorial, graphic, and diagrammatic forms, are an important part of mathematics education (Seeger, 1998). They are thought to increase understanding of mathematical concepts by helping children build relations among mathematical ideas (Hiebert & Carpenter, 1992). Consistent with this claim, the National Council of Teachers of Mathematics recommends that teachers include multiple forms of representations when teaching mathematical concepts (NCTM, 2000). Evidence supporting this recommendation is rather weak, however. For example, Sowell (1989) concluded that the use of physical manipulatives is beneficial, but (Beishuizen, 1993) found that their usefulness may be limited to children with low levels of achievement. Indeed, in some contexts, provision of manipulatives appears to be harmful rather than helpful to mathematics learning (Uttal, Liu, & DeLoache, 2006). Similarly, findings from some studies suggest that provision of pictorial information is beneficial to math learning (e.g., Beishuizen, 1993; Klein, Beishuizen, & Treffers, 1998), whereas findings from others indicate that many students do not benefit from them (e.g., Cobb, 1995; Gravemeijer, 1994). Which types of external representations are most helpful, and when such representations are helpful, are poorly understood.
To address these issues, the first graders who participated in the present study were randomly assigned to encounter one of several types of external representations of the magnitudes of addends and sums as they attempted to learn unfamiliar addition facts. Children in the *Computer Generate* group were presented 0-100 number lines with computer-generated (correct) representations of the magnitudes of addends and sums for four unfamiliar addition problems as they tried to learn the answers to the problems. Peers in the *Child Generate* group were asked to estimate for themselves on 0-100 number lines the magnitudes of the addends and the sums as they tried to learn the answers to the four trained problems. Children in the *Child and Computer Generate* condition first generated magnitude representations for themselves and then saw the computer-generated, correct representations as they tried to learn the answers to the problems. Finally, classmates in the *Control* group were given identical exposure to the problems and answers but were neither asked to generate a visual representation of them nor presented the accurate magnitude representations.

This design allowed us to test four main hypotheses about arithmetic learning. The first hypothesis was that linearity of numerical magnitude representations on the pretest predicts arithmetic learning. The logic underlying this prediction was that linear magnitude representations would lead children to activate a relatively narrow range of potential answers that included the correct sum, rather than activating either an overly broad range of numbers that did not provide much constraint on plausible answers or a range that did not include the sum. In contrast, children who relied on a logarithmic representation of the numbers between 0 and 100 would activate a broad range of numbers in the numerical range into which answers to the four trained problems fell.
(sums of 27-92), because the logarithmic representation does not discriminate strongly among them (Within a linear representation, the magnitudes of 92 and 27 differ by a ratio of 3.41 : 1, whereas in a logarithmic representation, the magnitudes differ by a ratio of only 1.37 : 1). Depending on how the representations of the addends were implemented in the context of solving arithmetic problems, the activated range might not even include the correct answer. Consider 17 + 29, one of the problems on which children received training. The natural logs of the addends sum to 6.20, whereas the natural log of 46 is 3.83. If children use the natural log values of 17 and 29 to generate an approximate answer, the correct sum might be activated minimally if at all.

The second hypothesis was that the predictive relation between pretest linearity of numerical representations and arithmetic learning would be present above and beyond the relations to arithmetic learning of math achievement test scores, prior knowledge of other addition problems, and working memory for numbers. All of these other variables were expected to be related to arithmetic learning, but the relation of the linearity of each child’s numerical magnitude representation to that child’s arithmetic learning was expected to be present even after the other variables were considered. This prediction was a particularly rigorous test of the hypothesized relation between numerical magnitude representations and arithmetic learning, because math achievement test scores share substantial common variance with the linearity of numerical magnitude representations (Booth & Siegler, 2006; Siegler & Booth, 2004). Moreover, math achievement tests have far better psychometric properties than the 5-minute assessment of children’s numerical magnitude representations that we used. Nonetheless, because math achievement and the other variables were hypothesized to reflect not only magnitude representations but also a
variety of other influences, and because magnitude representations were hypothesized to be particularly important for learning new arithmetic problems, linearity of numerical representations (as measured by linearity of number line estimates) was expected to explain variance in arithmetic learning beyond that which could be accounted for by the other three variables.

The third hypothesis was that children who were given accurate information about numerical magnitudes (those in the Computer Generate and Child and Computer Generate groups) would show more learning of the correct sums, and greater improvements in the distance of their answers from the correct sum, than children in the two groups that did not receive such information. Encountering accurate visual representations of the numerical magnitudes of the addends and sum of each problem would help children align the combined magnitudes of the addends with the magnitude of the corresponding sum. It also would provide a visual processing route to the approximate value of the sum, which would support the verbal association of problem and answer.

The fourth hypothesis was that children who generated their own estimates of the magnitudes of the addends and sum (those in the Child Generate and Computer and Child Generate conditions) would show greater improvement in knowledge of the trained addition problems than children who did not generate visual representations of the magnitudes of addends and sums for themselves. Asking children to generate representations of addends and sums was expected to lead to greater engagement in the task and deeper processing of the numerical magnitudes. Experimental manipulations that improve depth of processing have been found to enhance learning on a variety of
mathematical and scientific problems, including addition (Calin-Jageman & Ratner, 2005), balance scales (Pine & Messer, 2000), and mathematical equality (McNeill, 2007). A similar effect was anticipated in the present context.

Method

Participants

The children who participated, 105 first graders in the second half of their academic year (mean age = 7.20 years, SD = .40), were recruited from four public elementary schools in a predominately middle-income school district. The percentages of children at the schools who were eligible for free or reduced-cost lunches were 14%, 15%, 22%, and 24%, respectively. Almost all participants (96%) were Caucasian, as were both experimenters.

