Journal of Cognition and Development

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

Grounding the Symbols for Place Value: Evidence From Training and Long-Term Exposure to Base-10 Models

Kelly S. Mix, Linda B. Smith, Jerri DaSha Stockton, Yi-Ling Cheng & Justin A. Barterian

To cite this article: Kelly S. Mix, Linda B. Smith, Jerri DaSha Stockton, Yi-Ling Cheng & Justin A. Barterian (2017) Grounding the Symbols for Place Value: Evidence From Training and Long-Term Exposure to Base-10 Models, Journal of Cognition and Development, 18:1, 129-151, DOI: 10.1080/15248372.2016.1180296

To link to this article: http://dx.doi.org/10.1080/15248372.2016.1180296

Accepted author version posted online: 12 May 2016.
Published online: 12 May 2016.

JOURNAL OF COGNITION AND DEVELOPMENT, 18(1):129–151

ISSN: 1524-8372 print/1532-7647 online
DOI: 10.1080/15248372.2016.1180296

Routledge
Taylor & Francis Group

Grounding the Symbols for Place Value: Evidence From Training and Long-Term Exposure to Base-10 Models

Kelly S. Mix
University of Maryland

Linda B. Smith
Indiana University

Jerri DaSha Stockton and Yi-Ling Cheng
University of Maryland

Justin A. Barterian
The Ohio State University

Two experiments examined whether concrete models support place value learning. In Experiment 1 ($N = 149$), 7-year-olds were trained with either a) symbols alone or b) symbols and base-10 blocks. Children in both groups showed significant growth overall, but there were specific effects favoring one training type over another. Symbols-only training led to higher scores on a number line estimation task and was particularly effective among high-ability students, whereas blocks training led to better understanding of base-10 structure and was particularly effective among low-ability learners. In Experiment 2 ($N = 68$), Montessori students, for whom concrete models play a major role in mathematics instruction, also demonstrated better understanding of base-10 structure than did their matched peers enrolled in mainstream elementary schools.

The ability to link symbols to their referents is at the core of human cognition, yet the processes by which we make these linkages, particularly in childhood, remain poorly understood. Symbol grounding is complicated because symbols usually bear no obvious relation to their referents. Instead, their meaning derives from associations constructed between the symbol system and the perceptual world. In the realm of mathematics, these associations are particularly intricate, but we know very little about their precise nature, how they are formed, or what experiences with symbol systems are critical to their formation.

The present study examined these processes in young children learning the place value system, both with and without support from concrete models. From a symbol-grounding perspective, place value is

Correspondence should be sent to Kelly S. Mix, University of Maryland, 3304 Benjamin Bldg., College Park, MD 20742, USA. E-mail: kmix@umd.edu
Kelly S. Mix, Jerri DaSha Stockton, and Yi-Ling Cheng are formerly affiliated with Michigan State University.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hjcd.

of interest because it involves a complex symbol system with remote ties to its referents. The place value system represents large numbers using spatial position and the multiplicative relation between base-10 units and their counts (e.g., "429" stands for [4 × 100] + [2 × 10] + [9 × 1]). This results in a powerful representational system that converts otherwise intractable quantities into symbols we can read, write, compare, and combine with ease. Acquiring place value is a watershed in mathematical development. Children cannot progress very far without it, and those who struggle with place value in the early grades tend to face lower mathematics achievement throughout elementary school and beyond (Ho & Cheng, 1997; Moeller, Pixner, Zuber, Kaufmann, & Nuerk, 2011). Yet despite evidence of early-emerging, partial knowledge of place value (Byrge, Smith, & Mix, 2014; Mix, Prather, Smith, & Stockton, 2014), complete mastery eludes many children. Faulty place value concepts and rote, error-prone multidigit calculation are common and persistent problems (Cauley, 1988; Cobb & Wheatley, 1988; Fuson & Briars, 1990; Jesson, 1983; Kamii, 1986; Kouba et al., 1988; Labinowicz, 1985; Resnick & Omanson, 1987). For these reasons, place value has been targeted as a high priority for math education starting in kindergarten (National Council of Teachers of Mathematics, 2000; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010).

The source of these difficulties—and the most promising target for addressing them—may be a failure of symbol grounding. Symbol-to-referent mappings can fail when there is either a) an ambiguous correlational structure (Uttal, O'Doherty, Newland, Hand, & DeLoache, 2009) or b) low overall similarity between elements (Bassok & Medin, 1997; Bassok, Wu, & Olseth, 1995; Gentner & Markman, 1994; Gick & Holyoak, 1983). Both conditions hold for place value. First, the correlational structure for numbers is weak. The meaning of "4," for example, includes not only its whole number "count" meaning, but also its ordinal sense, as it appears in dates and addresses, a measurement sense, a time-telling sense, its part-whole meaning in rational numbers, and a purely arbitrary sense as it appears in phone numbers (Fuson, 1988). In a multidigit numeral, these correlations are even weaker because the same digits can have different meanings based on relative position (e.g., 14 vs. 41). Second, there is low surface similarity between the written symbols for large quantities and their referents. The written numeral "42" has no perceptual similarity linking it to a pile of 42 rocks. Without counting the rocks, there is literally nothing connecting the two, and even with counting, there are multiple layers of symbolic meaning to connect, including the counting sequence up to 42, the spoken number name "forty-two," and the written numeral.

Recognizing the problems children have with grounding multidigit numerals, teachers often provide concrete referents, such as base-10 blocks (see Figure 1). Base-10 blocks consist of small cubes to represent ones, sticks with 10 cubes to represent tens, flats made up of 10 sticks to represent hundreds, and so forth. Thus, the size and shape of each unit corresponds analogically to the increase in magnitude (i.e., 10 of the tens sticks lined up side by side are literally the same in size and shape as one of the hundreds flats). Children can use these materials to compare and transform quantities by moving and aligning sets of blocks. They also can map the written symbols and verbal names onto these physical models.

There are several reasons to expect base-10 blocks to be beneficial. First, they physically instantiate place value relations and thus provide referents that can be directly experienced. Both classic theories of cognitive development (e.g., Bruner, Oliver, & Greenfield, 1966; Piaget, 1951) and current views on symbol grounding and embodied cognition (e.g., Barsalou, 2008; Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004; Lakoff & Nunez, 2001) hold that direct perceptual experience is critical to interpreting abstract concepts. Second, base-10 blocks address both of the obstacles to symbol

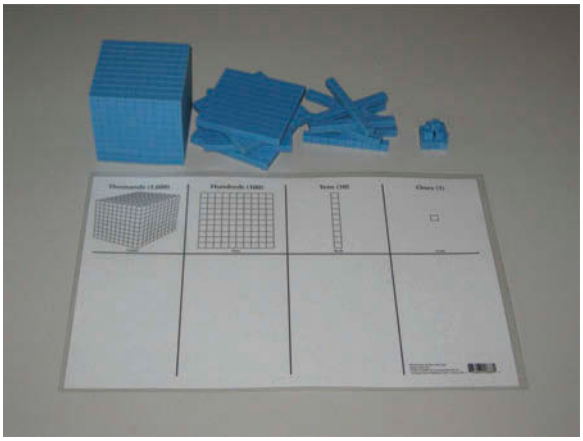

FIGURE 1. Training materials used in the blocks condition, Experiment 1.

grounding identified earlier. They have a strong, predictable internal structure, and they align well with written and spoken place value symbols.