Children were randomly assigned to one of four conditions: Child Generate (N=26); Computer Generate (N=25); Child and Computer Generate (N=27); or Control (N=27). The slightly unequal sample sizes in the four groups were due to two children being absent when the two-week follow-up session was conducted, and one child not completing the posttest number line task due to computer malfunction. Participation was completely voluntary, and no extrinsic rewards were provided.

Tasks

Number Line Estimation

Knowledge of numerical magnitudes was assessed using a computerized number line estimation task. On each problem, at the top of the screen was a number between 1 and 99, and in the middle of the screen was a horizontal number line with 0 at the left end and 100 at the right end. The experimenter explained that the number line always went
from 0 to 100 and that the child should click on the line to show where he or she thought the number at the top should go. Children were first given two practice trials, which involved marking 100 and then 0 on the number line. If children did not place each practice number at the appropriate endpoint, they were shown where the number belonged. After this orienting task, the remaining numbers were presented, one at a time without feedback.

To facilitate our ability to discriminate between linear and logarithmic estimation patterns, we slightly over-sampled the numbers at the low end of the 0-100 range by including four numbers from each decade below 30 and two numbers from each successive decade. The 26 numbers used on the number line task were: 2, 3, 6, 7, 11, 14, 15, 19, 21, 23, 24, 28, 32, 36, 44, 47, 51, 58, 63, 69, 72, 76, 84, 87, 91, and 98. Children received one of three random orderings of the numbers in each session, with the orderings counterbalanced across sessions.

Math Achievement Test

The math section of the Wide Range Achievement Test-Expanded (WRAT-Expanded) was administered to obtain a measure of each child’s general math knowledge. This standardized test includes questions on counting, reading numerals, measurement, spatial reasoning, and oral and written addition and subtraction. It was developed based on the curriculum content guidelines and standards developed by the National Council for Teachers of Mathematics (2000), and has been shown to be reliable in terms of internal consistency and stability (Robertson, 2001).

Working Memory Test

Working memory was measured using the standard forward digit-span task from
the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV) (Wechsler, 2002). The experimenter told children that she would say some numbers and that the child’s job was to repeat the numbers after she stopped. Children were given one practice trial with the string “1, 2”, and were corrected if they did not provide the correct response. On all trials, the presentation rate was 1/s. The experimental task began with two trials of three-digits each. If the child succeeded on at least one trial, the length of the list was increased by one digit after two presentations of that length. If the child failed on both presentations at a given level, the test was terminated. The child’s score was the value of the longest series of digits correctly repeated.

**Addition**

Thirteen addition problems with answers between 0 and 100 were used to examine the first graders’ knowledge of addition: $26 + 27, 1 + 4, 18 + 16, 38 + 39, 5 + 12, 49 + 43, 13 + 9, 48 + 17, 8 + 5, 17 + 29, 5 + 4, 46 + 35, and 9 + 18$. With the exception of the few problems with very small sums, these problems were likely to be unfamiliar to first graders. All of the problems with sums greater than 20 involved carrying; such problems were chosen to prevent children from solving these problems via strategies other than retrieval, and thus to require them to use retrieval. A time limit of 6 s/problem was imposed for the same reason.

The instructions also conveyed the information that children should not count or use algorithms to solve the problems but rather should try to retrieve the answer. The experimenter told children that their job was to say a number that they thought could be the answer, rather than trying to add the numbers in their head, and that all of the answers would be between 0 and 100. The experimenter presented each problem verbally and
asked the child to give the answer quickly. On the rare occasions when a child did not answer quickly (i.e. within six seconds after the experimenter finished reading the problem), the experimenter reminded the child not to calculate the answer, just to say a number that could be the answer if they did not know it. All children received the 13 addition problems in the same random order.

Four of the 13 addition problems were used as training items. Pilot testing indicated that these four problems were ones that the first graders rarely answered correctly on the pretest. The problems included one item that tended to elicit answers that were relatively close to the correct answer (9 + 18), two that elicited answers that tended to be incorrect by a moderate amount (26 + 27 and 17 + 29), and one that tended to elicit answers that were highly discrepant from the correct answer (49 + 43). In each session, after stating their answers for all of the problems, children in all conditions were told that they would be shown the answers to the problems and that they should try really hard to learn the correct answers. During each training session, students encountered four presentations of the first problem, followed by four presentations of the second problem, then the third, and finally the fourth. At the end of each training session, children were given a no-feedback test of the four trained addition facts, in which the experimenter presented each problem verbally and asked the child to give the answer.

*Instructional Procedures*

During the three training sessions, the particular procedures and types of visual information encountered by children varied with their instructional condition. Children in the *Computer Generate* group were shown horizontally oriented 0-100 number lines with red, blue, and purple bars that correctly represented the magnitudes of the addends and
sum. Peers in the Child Generate group generated bars on number lines for themselves to estimate the magnitudes of the addends and sums; therefore, their visual representations could be accurate or inaccurate. Children in the Child and Computer Generate group received both visual representations. Finally, children in the Control group were presented the same problems and answers on the screen but without any visual representation of the numerical magnitudes. Each instructional condition is described in more detail below.

**Computer Generate Condition.** Children were told that they would see the problem and the answer on the computer screen, that their job was to watch as the computer indicated the sizes of the numbers on the number line, and to try really hard to learn the answers to the problems. They were then given a practice problem to orient them to the important features of the task: that the red and blue bars on the computer screen represented how big the addends were, and that the purple bar for the answer equaled the length of the bars for the two addends laid end to end.

The practice problem had “10 + 10” at the top of the computer screen and a line with 0 and 100 at the two ends in the middle of the screen. The experimenter went over this problem twice, in somewhat different ways. First, the experimenter read the problem aloud and asked the child to repeat it. The first “10” was then highlighted in red and circled on the screen, and the child was asked to read the red number. Next, the correct magnitude for “10” was indicated in a three step procedure: First, a small vertical hatch mark appeared on the number line; then, a red bar from 0 to the hatch mark appeared; and then, “10” was displayed above the center of the red bar. The experimenter said, “The red line shows how big 10 is. Notice that the line is red, just like the number in the problem.
The first number is always going to be red.” The red bar and the number above it then disappeared, and the procedure was repeated for the second “10” (which was represented with a circled blue number and a blue line). Finally, the sum, which was represented with a circled purple number and a purple bar, was presented.

At the beginning of the second time through the practice problem, the computer screen was reset to its original state. This time, the child read the first “10”, which was again highlighted in red, and the screen displayed its magnitude with a red bar. The experimenter reminded the child that the red bar showed how big the first addend was. Then the child read the second “10,” and the screen displayed its magnitude with a blue bar under the number line, with the first addend and the red bar representing its magnitude remaining visible. The experimenter reminded the child that the blue bar indicated how big the second addend was, and said “Let’s add the red line for 10 to the blue line for 10.” At this point, the blue bar, which represented the second addend, was moved to the top of the number line, beginning where the red bar ended. Then the purple bar that represented the magnitude of the answer appeared just under the number line, and the experimenter explained that the purple bar indicated how big 20 was and that the child should notice that the length of the red bar plus the length of the blue bar equaled the length of the purple bar. Finally, with this display remaining visible, the child was asked to read the problem one final time.

After the practice problem, children were tested on their knowledge of what the red, blue, and purple bars meant. The experimenter asked the same three questions until the child answered all of them correctly. Then, the four training problems were presented, along with the correct visual representation of each one’s addends and sum.
Child Generate Condition. Children in this group were given the same instructions except that they were told that they (rather than the computer), would estimate the magnitudes of the numbers on the number line. The remainder of the procedure was identical to that in the Computer Generate condition, except for the children estimating the magnitudes of the addends and sum on the number line. Red, blue, and purple bars corresponding to the child’s estimates of the addends and sum for each problem (instead of the computer generated estimates) appeared on the screen as in the Computer Generate condition.

Child and Computer Generate Condition. Children in this group were given the same initial instructions and procedure as peers in the Child Generate group. However, the second time through the problem, the screen also showed computer-generated (correct) placements on a separate number line that was positioned directly underneath the child-generated estimates for the problem. For example, for the problem 17 + 29, after the child’s estimate of the magnitude of “17” appeared just above the upper number line, the computer generated the correct magnitude of 17 just above the lower number line. The experimenter then said “This red line (pointing) is how big you thought 17 was, but actually this red line (pointing) is how big 17 is.” Then, after the child’s estimate of the magnitude of 29 appeared in blue just below the upper number line, the correct magnitude of 29 appeared in blue just below the lower number line. Next, on both number lines, the blue bar representing the magnitude of the second addend was placed at the right end of the red bar that represented the magnitude of the first addend. After that, a purple bar representing the sum of the two addends appeared below the relevant number line. On the upper number line, the purple bar corresponded to the child’s estimates of the
sum, whereas on the lower number line, the purple bar corresponded to the correct sum. Thus, by the end of the trial, the computer screen displayed all three of the child’s estimates on the upper number line and all of the correct magnitudes on the lower line.

*Control Condition.* As in the other three conditions, children in the control group encountered each problem four times in each training session. The conditions under which children were presented each problem were also the same as in the other conditions. First, the experimenter read the addends and sum aloud. Then, the child repeated them. The next two times through the problem, the addends and sum were colored red, blue, and purple and circled as in the other conditions, and the child read them without the experimenter’s help.

*Procedure*

Children met one-on-one with the experimenter for three 10-15 minute sessions within a one-week period and for a 5-minute “follow-up” session two weeks after the third session. All sessions were conducted during the school day; children participated at a time selected by their teacher. An outline of the tasks presented in each of the four sessions is presented in Table 1.

In Session 1, children were presented the number line estimation task, the working memory task, and the 13-problem addition task. Then they were presented the instructional procedure for their experimental group, followed by a no-feedback test on the four trained addition problems at the end of the session. All children received the tasks in the same order. Session 2 consisted of a pretest, training, and a posttest on the four addition training problems, followed by the math section of the WRAT. Session 3 was identical to Session 2, except that children were given the number line estimation
task rather than the WRAT at the end of the session. During Session 4, (the follow-up session), children were again asked to provide sums for the same 13 addition problems as on the pretest.

No feedback regarding answers to the trained problems was given on the pretests and posttests for those problems, and only implicit feedback (i.e., the correct answer) was given during training. Children were told frequently during all parts of the experiment that they were doing well. At the end of each session, they were thanked for participating and returned to their classroom.

Results

The results are reported in two sections. In the first section, we examine for the entire sample the interrelations of performance on the four pretest tasks – addition, number line estimation, working memory span, and math achievement test – and examine each of their predictive relation to learning of the four trained addition problems. In the second section, we examine the causal influence of computer-generated and self-generated visual representations on the learning of the trained arithmetic problems and on number line estimates. The analyses examined performance at three times: pretest (Session 1), end of training (end of Session 3), and follow-up (Session 4).

Predictors of Learning and Performance

Relations Among Pretest Measures

As shown in Table 2, levels of performance on all four pretest measures of numerical proficiency were positively related. The relations among number line estimation, addition, and math achievement test scores were substantial; the relations between working memory span and the other three tasks were weaker. The correlation
between linearity of number line estimates and overall math achievement test score -- \( r = 0.57 \) -- was of the same magnitude as in previous studies (\( r \)'s = 0.50-.60).