It is puzzling, then, that empirical research on the effectiveness of concrete models has been mixed. Some studies have shown improved mathematical performance for children taught using concrete models and manipulatives (see Carbonneau, Marley, & Selig, 2013, for a recent meta-analysis), including base-10 blocks (Fuson & Briars, 1990; Peterson, Mercer, & O'Shea, 1988). Yet, other researchers have reported either no effect of concrete models, effects that fail to transfer, or even performance decrements (Ball, 1992; Goldstone & Sakamoto, 2003; Kaminski & Sloutsky, 2009; McNeil, Uttal, Jarvin, & Sternberg, 2009; Mix et al., 2014; Son, Smith, & Goldstone, 2011; Uttal, Amaya, Maita, Hand, et al., 2013; Vance & Kieren, 1971). It has been argued that concrete models are detrimental because they themselves are symbolic; they introduce extraneous, distracting details; and they lead to entrenched, context-specific learning (Goldstone & Sakamoto, 2003; Kaminski, Sloutsky, & Heckler, 2008; McNeil et al., 2009; Uttal, Scudder, & DeLoache, 1997).

What explains these discrepant findings? On one hand, there is reason to question whether recent research has given concrete models a fair test. The studies reporting poor outcomes offered either very brief exposure to concrete models or no training at all. In these studies, the models were mostly illustrative. In contrast, the studies reporting an advantage for concrete models usually involve many weeks of training (e.g., Fuson & Briars, 1990; Miller & Stigler, 1991; Reimer & Moyer, 2005). If concrete models impact learning by providing metaphors for symbol grounding (e.g., Lakoff & Nunez, 2001), it makes sense this mapping process could take time and effects would not be immediate. Another difference is that research demonstrating positive effects of concrete models tends to involve hands-on materials children manipulate. In contrast, some of the recent work showing a performance decrement varies the amount of rich detail across pictorial representations, rather than providing objects to move around (e.g., Kaminski & Sloutsky, 2009). Based on current symbol-grounding theory, direct contact and movement could be crucial (Glenberg et al., 2004; Lakoff & Nunez, 2001).

On the other hand, research claiming an advantage for concrete models may be overly optimistic due to weak controls. In many cases, the critical comparison has been from pretest to posttest, without comparison to an alternative treatment control group (e.g., Reimer &

Moyer, 2005). This approach has been used in several key studies showing an advantage of base-10 blocks in place value learning (Fuson, 1986; Fuson & Briars, 1990). The problem with this approach is that what might seem to be an effect of concrete models could actually be an effect of training in general, background developmental change, or even test–retest effects. Even in studies with comparisons to abstract symbol instruction, it is not always clear to what extent the instructional content was precisely parallel across the two conditions (vs. roughly targeting the same basic concept; see Carbonneau et al., 2013, for a review). A stronger test of concrete models would be to compare training with them to the same instruction using symbols alone.

Finally, it is possible the reported effects (or lack thereof) are related to individual differences in background knowledge or learning ability—differences that were not reported in previous research with children and thus may have been overlooked. For example, research with adults has suggested that high-ability learners actually perform worse with concrete models than with instruction based on abstract concepts. In contrast, novice or low-ability learners sometimes show an advantage from instruction with concrete models (Goldstone & Sakamoto, 2003; Kalyuga, Ayres, Chandler, & Sweller, 2003). If the same holds true for school-aged children, research could sometimes fail to show an effect of concrete models in samples composed of mostly high-ability children.

The present study examined whether and under what conditions concrete models for place value impact learning. The reported experiments combined assurances of adequate training and comparisons to training without concrete models to provide a rigorous but fair test of their potential effects. We also included a range of ability levels and probed specifically for performance differences across these groups.

EXPERIMENT 1

Method

Participants. The total sample was composed of 149 children with a mean age of 7;2 (range = 6;0–9;0). A power analysis indicated that a sample size of 75 children would be sufficient to detect a difference between conditions (f = .42) at the .90 level (Faul, Erdfelder, Buchner, & Lang, 2009). An additional 29 children were recruited but excluded because their pretest scores were greater than 90% correct.

Children were randomly assigned to either the blocks training group (n = 52; M_{age} = 7;2, range = 6;0–8;7; 23 boys) or the symbols-only training group (n = 49; M = 7;2, range = 6;0–8;0; 31 boys). Random assignment was made for individual children within each class or childcare group and within each school or camp program. A no-training control group (n = 48; M = 7;0, range = 6;0–9;0; 26 boys) was recruited from the same schools and classrooms during the following academic year.

The children in all three conditions attended a public elementary school in one of six school districts that served the same middle-socioeconomic status (SES) population (86% White, 6% Black, 5% Hispanic or Latino; median income = $54,087). Teachers and afterschool care providers distributed consent forms to parents, and only children whose parents returned the signed consent forms were included. Data were collected in the spring of first grade (n = 33), the

fall of second grade (n = 65), or during the intervening summer via day camp programs (n = 51). All participating school districts had adopted *Everyday Mathematics* (McGraw-Hill Education) as their mathematics curriculum. Within this curriculum, place value and multidigit calculation are taught in first and second grade, so it is reasonable to assume children in the study had received some exposure to these topics in school. In this regard, the training provided in our experiment could be viewed as supplemental. However, as the results will show, children did not perform near ceiling on our posttests, so there remained much room for improvement. Also, we could not completely eliminate the possibility of concurrent exposure without focusing on a younger age group, for whom the subject matter may be too advanced.

Methods and Procedure. Children in the blocks condition used individual sets of base-10 blocks (15 ones, 15 tens, 15 hundreds, and 2 thousands blocks) during the training sessions. They also used mats that showed how written numerals, place value names, and blocks aligned (see Figure 1). Children in the symbols-only condition completed the same lessons and activities as the blocks training group but without concrete models. In this group, the problems and activities were completed either in writing or using a set of plain white note cards with hand-written single-digit numerals on them (range = 0–9). Children in the no-training group completed the pretests and posttests at a 4-week interval without any intervening training.

The training sessions took place over 4 to 6 weeks and focused on six content lessons that introduced children to multidigit number meanings and calculation. (See the Appendix for a detailed description.) Learning was assessed with three written measures: a place value test, the school sale problem, and a number line estimation task. Number line estimation was administered individually, but the other two tests were completed in small groups (n = 4). All 149 children completed the place value test, and most, but not all, completed the other two tests (number line, n = 113; school sale, n = 111) because these measures were added later. As noted, a sample size of 75 would be sufficient to detect a medium effect (f = .42) at the .90 level.

The place value test consisted of 12 to 16 items distributed among three item types: a) numeral ordering (6 items); b) numeral interpretation (3 to 7 items[1]); and c) multidigit addition (3 items). For numeral ordering, children a) saw a pair of three- or four-digit numerals and indicated which of the pair was either larger or smaller (e.g., 567 vs. 439, 2,523 vs. 2,851; 4 items); b) ordered three three-digit numerals, either by arranging cards with the numerals written on them or indicating which of four lists had the correct smaller-to-larger ordering (e.g., 135, 153, 315; 1 item); or c) saw two three-digit numbers with a gap in between and indicated which of four three-digit numerals would fit into the series (1 item). For numeral interpretation, children either a) identified which of four multidigit numerals was named by the experimenter (2 items); or b) identified which of four numerals had a specific place value meaning (e.g., "Which number has a 7 in the hundreds place?"; 2 items). Most children (n = 83) also identified which of four expanded notations matched a multidigit numeral (e.g., 152 = 100 + 50 + 2; 4 items). Multidigit addition items were vertically oriented two- and three-digit problems that required carrying. Children responded by writing their solutions in the test booklet (3 items). Scores on each item type were converted to percent correct so as to equate for differences in the total number of items presented. Children's actual scores on the place value test

[1] We combined the data from two training conditions in this experiment. Children in one condition received more numeral interpretation items than children in the other condition. See the Appendix for details.

ranged from 0% to 100% correct. The test–retest reliability for this measure, based on scores in the no-training condition, was .85.