**Pretest Predictors of Arithmetic Learning**

We examined pretest predictors of performance on the trained addition problems at the end of training and at the follow-up session for the whole sample and also for each instructional condition separately. The predictive relations were similar for all conditions; therefore, only the results for the full sample are reported.

**Addition Knowledge.** Children answered 21% of pretest items correctly. The large majority of correct answers on the pretest were on the four problems that had sums below 20 --55%, vs. 6% on the nine larger problems. On the four large addend problems on which children subsequently received training, they answered correctly on 7% of trials.

Knowledge of addition was also assessed through examining percent absolute error (PAE):

\[
\text{PAE} = \left( \frac{\text{Child's Answer} - \text{Correct Answer}}{\text{Scale of Answers}} \right)
\]

For example, if a child answered “80” for the problem 26 + 27, the PAE would be .27 \([(80 - 53)/100]\), or 27%. PAE provides a continuous measure of the degree to which answers are of the right magnitude, which was a central issue in the present study.

Mean PAE on the pretest was 12%. Again, there was a large discrepancy between performance on the four small addend problems (PAE = 2%) and on the nine large addend ones (PAE = 17%). On the four training items, pretest PAE was 16%.

We next examined the stability of individual differences in addition knowledge, in particular whether children’s pretest addition knowledge was predictive of their posttest knowledge of the trained addition problems. To avoid the auto-correlations that would
have been entailed by including the same items as predictors and predicted variables, we correlated each child’s mean pretest PAE on the nine non-trained addition problems with that child’s performance following training on the four trained addition problems. (Too few children answered any of these four problems correctly on the pretest for pretest number correct to be a meaningful predictor.)

PAE on the nine untrained problems on the pretest was positively related to PAE on the four trained addition problems at the end of training \((r(103) = .47, p < .01)\) and also on the follow-up \((r(101) = .62, p < .01)\). PAE on the nine untrained problems on the pretest also was correlated with percent correct answers on the four trained problems both at the end of training \((r(103) = -.23, p < .05)\) and on the follow-up \((r(101) = -.37, p < .01)\).

We next conducted a regression analysis to determine whether pretest PAE on the nine untrained addition problems accounted for variance in learning of the four training set items, beyond that which could be explained by pretest PAE on the training set items themselves. In this analysis, we first entered each child’s pretest PAE on the four training set items and then their pretest PAE on the nine untrained addition problems.

As shown in Table 3, pretest PAE on the nine untrained addition problems added significant variance at both the end of training (8%) and on the follow-up (11%) beyond that which could be explained by pretest PAE on the trained problems. This finding suggests that children’s general sense of the likely magnitudes of answers to addition problems predicts their subsequent learning, beyond their initial knowledge of answers to the specific problems that they are learning.
Number Line Estimation. To assess the relation between pretest numerical magnitude representations and addition learning, we correlated each child’s $R^2_{Lin}$ on the pretest number line estimation task with that child’s percent correct answers and PAE on the trained addition facts at the end of training and on the follow-up.

$R^2_{Lin}$ on the number line pretest predicted PAE on the trained addition facts at both the end of training ($r(103) = -.45, p < .01$) and at the follow-up ($r(101) = -.54, p < .01$). Pretest $R^2_{Lin}$ on the number line pretest also predicted percent correct answers on the trained addition problems at the end of training ($r(103) = .23, p < .05$) and at the follow-up ($r(101) = .30, p < .01$). Thus, initial knowledge of numerical magnitudes predicted both learning of unfamiliar addition facts and quality of errors on them.

To determine whether pretest number line estimation was related to improvement in percent absolute error on the addition facts, above and beyond the quality of initial answers on the problems, we conducted a regression analysis in which we first entered each child’s pretest PAE on the trained addition facts and then their $R^2_{Lin}$ values on the number line pretest. As shown in Table 3, $R^2_{Lin}$ on the number line pretest added a substantial amount of variance to that which could be explained by pretest PAE on the trained addition facts: 8% at the end of training and 11% at the two-week follow-up.

Math Achievement. Children’s math achievement, as measured at the time of the pretest, also was predictive of their PAE on the trained addition problems both at the end of training ($r(103) = -.49, p < .01$) and on the follow-up ($r(101) = -.52, p < .01$). Their achievement test scores also were related to percent correct answers on the trained problems at the end of training ($r(103) = .28, p < .01$) and on the follow-up ($r(101) = .33, p < .01$). As shown in Table 3, the math achievement test scores, like pretest performance
on the untrained addition problems and number line estimates, added to the variance that could be explained by pretest PAE on the trained addition facts: 10% additional variance at the end of training and 6% at the follow-up.

*Working Memory.* Working memory span was predictive of PAE on the trained addition facts both at the end of training ($r(103) = -.20, p < .05$) and at the follow-up ($r(101) = -.23, p < .05$). However, it was not related to percent correct on the trained addition problems either at the end of training ($r(103) = .11, ns$) or at the follow-up ($r(101) = .08, ns$). It also did not account for significant variance in improvement in PAE at the end of training or at the follow-up, beyond that which could be explained by pretest PAE on the four trained addition problems (Table 3).

*Linear Representations as an Independent Predictor of Addition Learning.* One of the main purposes of the study was to test the prediction that even after the effects of other aspects of numerical knowledge were considered, linearity of numerical magnitude representations would still predict arithmetic learning. Therefore, we conducted regression analyses on reduction of error on the trained addition facts by the end of training and on the follow-up. We first entered each child’s pretest PAE on the trained addition facts, pretest PAE on untrained addition facts, pretest working memory span, and math achievement test scores. Then we entered the child’s linearity of pretest number line estimates, to see if it explained additional variance in the child’s post-training performance on the trained addition problems.