The school sale problem (Bednarz & Janvier, 1982) assessed children's understanding of the hierarchical relations underlying place value. For example, children were told, "Pretend you are packing things in bags and boxes to sell at school. First, you are going to pack erasers. You have 38 erasers. You can fit 5 erasers in each bag. You can fit 5 bags in each box. How many full boxes can you make with the 38 erasers you have?" Note that this problem is structured to reflect the same kind of nested relations represented by base-10 notation and base-10 blocks (i.e., ones, tens, and hundreds). We presented three versions of the problem that varied in terms of the objects (candies or erasers), order of the problem (start with total quantity and divide into units or start with units and colligate into the total), and the numbers involved. Each child completed all three items in a random order. They were given a paper and pencil and told they could write or draw anything that would help them solve the problem. Scores were the total number correct out of 3 possible points and ranged from 0 to 3 correct. The test–retest reliability in the no-training group was low (.22), most likely due to widespread floor performance; however, 68% of children in the no-training group had the same scores at pretest and posttest.

In each trial of the number line estimation task (Siegler & Booth, 2004), children were shown a horizontal 0-to-1,000 number line with the anchors labeled and a stimulus number printed above the center of the line in a circle. The stimulus numbers ranged from "2" to "983." Children were asked to mark the number line with a pencil to indicate where the stimulus number would be placed (22 total trials). In previous research, Siegler and colleagues have found that kindergarten and first-grade students initially bunched magnitudes together such that the best fit for their responses was a logarithmic function. However, by second grade, most children correctly place different magnitudes along the line, resulting in a linear function being the best fit. This improvement has been measured using either model fit (linear vs. logarithmic) or percent absolute error (PAE; Booth & Siegler, 2006). We used PAE in our analyses because this approach yielded a continuous distribution of accuracy (range = .04–.51). The test–retest reliability on this task was .69.

Results

Our analyses focused on three main questions. First, to evaluate whether either training was effective, paired-samples *t* tests were conducted on the pretest and posttest scores for each condition. Second, we compared the effectiveness of one training condition to the other and to the control using analyses of covariance (ANCOVAs) that controlled for pretest differences. Third, to examine whether children of different prior ability benefitted differently from training, we conducted additional ANCOVAs with samples that were split at the median. All post-hoc comparisons were corrected using the Bonferroni method.

Place Value Test. We first confirmed that children's place value test scores did not differ across conditions at pretest using a one-way analysis of variance (ANOVA; M_{blocks} = 42%, $M_{\text{symbols-only}}$ = 46%, $M_{\text{no-training}}$ = 42%), $F(2, 146) = 0.80$, $MSE = .04$, $p = .45$, $\eta^2_p = .01$. Next, we compared children's pretest and posttest scores and found significant improvement in both training conditions (posttest scores, M_{blocks} = 57%, $t[51] = 6.37$, $p < .0001$, Cohen's $d = 0.88$; $M_{\text{symbols-only}}$ = 62%, $t[48] = 5.40$, $p < .0001$, Cohen's $d = 0.77$), but not in the no-training control

group ($M_{\text{no-training}}$ = 44%, t[47] = 1.05, p = .15, Cohen's d = 0.15). The same pattern emerged in the ANCOVA. Specifically, there was a main effect of condition, F(2, 145) = 10.55, MSE = .03, p < .0001, η^2_p = .13, such that the two training groups had higher posttest scores than the no-training control group (blocks vs. no training, M_{diff} = 13%, p = .001; symbols-only vs. no training, M_{diff} = 15%, p < .0001) but did not differ from each other (M_{diff} = 2%, p = 1.00).

Although neither of the training conditions seemed more effective than the other for the place value test as a whole, t tests comparing the pretests and posttests for each item type suggested a different pattern (see Table 1). Children in the blocks condition showed significant improvement on numeral interpretation items, t(51) = 3.34, p = .001, Cohen's d = 0.46, whereas children in the symbols-only condition did not, t(48) = 0.98, p = .16, Cohen's d = 0.14. In contrast, children in both groups showed improvement on both numeral ordering (blocks, t[51] = 2.79, p = .004, Cohen's d = 0.39; symbols only, t[48] = 1.85, p = .04, Cohen's d = 0.26) and multidigit addition (blocks, t[51] = 4.06, p < .0001, Cohen's d = 0.56; symbols only, t[48] = 6.04, p < .0001, Cohen's d = 0.86). Unlike numeral interpretation, these latter two item types were close to the material and tasks covered in the training (see the Appendix). Thus, both groups improved on outcome measures that were similar to the training tasks, but only blocks training led to significant gains on numeral interpretation items, which were both novel and most clearly focused on base-10 structure. In the no-training group, no pretest to posttest comparisons reached significance (lowest p = .11, one-tailed, Cohen's d = 0.18).

Recall that previous work has shown differential responses to concrete models depending on overall ability. To probe for such differences, we divided children into ability groups based on their pretest scores (median = 44% correct; low ability, pretest = 0%–43%, n = 72; high ability, pretest = 44%–100%, n = 77), and we used this grouping as a between-subjects factor in an ANCOVA with training condition (blocks, symbols only, no training) as the other between-subjects factor, pretest scores as the covariate, and posttest scores as the dependent measure.[2] There was a small but significant interaction between ability and condition, F(2, 142) = 3.57, MSE = .03, p = .03, η^2_p = .05 (see Figure 2). Post-hoc comparisons showed that for the high-ability children, only symbols-only training led to greater gains than no training (M_{diff} = 16%, p = .004). For high-ability children in the blocks condition, this difference was not significant (M_{diff} = 7%, p = .51). In contrast, for children with low ability, there was a large blocks training advantage in comparison with no training (M_{diff} = 19%, p < .0001), but this was not so following symbols-only training (M_{diff} = 11%, p = .12). These performance differences are consistent with previous work showing an advantage for abstract symbols among other higher-ability learners (e.g., undergraduates learning a difficult science concept; Goldstone & Sakamoto, 2003) but also suggest that concrete models may be particularly important for children who are struggling with mathematical symbol meanings. Still, it should be noted that performance in the two training

[2] The conditions were not represented equally in these ability groups because the median pretest score for the symbols-only group (median = 50%) was higher than that for the blocks group (median = 41%) and no training (median = 31%). This resulted in relatively more symbols-only children in the high-ability group versus the other conditions and relatively fewer symbols-only children in the low-ability group (symbols only, high ability, n = 30, low ability, n = 19; blocks, high ability, n = 26, low ability, n = 26; no training, high ability, n = 21, low ability, n = 27). Because such an imbalance might lead to spurious effects due to increased power in one ability group and decreased power in the other, we repeated the analyses after shifting the 5 symbols-only children who performed directly on the median (44%) from the high- to the low-ability group. This resulted in even ability group sizes (high ability, n = 25; low ability, n = 24) and yielded the same pattern of findings reported here.

TABLE 1
Mean performance by training condition and item type, Experiment 1

	Numeral Ordering		*Numeral Interpretation*		*Multidigit Addition*		*School Sale*		*Number Line Estimation PAE*	
Condition	*Pretest*	*Posttest*	*Pretest*	*Posttest*	*Pretest*	*Posttest*	*Pretest*	*Posttest*	*Pretest*	*Posttest*
Symbols only	.71	.76*	.62	.67	.22	.54*	.19	.29	.21	.18*
	(.19)	(.20)	(.34)	(.36)	(.25)	(.39)	(.25)	(.39)	(.10)	(.08)
Blocks	.64	.73*	.45	.58*	.19	.42*	.18	.41*	.19	.18
	(.24)	(.24)	(.30)	(.33)	(.25)	(.36)	(.24)	(.43)	(.09)	(.08)
No training (Control)	.60	.63	.35	.39	.23	.18	.10	.14	.30	.29
	(.22)	(.24)	(.31)	(.35)	(.38)	(.34)	(.20)	(.25)	(.11)	(.11)

Note. For each cell, we report the mean and standard deviation. Significant gains from pretest to posttest based on one-tailed t tests are indicated with an asterisk ($p < .05$). For *number line estimation*, the means are reported as percent absolute error (PAE) instead of percent correct, so the asterisk indicates a significant decrease.

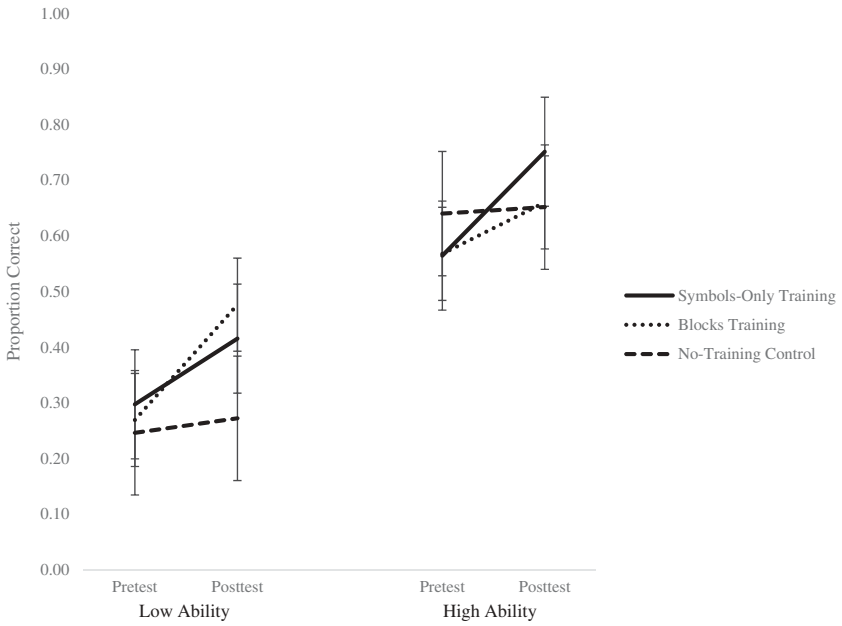

FIGURE 2. Mean changes from pretest to posttest, by condition and ability group, Experiment 1.

conditions did not differ for either ability group (high ability, M_{diff} = 9%, p = .12; low ability, M_{diff} = 8%, p = .40), suggesting that both types of training improved performance in both groups somewhat.

School Sale Problem. Recall that the school sale problem asked children to solve a challenging word problem using nested hierarchies parallel to those used in multidigit numerals (i.e., ones, tens, hundreds). Two of the three problems were worded so that children started with a total number of candies or erasers and had to report how many units resulted after packing. The third problem used the reverse order, such that children started with the total number of boxes and leftover candies/erasers and had to calculate the total. Because performance on this problem was at floor across conditions, even following training, our analyses focused on only the first two problems.

Children received a score of 0, 1, or 2, based on the number of these two problems they solved correctly. Six children performed at ceiling on the pretest (i.e., obtained a score of 2 correct) and were excluded from further analysis. The mean percent correct for the remaining children is presented in Table 1. We first used a one-way ANOVA to check the equivalence of the pretest scores across training groups and found them to be comparable (M_{blocks} = 18%, SD = 24%; $M_{symbols\text{-}only}$ = 19%, SD = 25%; $M_{no\text{-}training}$ = 10%, SD = 20%), $F(2, 108) = 1.79$, MSE = .05, p = .17, η^2_p = .03. Comparisons of children's pretest and posttest averages revealed significant improvement following blocks training ($M_{posttest}$ = 41%), $t(37) = 4.02$, $p < .0001$, Cohen's d = 0.65. Neither of the other pretest–posttest comparisons reached significance, though there was marginal improvement in the symbols-only group, $M_{posttest}$ = 29%, $t(28) = 1.65$, p = .06, Cohen's d = 0.31; no training, $M_{posttest}$ = 14%, $t(43) = 0.72$,

p = .25, Cohen's d = 0.11. The superiority of blocks training also was evident in an ANCOVA using children's posttest scores as the dependent variable, pretest scores as the covariate, and training condition as the between-subjects factor.[3] Specifically, there was a significant main effect of condition, $F(2, 107) = 4.42$, $MSE = .10$, $p = .01$, $\eta^2_p = .08$, that was due to better performance in the blocks group versus the no-training condition ($M_{diff} = 22\%$, $p = .01$); no significant difference was obtained between symbols-only training and no training ($M_{diff} = 10\%$, $p = .66$). Neither did the blocks group significantly outperform the symbols-only group in a direct comparison ($M_{diff} = 12\%$, $p = .42$). However, the finding of a blocks advantage in comparison to the no-training group is consistent with their performance on the place value test. That is, children in the blocks group demonstrated better understanding of both the school sale problem and the numeral interpretation items from the place value test—measures that specifically tap understanding of the underlying structure of multidigit numerals.

To examine whether performance varied by ability, we divided children into two groups based on a median split of the place value pretest scores. We could not use the school sale problem pretest scores because two thirds of the children had scores equal to 0. Also, using the place value pretest scores had the advantage of keeping the ability groups constant across analyses. The resulting ANCOVA, with ability group and the three training conditions as between-subjects factors, indicated that children did not benefit differentially from one training condition or the other, $F(2, 104) = 1.14$, $MSE = .11$, $p = .32$, $\eta^2_p = .02$.

Number Line Estimation. Recall that children were asked to mark the correct locations of multidigit numerals on a 0-to-1,000 number line. In each trial, we measured the distance between children's actual responses and the correct placement for each numeral and used the PAE averaged across trials as the dependent variable in all analyses. We first examined children's pretest performance and found that a condition difference existed prior to training, $F(2, 110) = 14.38$, $MSE = .01$, $p < .001$, $\eta^2_p = .21$. Children in both training conditions ($M_{blocks} = 19\%$, $SD = 9\%$; $M_{symbols\text{-}only} = 21\%$, $SD = 10\%$) had lower error rates than children in the no-training control ($M_{no\text{-}training} = 30\%$, $SD = 11\%$). During testing, we monitored performance on the place value test and school sale problem to ensure roughly equivalent performance across groups, but because the number line estimation task is more time-consuming to code, we were not aware of this discrepancy until after testing was completed. However, though not ideal, this outcome was less concerning than pretest differences between the training groups would have been.