The linearity of pretest number line estimates added significant variance at both the end of training and on the follow-up to that which could be explained by the other variables. Pretest performance on the other four tasks explained 45% of the variance in
PAE on the trained addition problems at the end of training; pretest $R^2_{lin}$ on the number line explained an additional 2% ($t(102) = 2.09, p < .05, d = .41$). The same four pretest variables explained 33% of the variance in PAE on the trained problems on the follow-up; pretest $R^2_{lin}$ on the number line estimation task explained an additional 6% ($t(100) = 2.88, p < .01, d = .58$). Thus, despite $R^2_{lin}$ on the number line task being correlated with three of the four other predictors, and all of those other measures being correlated with PAE on the trained addition problems at the end of training and at the follow-up, linearity of number line estimates accounted for variance in arithmetic learning that could not be explained by the other measures of numerical knowledge.

Effects of Instructional Condition

To examine the effects on arithmetic learning of the experimental manipulations, we contrasted performance in the four experimental conditions. Comparisons were conducted on both number of correct answers and PAE.

Percent Correct on Trained Addition Problems

A 2 (Computer generated magnitude representations: Present or absent) X 2 (Child generated magnitude representations: Present or absent) X 2 (Time of test: Pretest or end of training) ANOVA on the number of correct answers was conducted to determine whether the visual representations of numerical magnitude influenced children’s learning up to the end of training. The analysis yielded main effects for the time of test variable, $F(1,101) = 70.14, p < .01, \eta^2 = .41$, and for exposure to computer-generated visual representations, $F(1,101) = 4.08, p < .05, \eta^2 = .04$. Children generated more correct answers at the end of training than on the pretest (29% vs. 7%), and children who encountered the accurate, computer-generated representations of numerical
magnitudes generated a greater number of correct answers than did children who did not
encounter them (21% vs. 15%).

The interaction among the three variables also was significant, $F(1,101) = 5.01, p < .05$, $\eta^2 = .05$. To interpret this interaction, we first conducted separate ANOVAs for
pretest and end of training performance. No differences among the four groups were
apparent on the pretest ($F(3,101) = 1.02, ns, \eta^2 = .03$), but differences were present at
the end of training, $F(3,101) = 2.62, p = .05$, $\eta^2 = .07$. LSD Post Hoc tests indicated that
at the end of training, children in the Computer Generate condition advanced a greater
number of correct answers than peers in both the Control and the Computer and Child
Generate conditions ($p$’s < .05, $d$’s = .75 and .61, respectively). These findings indicated
that the interaction was due to the Computer Generate procedure increasing learning
when presented alone but not when presented in combination with the Child Generate
procedure.

Analyses of pretest-posttest changes for each of the four groups supported this
interpretation. As shown in the top half of Table 4, children in all four groups answered
correctly more often at the end of training than on the pretest: Control ($t(26) = 3.43, p <
.01, d = .88$); Child Generate $t(25) = 4.41, p < .01, d = 1.22$; Computer Generate ($t(25)
= 6.53, p < .01, d = 1.51$); and Child and Computer Generate ($t(27) =2.67, p < .01, d =
.65$). However, the changes were larger in the Computer Generate condition (32%
difference between pretest and posttest percent correct) than in the other three conditions
(15%, 17%, and 24%).

A parallel 2 X 2 X 2 ANOVA on changes in correct answers between pretest and
follow-up on the training problems indicated only a single main effect for time of testing,
Children advanced a greater number of correct answers on the follow-up than on the pretest (13% vs. 7%).

**Percent Absolute Error on Trained Addition Problems**

To determine whether encountering accurate visual representations of the magnitudes of the addends and sum influenced the quality of students’ errors as well as the frequency of correct answers, we conducted a 2 (Computer generated magnitude representations: Present or absent) X 2 (Child generated magnitude representations: Present or absent) X 2 (Time of testing: Pretest or end of training) ANOVA on PAE on the addition problems. As on the parallel analysis of correct answers, the ANOVA revealed a main effect for time of testing, \(F(1,101) = 6.59, p < .01, \eta^2 = .06\). From the pretest to the end of training, PAE on the addition problems dropped from 16% to 13%.

Also as on the parallel analysis of changes in correct answers, the ANOVA for the PAE data revealed an interaction among the three variables, \(F(1,101) = 4.36, p < .05, \eta^2 = .04\). The source of the interaction appeared to be the same as in the analysis of correct answers. Again, no differences among the four groups were detected on the pretest \(F(3,101) = 0.62, \text{ns}, \eta^2 = .02\), but the groups tended to differ at the end of training, \(F(3,101) = 2.12, p = .10, \eta^2 = .06\). LSD Post Hoc tests indicated that children in the *Computer Generate* condition had a lower posttest PAE than peers in the *Control* condition \((p < .05, d = .66)\), and tended to have a lower posttest PAE than children in the *Computer and Child Generate* condition \((p < .10, d = .63)\).

Analyses of pretest-posttest changes for each of the four groups supported this interpretation (bottom half of Table 4.) Children in the *Computer Generate* group reduced their PAE on the trained addition problems between the pretest and the end of
training \((t(24) = 2.13, p < .05, d = .60)\). There was also a trend toward reduced error for children in the *Child Generate* group between the pretest and the end of training \((t(25) = 1.88, p < .10, d = .47)\). In contrast, PAE did not decrease in either the *Child and Computer Generate* group \((t(26) = 0.32, ns, d = .11)\) or in the *Control* group \((t(26) = 0.84, ns, d = .18)\).

A parallel 2 X 2 X 2 ANOVA on changes in PAE between pretest and follow-up on the training problems yielded no significant main effects or interactions.