In comparisons of the pretest scores to posttest scores, only symbols-only training yielded significant improvement ($M_{symbols\text{-}only} = 18\%$, $SD = 8\%$, $t[33] = 2.07$, $p = .02$, Cohen's $d = 0.36$; $M_{blocks} = 18\%$, $SD = 8\%$, $t[35] = 0.17$, $p = .43$, Cohen's $d = 0.03$; $M_{no\text{-}training} = 29\%$, $SD = 11\%$, $t[42] = 0.90$, $p = .19$, Cohen's $d = 0.14$). Because extreme pretest differences can bias the results using an ANCOVA (Allison, 1990; Oakes & Feldman, 2001), we compared performance in the three conditions using an ANOVA on the gain scores, with condition as the between-subjects variable. This analysis yielded no significant effects, $F(2, 110) = 0.75$, $MSE = .01$, $p = .47$, $\eta_p^2 = .01$. Nor was there a significant interaction when ability was added as a second between-subjects

[3] Because scores on the school sale problem have a narrow range (0–2), we repeated the reported analyses using nonparametric ANOVAs (i.e., Kruskal-Wallis tests). One disadvantage to this approach is that it does not permit covariates, such as the pretest scores in Experiment 1 or the Peabody Picture Vocabulary Test-4 scores in Experiment 2, to be included. Still, the results of these analyses were the same as those reported here and in Experiment 2 using ANCOVAs.

factor, $F(2, 107) = 1.01$, $MSE = .01$, $p = .37$, $\eta_p^2 = .02$. Thus, there was only weak evidence that children improved on this task after training, and only in the symbols-only group.

Discussion

In this experiment, we provided instruction on place value operations, either with or without the aid of base-10 blocks. Overall, children improved in both training conditions on numeral ordering and multidigit addition, suggesting that either symbol training or symbol training with concrete models is beneficial. This result challenges the notion that concrete models interfere with learning because they are distracting and difficult to interpret (Goldstone & Sakamoto, 2003; Kaminski et al., 2009; McNeil et al., 2009; Uttal et al., 1997). Despite the general improvement, however, there were several indications that exposure to symbols and exposure to concrete models conferred unique advantages.

First, children who received blocks training demonstrated an advantage on measures that required understanding of base-10 structure—namely, numeral interpretation items and the school sale problem. Perhaps because blocks training aligned more closely with this structure and exposed it more fully, it served to support better performance on these measures. One might wonder why, then, the same advantages were not evident on other place value tasks, such as number ordering and multidigit addition. One reason may be that knowledge of base-10 structure is helpful, but not necessary, in these tasks. Numbers may be ordered using a rough approximation of magnitude without decomposition. Similarly, children can apply the procedures for multidigit calculation by rote. Thus, it is possible to show improvement on these tasks without necessarily understanding base-10 structure. Indeed, research has suggested that the base-10 structure of tasks like numeral ordering or multidigit calculation is not immediately apparent to learners and actually requires deliberate scaffolding to be recognized (Kurtz, Miao, & Gentner, 2001; Richland & Hansen, 2013).

Second, only symbols-only children improved on number line estimation, suggesting that exposure to written symbols may be sufficient to induce their meaning. This was a small effect, and children in both groups were far from ceiling performance. However, the small difference evident in relatively brief training regimes could be meaningful in accounts of how children learn about the spoken and written representations of large numbers: Specifically, understanding how numerical representations are ordered on a number line may depend mostly on experience with numerical symbols. This result is consistent with previous work showing that young children begin to induce the meanings of multidigit numerals from the statistical patterns available from everyday exposure (e.g., two-digit numerals stand for smaller quantities than three-digit numerals; Byrge et al., 2014; Mix et al., 2014). On this view, one reason children in the blocks condition failed to show the same effect may be that they did not receive the same amount of exposure to written symbols as symbols-only group children, or it may be because the cognitive resources needed to understand the blocks limited what they could learn from the symbols themselves. This may be one sense in which concrete models were a detriment to learning, though this difference did not lead to general interference and rather, may have contributed to a specific pattern of strengths and weaknesses.

We also found different responses to training based on initial ability. High-ability children had greater gains from symbols-only training than they did from training with concrete models. This finding hints that base-10 blocks were a hindrance to children in the high-ability group,

consistent with previous research using other models (Goldstone & Sakamoto, 2003; Kalyuga et al., 2003). Perhaps once learners have achieved some understanding of symbols, they find these a more efficient medium for mathematical thought than base-10 blocks, and in this situation, base-10 blocks have distracting or superfluous features, as others have argued (e.g., Kaminski et al., 2008; McNeil et al., 2009). This may be why low-ability children showed the opposite pattern—exhibiting greater gains with blocks training—because they lack this competence and need scaffolding from concrete models to comprehend multidigit numerals at all.

The differential effects of the two training conditions, though small, remind us that understanding place value is multifaceted and there may not be a single answer to the question of whether concrete models help. Instead, the answer may depend on the specific knowledge component being measured and the ability level of the child. Also, small effects are perhaps not surprising given the relatively brief exposure to training children received in this experiment. If children gradually internalize concrete models, as theories of embodied cognition predict, then they could require months or even years of exposure for a clear advantage of concrete models in addition to symbol training to emerge.

EXPERIMENT 2

In Experiment 1, we found significant effects of blocks training, but these effects were limited. Although there were some large pretest versus posttest differences and significant contrasts with children in the no-training group, there were not significant contrasts with children in the symbols-only group. Perhaps these findings were limited because 4 to 6 weeks is too short of an exposure period. Although it would be difficult to carry out a controlled training experiment over months or years, there are naturally occurring educational situations that result in different exposure to concrete models. For place value concepts, the Montessori mathematics curriculum offers a useful test case because it centers on repeated, direct contact with concrete models. Indeed, a defining feature of Montessori education is its consistent, long-term use of sensorial materials, action-based problem solving, and careful scaffolding of written symbols to these materials and actions (Lillard, 2012). Children begin to work with base-10 beads, for example, starting in the prekindergarten years and continue well into the elementary years (Lillard, 1997). If children benefit from extensive exposure to concrete models for place value concepts, this benefit is likely to appear for Montessori students.

This experiment is also of interest because previous research on the effects of Montessori schooling has been mixed. Some studies have shown no performance differences for children in Montessori or traditional school programs (e.g., Lopata, Wallace, & Finn, 2005). However, several carefully controlled comparisons have demonstrated an advantage of Montessori education (e.g., Lillard, 2012; Lillard & Else-Quest, 2006). For example, Angeline Lillard (2012) found that 3- to 6-year-olds in classic, high-fidelity Montessori programs showed larger gains over the academic year in a range of subject areas, including mathematics, when compared with children in either supplemental Montessori or non-Montessori programs. In older children, these effects have not always been obtained (e.g., Lillard & Else-Quest, 2006), but sometimes they have. For example, children who received Montessori education from preschool through fifth grade went on to have significant advantages in high school mathematics and science versus their matched classmates without prior Montessori schooling (Dohrmann, Nishida, Gartner,

Lipsky, & Grimm, 2007). Interestingly, this study did not reveal similar advantages for Montessori students in English or overall grade point average, as one might expect if general curricular differences (e.g., greater classroom autonomy) were driving across-the-board performance differences. However, because this study assessed a broad range of skills with limited depth, it is hard to know what particular aspects of the mathematics and science curriculum might have accounted for the difference. In the present study, we probed for specific performance differences that are linked to the salient differences in Montessori curriculum content and instructional approach (i.e., performance on place value tests in particular). Perhaps this targeted approach will yield more clear-cut effects than previous studies that cast a broader net.

Method

Participants. The sample consisted of 68 children. Half the sample (n = 34; 13 boys) was composed of children who had been continuously attending a Montessori school from the age of 3 years. The remaining children (n = 34; 14 boys) had attended a non-Montessori preschool and at the time of the study were attending one of three elementary schools (two public and one private) from the same community. Children were divided into two age groups (kindergarten, M_{age} = 5;6, range = 4;8–6;3, n = 36; and second grade, M_{age} = 7;3, range = 6;0–8;0, n = 32). The number of Montessori and non-Montessori students within each age group was equal. An a priori power analysis indicated that a sample of 62 children would be sufficient to reveal a medium effect (f = .42) at 90% power (Faul et al., 2009).

To guard against possible population differences between the two school groups, we matched children using Peabody Picture Vocabulary Test-4 (PPVT-4) scores. Vocabulary was chosen because it is known to correlate with SES, parent input, and IQ (Bornstein, Haynes, & Painter, 1998; Hart & Risley, 1995; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Pan, Rowe, Singer, & Snow, 2005). Of the total sample, 38 children were matched exactly (i.e., the same stanine), and the remaining 30 children were matched within one stanine. The difference in the raw PPVT-4 scores for the two groups was not statistically significant ($M_{Montessori}$ = 128.85, $M_{non\text{-}Montessori}$ = 120.18), $t(66) = 1.43$, p = .16, two-tailed, Cohen's d = 0.35, but as an added precaution, we controlled for vocabulary differences in our statistical analyses (see Results section). An additional 58 children were tested but excluded due to lack of match.

Data were collected during winter and spring, so children had nearly completed each grade at the time of testing. Thus, the two grade-level groups (kindergarten and second grade) allowed us to compare children who had received either 3 or 5 years of exposure to either Montessori or non-Montessori schooling. As in Experiment 1, the three non-Montessori elementary schools had adopted *Everyday Mathematics* (McGraw-Hill Education) as their mathematics curriculum.

Fidelity to the Montessori method varies across schools, and this variation has complicated interpretations in previous work (Lillard, 2012). Although we did not measure fidelity objectively, we selected Montessori schools with a strong local reputation for adhering to Montessori methods. All three were private schools, and each had been in operation for at least 25 years. One school was accredited by the Association Montessori Internationale (AMI), and all three employed teachers with advanced training (e.g., master's degrees) from AMI- and American Montessori Society (AMS)-accredited programs.

Materials and Procedures. Learning was assessed using the same measures as in Experiment 1. However, Experiment 2 was not an intervention study, so the measures were given only once rather than used as a pretest and posttest. The school sale problem and number line estimation tasks were exactly the same as in Experiment 1. The place value test was in the same format, with the same three item types as before, but items were added to ensure a sufficiently challenging range of difficulty. For example, whereas the previous multidigit calculation test included only two- and three-digit problems, the revised test included problems up to six digits. Also, whereas only addition items were analyzed in Experiment 1, both addition and subtraction items were included in Experiment 2 (note that the results of Experiment 2 were the same whether or not subtraction items were included). As a result, the place value test was also longer (52 items total vs. 16 items total in Experiment 1).

Results

Place Value Test. Children's mean performance is presented in Table 2. To compare the groups, we first carried out an ANCOVA with grade (kindergarten vs. second grade) and curriculum (Montessori vs. non-Montessori) as between-subjects factors, children's PPVT-4 (vocabulary) raw scores as the covariate, and their percent correct on the place value test as the dependent variable. Not surprisingly, there was a significant effect of grade such that second-grade students outperformed the kindergarteners, $F(1, 63) = 24.93$, $MSE = .01$, $p < .0001$, $\eta_p^2 = .28$. But there also was a significant main effect of school type, $F(1, 63) = 16.76$, $MSE = .01$, $p < .0001$, $\eta_p^2 = .21$, that favored the Montessori students (26% vs. 14%). This finding indicated that even with vocabulary scores matched and controlled, children who had received a Montessori education performed better overall. There also was a significant interaction between school type and grade, $F(1, 63) = 21.86$, $MSE = .01$, $p < .0001$, $\eta_p^2 = .26$, due to equal performance in the two school groups in kindergarten ($M_{diff} = 1\%$, $p = .49$) but significantly better performance in the Montessori group versus the non-Montessori group in second grade ($M_{diff} = 24\%$, $p < .0001$).

We next carried out separate ANCOVAs for each item type, with school type as the between-subjects variable and PPVT-4 (vocabulary) scores as the covariate. We analyzed only the second-grade test scores as there was no evidence in the previous analysis for a school-type difference in kindergarten. The analyses revealed a large advantage for Montessori students on multidigit calculation, $F(1, 29) = 20.23$, $MSE = .03$, $p < .0001$, $\eta_p^2 = .41$, and a moderate advantage on numeral interpretation, $F(1, 29) = 3.13$, $MSE = .07$, $p = .09$, $\eta_p^2 = .10$. Recall that in Experiment 1, children who received blocks training showed a similar advantage on numeral interpretation problems. Interestingly, performance on numeral ordering did not differ across the two groups, $F(1, 29) = 0.27$, $MSE = .05$, $p = .61$, $\eta_p^2 = .009$, just as we found for the blocks and symbols-only groups in Experiment 1.

School Sale Problem. As in Experiment 1, none of the children responded correctly on the third question for which the candies/erasers are already packed and from this, children have to compute the total number. We therefore used children's scores out of 2 on the remaining questions as the dependent variable in an ANCOVA with school and grade as between-subjects variables and PPVT-4 raw scores as the covariate.

TABLE 2
Mean performance by grade, school type, and item type, Experiment 2

Grade	*School Type*	*Number Ordering*	*Numeral Interpretation*	*Multidigit Calculation*	*School Sale*	*Number Line Error Rate*
Kindergarten	Non-Montessori	.42 (.23)	.23 (.23)	.00 (.01)	.03 (.12)	.34 (.10)
	Montessori	.36 (.18)	.22 (.13)	.01 (.02)	.03 (.12)	.37 (.07)
Second grade	Non-Montessori	.64 (.24)	.42 (.23)	.04 (.10)	.06 (.17)	.25 (.07)
	Montessori	.68 (.24)	.63 (.29)*	.35 (.23)*	.28 (.31)*	.21 (.12)

Note. Means and standard deviations are reported. Significant group differences between Montessori and non-Montessori based on two-tailed *t* tests are indicated with an asterisk, $p < .05$.