**Numerical Magnitude Estimation**

Exposure to visual representations of numerical magnitudes was expected to improve children’s number line estimates as well, due to the training providing experience with number line representations of numerical magnitudes. Separate 2 (Computer generated magnitude representations: Present or absent) X 2 (Child generated magnitude representations: Present or absent) X 2 (Time of testing: Pretest or end of training) ANOVAs were performed on percent absolute error (PAE) and linearity \(R^2_{\text{Lin}}\) of number line estimation.

Main effects of viewing computer-generated visual representations were found on PAE, \((F(1,100) = 7.23, p < .01, \text{eta}^2 = .07)\). The number line estimates of children who encountered accurate, computer-generated representations had a lower PAE than did peers who did not encounter the accurate visual representations (9% vs. 11%).

In addition to this main effect, an interaction between exposure to computer-generated representations and test session was present on the PAE measure \((F(1,100) = 4.87, p < .05, \text{eta}^2 = .05)\). The number line estimates of children who viewed the computer-generated (correct) visual representations improved from a PAE of 10% on the
pretest to 9% at the end of training. In contrast, the number line estimates of peers who did not see correct visual representations of numerical magnitude failed to improve: PAE increased from 11% to 12%.

Discussion

The results of this study supported three of the four hypotheses that motivated it. As hypothesized, the linearity of children’s number line estimates correlated positively with their existing knowledge of addition. Also as hypothesized, the degree of linearity of children’s pretest estimates was predictive of their learning of answers to novel addition problems. Moreover, providing accurate visual representations of the magnitudes of addends and sums increased children’s learning of the novel addition problems beyond the level produced by simply presenting problems and answers. The one result that was contrary to expectation was that having children generate their own representations of addends and sums did not enhance learning and may have interfered with it. In this concluding section, we examine implications of these findings for understanding the role of numerical magnitude representations in childhood numerical competence, for understanding arithmetic learning, and for improving mathematics instruction.

Numerical Representations and Numerical Competence

The pretest results both replicated and extended previous findings that the linearity of elementary school children’s representations of numerical magnitudes is related to their performance on a wide range of numerical tasks. Prior studies indicated that the linearity of number line estimates correlates positively with the linearity of measurement estimation, numerosity estimation, numerical categorization, and numerical magnitude comparison, as well as with math achievement test scores (Booth & Siegler,
A very recent study added the findings that preschoolers’ estimates on a 0-10 number line correlate positively and quite strongly with their counting, numeral identification, and magnitude comparison skills (Ramani & Siegler, in press).

The present study replicated the relation of number line estimation to standardized achievement scores in those studies and indicated that first graders’ linearity of numerical magnitude representations also is related to their arithmetic proficiency and working memory span. Two other very recent studies have linked the linearity of number line estimation to yet other aspects of cognition: several measures of executive functioning and IQ test scores (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Geary, Hoard, Nugent, & Byrd-Craven, 2007). Geary and his colleagues advanced the intriguing hypothesis that the strong relations of executive functioning to linearity of number line estimates reflect the need for children to inhibit the previously dominant logarithmic representation when they begin to rely on the linear representation. Geary et al.’s research also demonstrated that the number line estimates of first and second graders diagnosed as having math learning disabilities lag far behind those of typically achieving peers in both linearity and accuracy.

Taken together, the present and previous findings indicate that preschoolers’ through fourth graders’ numerical magnitude representations are positively correlated with performance on specific tasks measuring a wide range of specific numerical competencies, as well as with performance on overall tests of mathematical achievement and IQ. The linearity of the numerical magnitude representations also shows robust and stable patterns of individual differences and consistent developmental trends (Booth &
Siegler, 2006; Geary et al, 2007). Thus, as hypothesized by Case and colleagues (Case, 1992; Case & Okamoto, 1996; Griffin, Case, & Sandieson, 1992; Marini, 1992), the mental number line appears to provide a central conceptual structure for organizing a wide range of numerical knowledge.

**Arithmetic Learning**

The present findings indicated that numerical magnitude representations are not only positively related to a variety of types of numerical knowledge; they also are predictive of success in acquiring new numerical information, in particular answers to arithmetic problems. Prior studies (e.g., Geary, et al., 2007; Jordan, et al., 2006) had shown that incoming characteristics of students predict their learning of arithmetic and other numerical skills during the school year. However, the present study appears to be the first to demonstrate that such predictive relations are present when learning occurs under experimentally controlled conditions.

An important implication of the present findings is that arithmetic learning, even in the sense of memorizing answers to unfamiliar problems, is not a rote activity but rather a meaningful one. If memorization of answers to arithmetic problem were a rote activity, there would be no reason to expect that being presented accurate visual representations of the magnitudes of addends and sums would increase learning of correct answers. Nor would there be any reason to expect that pretest linearity of number line estimation would predict learning of novel arithmetic problems, even after prior knowledge of arithmetic, working memory span, and overall math achievement test scores were statistically controlled. These results cannot be explained in terms of exposure to the addends and answers, which were the same in all four conditions, nor in
terms of time on task, which was greater for children generating their own representations of numerical magnitudes than for children who encountered the computer generated, accurate ones.

Several previous investigators have argued that backup strategies, such as counting fingers, counting from the minimum addend, and repeated addition, make arithmetic a meaningful activity (Brownell & Chazal, 1935; Cowan, 2003; Cowan, Dowker, Christakis, & Bailey, 1996; Cowan & Renton, 1996). The present findings add evidence that linear representations of numerical magnitudes make arithmetic meaningful even when children learn answers via retrieval.