Overall, there was a trend toward improvement on these problems from kindergarten to second grade, $F(1, 63) = 3.77$, $MSE = .04$, $p = .06$, $\eta_p^2 = .06$. For the comparisons involving school type, there was a significant main effect, $F(1, 63) = 4.66$, $MSE = .04$, $p = .04$, $\eta_p^2 = .07$, that favored Montessori students (15% vs. 4%) and a significant interaction between school type and grade, $F(1, 63) = 5.07$, $MSE = .04$, $p = .03$, $\eta_p^2 = .07$. This interaction was due to equal, near-floor performance in both grades in the non-Montessori group ($M_{diff} = 0.4\%$, $p = .92$) but significantly better performance for second graders than kindergarteners in the Montessori group ($M_{diff} = 23\%$, $p = .03$).[3] This finding demonstrates that Montessori students ultimately achieved a better understanding of the school sale problem, but not until second grade.

Number Line Estimation. An ANCOVA that used children's error rates as the dependent variable, with grade and curriculum as between-subjects factors and their PPVT-4 raw scores as a covariate, revealed no significant differences based on schooling. There was a main effect of grade, $F(1, 63) = 4.30$, $MSE = .01$, $p = .04$, $\eta_p^2 = .06$, such that second-grade students had lower error rates than those in kindergarten ($M_{\text{kindergarten}} = 35\%$, $M_{\text{second grade}} = 23\%$); however, there were no other significant main effects or interactions. Thus, there was improvement with age, as in previous research on the number line task (e.g., Booth and Siegler, 2006); however, unlike the other two measures, there was no Montessori advantage. This finding is consistent with the lack of improvement in the blocks group on number line estimation (Experiment 1) and suggests there is a distinction between the developing knowledge that underlies success on the school sale problem and number line estimation. Specifically, the school sale problem requires decomposition and composition across units whereas number line estimation does not. This pattern again points to a curriculum dense in experience with manipulatives as possibly providing special support for a developing structural understanding of the place value system.

Discussion

There was a clear advantage for Montessori students on a range of place value tasks. Because the main distinguishing feature of Montessori mathematics instruction is its early, consistent integration of concrete models, this advantage likely reflects exposure to these materials, at least in part. Interestingly, however, this advantage did not appear in kindergarten, but rather, it emerged between kindergarten and second grade. This pattern may reflect the fruition of a long incubation period during which children gradually internalize their experiences with concrete models and link them to symbolic procedures, as one would predict based on the embodied cognition view (Barsalou, 2008; Glenberg & Robertson, 2000; Lakoff & Nunez, 2001).

In relation to previous research on the effects of Montessori instruction, our results are consistent with several findings. First, as in other studies (e.g., Lillard, 2012), we showed a performance advantage for Montessori students versus those in traditional schooling. Also, we showed this effect for mathematics in particular, consistent with prior work (e.g., Dohrmann et al., 2007). However, unlike previous studies demonstrating these effects in younger children and with less exposure (e.g., Lillard, 2012, Lillard & Else-Quest, 2006), our effects emerged in the second grade, after 5 years of exposure to Montessori education. One possibility is that the particular content we measured—place value concepts—requires more exposure or time to develop than the particular mathematics content measured in these previous studies (i.e., the Applied Problems subtest from the *Woodcock Johnson-3*).

Though our results are suggestive of a link between mathematics outcomes and instruction with concrete models, there is no way to directly connect the two in a long-range retrospective study such as the present study, raising the possibility of several alternative interpretations. For example, one might argue that Montessori teachers simply spend more time teaching place value than do teachers in traditional classrooms and this explains the observed advantage. Alternatively, the performance differences might not be related to mathematics at all, but rather to more general Montessori advantages, such as fostering better executive function (Lillard, 2012; Lillard & Else-Quest, 2006), that have cascading effects across academic areas (e.g., Blair & Razza, 2007; Bull & Scerif, 2001; Clark, Pritchard, & Woodward, 2010). Finally, one might question whether SES or overall intelligence contributed to the apparent Montessori advantage, despite our efforts.

These interpretations seem unlikely, however. First, the Montessori and non-Montessori samples were matched on vocabulary scores, and these scores were used as a covariate in our analyses. This approach should have reduced or eliminated any population differences or general effects of instructional method (e.g., executive function). Second, there was not a Montessori advantage across the board. Children's performance was similar in the two school groups at kindergarten and on number line estimation for both grades. These findings argue against the notion that Montessori students simply receive more place value instruction or that they perform better due to fundamental differences in SES, income, cognitive function, and so forth.

Indeed, the correspondence between the patterns in Experiment 1 and Experiment 2 suggests that long-term experience with manipulatives may be the core difference between Montessori and non-Montessori mathematics instruction. In multiple weeks of training in Experiment 1, children trained with blocks were better at interpreting multidigit numerals, and they showed better understanding of the structure underlying place value in the school sale problem; in the months and years of instruction received by Montessori students in Experiment 2, we observed the same advantages in comparisons to matched peers in other schools. In the multiple-week training of Experiment 1, children trained only with symbols performed better on number line estimation than did children trained with blocks; in the months and years of instruction received by Montessori students, and despite their overall higher achievement on the place value test, they did not exhibit better performance on number line estimation compared with matched peers in other schools.

Clearly, further research is needed to disentangle the subtle differences in instruction that may contribute to the Montessori advantage we observed. However, the present study provides evidence suggestive of a late-emerging benefit for place value learning and, at the least, demonstrates that long-term exposure to concrete models is not detrimental.

GENERAL DISCUSSION

Research has shown that place value concepts are difficult for children to acquire and thereby create serious obstacles to later mathematics achievement. In the present study, we considered components of symbol grounding that might explain where and how children become stuck. In particular, we examined the effects of instruction using concrete models and asked whether differences in outcome measures, age, and ability might explain related discrepancies in the literature.

In a multiweek training study that mimicked the complex training conditions of school instruction, children benefitted from training with either concrete models or symbols alone on some measures. However, for certain children and on certain measures, there was a clear advantage of training with concrete models, and for other children and other measures, there was an advantage of training with symbols alone. In a second experiment that capitalized on naturally occurring educational variation, we found that children exposed to years of instruction with concrete models showed advantages akin to those exhibited by children trained with concrete models in Experiment 1, but the effects were even larger and more clear-cut.

This pattern of findings helps to explain discrepancies in the existing literature. First, previous studies reporting a disadvantage of exposure to concrete models may not have provided adequate exposure, as our effects were strongest in Montessori students who had received 5 years of instruction with concrete models. Second, previous studies did not consider initial ability, but we found an interaction whereby high-ability children fared better with symbol-based instruction and low-ability children benefitted more from concrete models. Thus, how much a child already knows about a symbol system may determine which method is most effective.

That said, we know from previous research that most children struggle to master place value notation and typically exhibit stubborn errors and misconceptions despite limited competence on some tasks (Cauley, 1988; Cobb & Wheatley, 1988; Fuson & Briars, 1990; Jesson, 1983; Kamii, 1986; Kouba et al., 1988; Labinowicz, 1985; Resnick & Omanson, 1987). Perhaps children can make surface-level gains based on exposure to symbols, but these gains only take children so far. The ability to judge the rough ordinality of multidigit numbers, for example, may not be enough to support performance on more challenging tasks where understanding the multiplicative structure of base-10 notation is required (Laski, Ermakova, & Vasilyeva, 2014; Moeller et al., 2011). Concrete models may play an important role in supporting this specific process (i.e., grounding the meaning of this multiplicative structure). The finding that blocks training was particularly advantageous on novel transfer tasks that targeted base-10 structure is consistent with this notion.