Although the linearity of numerical magnitude representations on the pretest was related to improvements on the trained arithmetic problems in both percent correct and percent absolute error, the relations to reductions in PAE were stronger, accounting for more than twice as much variance both at the end of training and on the follow-up. This finding was consistent with the theory underlying the experiment. The most straightforward way in which numerical magnitude representations would influence arithmetic learning is by constraining the set of possible answers to ones of approximately the right magnitude. Such constraints would have the direct effect of improving the quality of errors; they also would increase the likelihood of the exact correct answer, because by definition, it has exactly the right magnitude.

Linearity of numerical magnitude representations was not the only type of numerical knowledge that predicted arithmetic learning. Mathematical achievement test scores and addition knowledge also did. Yet linearity of magnitude representations was
predictive of arithmetic learning even after the relations of the other variables to arithmetic learning were statistically controlled. The question was why.

Our interpretation is that linearity of number line estimates varies more closely with the underlying variable of linearity of numerical magnitude representations than do the other two variables. Addition knowledge and achievement test scores vary in part with linearity of numerical magnitude representations, but they also vary with other influences (e.g., counting skills, memory for computational procedures, working memory capacity). Thus, linearity of number line estimation may predict learning of novel arithmetic problems, above and beyond the prediction of the other variables, because it provides a purer measure of the underlying magnitude representation.

The present findings ruled out the possibility that the predictive relation of linearity of number line estimation to arithmetic learning simply reflected general familiarity with numbers or the richness of the home intellectual environment. If either of these interpretations were correct, there would be no reason to expect that linearity of number line estimates would remain a significant predictor of learning of novel arithmetic problems even when achievement test scores and prior addition knowledge were statistically controlled. Indeed, there was reason to expect the opposite. Some parents teach their young children to solve arithmetic problems, whereas others do not. First graders’ knowledge of addition would be expected to reflect this parental input. In contrast, to the best of our knowledge, no parents (not even the present authors) are so pedantic that they ask their children to locate numbers on number lines. Thus, the fact that the linearity of number line estimation predicts arithmetic learning above and beyond
prior arithmetic knowledge indicates that learning of new arithmetic problems is influenced by numerical magnitude representations.

The one factor that, to our surprise, did not predict learning of addition facts was working memory. One possible explanation is that the training in this study allowed children to view the addends and the answers simultaneously as they were trying to learn them; thus, the task format may have reduced the usual burden on working memory of learning answers to arithmetic problems. Another possible explanation was that the digit span task was an inadequate measure of working memory capacity; a different measure of working memory might have revealed stronger relations.

Instructional Implications

Results of the experimental manipulations yielded causal evidence that converged with the correlational data. As hypothesized, exposure to accurate visual representations of numerical magnitude improved children’s arithmetic learning. This was true both when learning was measured in terms of percent correct and when it was measured in terms of percent absolute error. Thus, the patterns of correlations on the pretest, the predictive relations from pretest performance to performance at the end of training, and the effects of providing randomly chosen children with accurate linear representations of numerical magnitude converged on the conclusion that linear representations of numerical magnitude promote arithmetic learning.

Self-generated representations of numerical magnitudes did not have a comparable effect. In fact, they appeared to interfere with the positive effects of being shown accurate visual displays when children received both. Thus, it appears that pictorial representations of numerical magnitudes must be accurate to enhance learning.
The pattern of results dovetails with previous findings indicating that being asked to explain correct answers improves learning of mathematical and scientific information, but being asked to explain the reasoning underlying ones’ own answers, right or wrong, does not enhance learning (Siegler, 1995; 2002).

Perhaps the most surprising finding of the study was that the combination of computer-generated and self-generated magnitude representations was less effective in promoting learning than the computer-generated representations alone. One plausible interpretation was that the first graders in this condition simply could not take in all of the information when both representations were present. Seeing both visual representations on the screen may have provided a difficult source monitoring challenge, in that it required remembering that the computer generated representations were the accurate ones and that the self-generated ones were not necessarily accurate. Source monitoring is generally a difficult task for young children (Ratner, Foley, & Gimpert, 2002). The challenge might have been especially great in the current context, because the correct interpretation required children to discount their own magnitude estimates.

The pattern of effects and non-effects in the present study highlight the value of presenting accurate visual representations of mathematical relations. Contemporary mathematics education for elementary school children makes a great deal of use of concrete objects that can be manipulated, such as Dienes Blocks and Cuisenaire Rods. Recent research, however, suggests that children often have difficulty viewing physical objects purely as symbols, even when that is the teacher’s intent. This is especially the case when children handle or play with the objects (DeLoache & Burns, 1994; Uttal, Liu, & DeLoache, 2006).
The findings and analyses of DeLoache, Uttal, and their colleagues, along with the present findings, provide a basis for hypothesizing that accurate pictorial representations, such as diagrams and graphs, may be particularly useful for promoting math learning. Like manipulatives, they provide a visual representation of the meaning of mathematical operations. Unlike manipulatives, they are not tempting objects for play and other activities that may interfere with their connection to mathematical operations. This analysis, together with the present finding that accurate visual representations of numerical magnitude promote arithmetic learning, suggests that providing pictorial information that illustrates mathematical relations in a transparent way can improve math learning as early as first grade.
References


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* (Nos. 1-2).

Research in Mathematics Education, 26*, 362-385.

Cowan, R. (2003). Does it all add up? Changes in children’s knowledge of addition
combinations, strategies, and principles. In A. J. Baroody & A. Dowker (Eds.), *The
development of arithmetic concepts and skills: Constructing adaptive expertise* (pp.

assessing children’s understanding of the order-irrelevance principle. *Journal of

Cowan, R., & Renton, M. (1996). Do they know what they are doing? Children’s use of
economical addition strategies and knowledge of commutativity. *Educational
Psychology, 16*, 407-420.


Lerner (Series Eds.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child
Hoboken, NJ: Wiley.


distributed processing. In D. E. Rumelhart, J. L. McClelland, and the PDP Research
Group, Parallel distributed processing. Explorations in the microstructure of

development in children’s performance on equivalence problems. In J. H. Bisanz
(Chair), Overcoming misconceptions: Mechanisms of positive change for a common
mathematical misunderstanding. Symposium conducted at the biennial meeting of the
Society for Research in Child Development, Boston, MA.


of fractional magnitude than adults. In McNamara, D. S., & Trafton, G. (Eds.),
Mahwah, NJ: Erlabum.

Pine, K. J., & Messer, D. J. (2000). The effect of explaining another’s actions on

Ramani, G. B., & Siegler, R. S. (in press). Promoting broad and stable improvements in
low-income children’s numerical knowledge through playing number board games.
Child Development.

Ratner, H. H., Foley, M. A., & Gimpert, N. (2002). The role of collaborative planning in
children’s source-monitoring errors and learning. Journal of Experimental
Psychology, 81, 44-73.


Table 1

*Measures Included in Each of the Four Experimental Sessions*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Session 1</th>
<th>Session 2</th>
<th>Session 3</th>
<th>Session 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Line Estimation</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Math Achievement Test</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working Memory Span</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-Item Addition Test</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Item Addition Pretest, Training, and Posttest</td>
<td>X&lt;sup&gt;1&lt;/sup&gt;</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

<sup>1</sup> In Session 1, the 4-item addition pretest reflected performance on the 4 items on which children subsequently received training from within the 13-item addition test.
Table 2

Correlations Among Number Line Estimation, Addition, Working Memory Span, and math Achievement Test Performance: Pretest Data

<table>
<thead>
<tr>
<th>Measure</th>
<th>Number Line Estimation ($R^2_{Lin}$)</th>
<th>Math Achievement (WRAT)</th>
<th>Addition (PAE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math Achievement (WRAT)</td>
<td>.57**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addition (PAE)</td>
<td>.41**</td>
<td>.48**</td>
<td></td>
</tr>
<tr>
<td>Working Memory Span</td>
<td>.07</td>
<td>.21*</td>
<td>.21*</td>
</tr>
</tbody>
</table>

*Notes: df = 103, ** $p < .01$, * $p < .05$*
Table 3

*Pretest Predictors of Improvement in PAE on Trained Addition Problems in End of Training and Follow-up Performance*

<table>
<thead>
<tr>
<th>Pretest Predictor (s)</th>
<th>% Variance Explained in Reduction of PAE on Trained Problems at End of Training</th>
<th>% Variance Explained in Reduction of PAE on Trained Problems at Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trained Addition Problems (PAE)</td>
<td>31</td>
<td>18</td>
</tr>
<tr>
<td>Trained Addition Problems (PAE) + Untrained Addition Problems (PAE)</td>
<td>39*</td>
<td>29*</td>
</tr>
<tr>
<td>Trained Addition Problems (PAE) + Number Line Estimation ($R^2_{Lin}$)</td>
<td>39*</td>
<td>29*</td>
</tr>
<tr>
<td>Trained Addition Problems (PAE) + WRAT Score</td>
<td>41*</td>
<td>24*</td>
</tr>
</tbody>
</table>

1 First row of data indicates variance in improvement in PAE on trained addition problems after training and on the follow-up that could be explained by pretest PAE on those problems. Subsequent rows indicate the variance in improvement in PAE at the end of training and on the follow-up on the trained addition problems that could be explained by considering pretest PAE on the same problems and also the additional variable indicated in the left column. Thus, considering pretest number line estimation ($R^2_{Lin}$) as well as pretest PAE on the trained addition problems accounted for 8% more variance after training (39% vs. 31%) and 11% more variance on the follow-up (29% vs. 18%) than only considering pretest PAE on the same problems (*p < .01).
<table>
<thead>
<tr>
<th>Trained Addition Problems (PAE) + WM</th>
<th>Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>20</td>
</tr>
</tbody>
</table>
Table 4

Percent Correct and PAE for Trained Addition Problems on Pretest, End of Training, and Follow-up Sessions

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>End of Training</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Percent Correct</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Control</em></td>
<td>6 (13)</td>
<td>23 (24)</td>
<td>10 (16)</td>
</tr>
<tr>
<td><em>Child Generate</em></td>
<td>4 (12)</td>
<td>28 (25)</td>
<td>13 (20)</td>
</tr>
<tr>
<td><em>Computer Generate</em></td>
<td>9 (18)</td>
<td>41 (24)</td>
<td>13 (22)</td>
</tr>
<tr>
<td><em>Child and Computer Generate</em></td>
<td>10 (17)</td>
<td>25 (28)</td>
<td>14 (9)</td>
</tr>
<tr>
<td><strong>PAE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Control</em></td>
<td>18 (10)</td>
<td>16 (12)</td>
<td>18 (11)</td>
</tr>
<tr>
<td><em>Child Generate</em></td>
<td>17 (12)</td>
<td>12 (9)</td>
<td>16 (13)</td>
</tr>
<tr>
<td><em>Computer Generate</em></td>
<td>15 (11)</td>
<td>9 (9)</td>
<td>13 (10)</td>
</tr>
<tr>
<td><em>Child and Computer Generate</em></td>
<td>14 (9)</td>
<td>15 (10)</td>
<td>14 (9)</td>
</tr>
</tbody>
</table>

*Note:* All statistics are reported as Mean (SD)
Figure Captions

*Figure 1.* Median number line estimates and best fitting function in Siegler & Booth (2004).
(Figure 1)

Kindergarten  
First Grade  
Second Grade