Although we did not continue our training long enough to demonstrate further transfer, it seems plausible children could use this initial mapping to bootstrap their way into the more obscure nested structure of the written place value system, based on the extensive literature on analogical learning (e.g., Gentner, 2010). One indication is the large Montessori advantage we observed for both conventional symbolic problems (e.g., multidigit calculation) and specific measures of base-10 structure (e.g., the school sale problem). This pattern, along with the evidence for smaller improvements on the same measures from blocks training in Experiment 1, suggests that concrete models might enhance a structural understanding of place value that ultimately leads to better overall mathematics knowledge.

Pertinent to this hypothesis is the finding that the Montessori advantage was not apparent in kindergarten students; it only emerged in second grade. It makes sense that this process would be protracted because the mapping problem for place value is quite challenging. Indeed, though symbol grounding does not necessarily require lengthy exposure to a model (e.g., when learning words like “cookie” or “cat”), it might in the case of place value because of its complex internal structure. Concrete models may help bridge the gap, but as others have pointed out (e.g., McNeil et al., 2009; Uttal et al., 1997), these models have their own representational structures that must be unpacked and mapped to be useful. It could take several years to complete this mapping, even with carefully structured models and extensive scaffolding of comparisons. If so, then research

on concrete models that uses only a few brief training sessions may not provide enough exposure and incubation time for positive effects to be measured.

In sum, much like the existing literature related to place value acquisition and concrete models, the present study yielded mixed results. There were ways in which children learned about symbols from symbols and ways in which exposure to base-10 blocks yielded unique insights. It seems likely that children benefit from both streams of input, and it is probably incorrect to frame this research problem in terms of absolutes. Instead, our results indicate how nuanced and context-sensitive the process of symbol grounding for place value acquisition might be. Additional research aimed at delineating the interplay of various inputs, rather than simply pitting one against the other, is needed to clarify the complex mechanisms that likely drive this learning. The most critical question may not be whether concrete models are beneficial, but rather, what underlying skills and concepts are supported by different kinds of input and what cascading consequences they have for later mathematics learning.

ACKNOWLEDGMENTS

We are grateful to all of the children and school personnel who participated in this study, including those from the Holt Public Schools, Okemos Public Schools, Howell Public Schools, Fowlerville Public Schools, Leslie Public Schools, Montessori Children's House of Lansing, Stepping Stones Montessori School of East Lansing, Montessori Radmoor School of Okemos, St. Thomas Elementary School, and community programs in the cities of Lansing, Mason, and Grand Ledge, MI.

FUNDING

This research was supported by a generous grant to the first and second authors from the Institute of Education Sciences (#R305A080287). The opinions and positions expressed in this article are the authors' and do not necessarily represent the opinions and positions of the Institute of Education Sciences or the U.S. Department of Education.

REFERENCES

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology*, *20*, 93–114. doi:10.2307/271083

Ball, D. L. (1992). Magical hopes: Manipulatives and the reform of math education. *American Educator*, *16*(2), 14–18, 46–47.

Barsalou, L. B. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. doi:10.1146/annurev.psych.59.103006.093639

Bassok, M., & Medin, D. L. (1997). Birds of a feather flock together: Similarity judgments with semantically rich stimuli. *Journal of Memory and Language*, *36*, 311–336. doi:10.1006/jmla.1996.2492

Bassok, M., Wu, L., & Olseth, L. K. (1995). Judging a book by its cover: Interpretative effects of content on problem solving transfer. *Memory & Cognition*, *23*, 354–367. doi:10.3758/BF03197236

Bednarz, N., & Janvier, B. (1982). The understanding of numeration in primary school. *Educational Studies in Mathematics*, *13*, 33–57. doi:10.1007/BF00305497

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*, 647–663. doi:10.1111/cdev.2007.78.issue-2

Booth, J., L. & Siegler, R.S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *41*, 189–201. doi:10.1037/0012-1649.6.189

Bornstein, M. H., Haynes, M. O., & Painter, K. M. (1998). Sources of child vocabulary competence: A multivariate model. *Journal of Child Language*, *25*, 367–393. doi:10.1017/S0305000998003456

Bruner, J. S., Oliver, R. R., & Greenfield, P. M. (1966). *Studies in cognitive growth*. New York, NY: Wiley.

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

Byrge, L., Smith, L. B., & Mix, K. S. (2014). Beginnings of place value: How preschoolers write three-digit numbers. *Child Development*, *85*, 437–443. doi:10.1111/cdev.12162

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology*, *105*, 380–400. doi:10.1037/a0031084

Cauley, K. M. (1988). Construction of logical knowledge: Study of borrowing in subtraction. *Journal of Educational Psychology*, *80*, 202–205. doi:10.1037/0022-0663.80.2.202

Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology*, *48*, 1176–1191. doi:101037/a0019672

Cobb, P., & Wheatley, G. (1988). Children's initial understandings of ten. *Focus on Learning Problems in Mathematics*, *10*(3), 1–28.

Dohrmann, K. R., Nishida, T. K., Gartner, A., Lipsky, D. K., & Grimm, K. J. (2007). High school outcomes for students in a public Montessori program. *Journal of Research in Childhood Education*, *22*, 205–217. doi:10.1080/02568540709594622

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. doi:10.3758/BRM.41.4.1149

Fuson, K. C. (1986). Roles of representation and verbalization in the teaching of multi-digit addition and subtraction. *European Journal of Psychology of Education*, *1*, 35–56. doi:10.1007/BF03172568

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

Fuson, K. C., & Briars, D. J. (1990). Using a base-ten blocks learning/teaching approach for first- and second-grade place-value and multidigit addition and subtraction. *Journal for Research in Mathematics Education*, *21*, 180–206. doi:10.2307/749373

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*, 752–775. doi:10.1111/cogs.2010.34.issue-5

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science*, *4*, 152–158. doi:10.1111/j.1467-9280.1994.tb00652.x

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38. doi:10.1016/0010-0285(83)90002-6

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, *96*, 424–436. doi:10.1037/0022-0663.96.3.424

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language*, *43*, 379–401. doi:10.1006/jmla.2000.2714

Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, *46*, 414–466. doi:10.1016/S0010-0285(02)00519-4

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.

Ho, C. S. H., & Cheng, F. S. F. (1997). Training in place-value concepts improves children's addition skills. *Contemporary Educational Psychology*, *22*, 495–506. doi:10.1006/ceps.1997.0947

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*, 236–248. doi:10.1037/0012-1649.27.2.236

Jesson, D. S. (1983). The development of place value skills in primary and middle school children. *Research in Education*, *29*, 69–79.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*, 23–31. doi:10.1207/S15326985EP3801_4

Kamii, C. (1986). Place value: An explanation of its difficulty and educational implications for the primary grades. *Journal of Research in Childhood Education*, *1*, 75–86. doi:10.1080/02568548609594909

Kaminski, J. A., & Sloutsky, V. M. (2009). The effect of concreteness on children's ability to detect common proportion. In N. Taatgen & H. Van Rijn (Eds.), *Proceedings of the XXXI Annual Conference of the Cognitive Science Society* (pp. 335–340). Mahwah, NJ: Erlbaum.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). Learning theory: The advantage of abstract examples in learning math. *Science*, *320*, 454–455. doi:10.1126/science.1154659

Kouba, V. L., Brown, C. A., Carpenter, T. P., Lindquist, M. M., Silver, E. A., & Swafford, J. O. (1988). Results of the fourth NAEP assessment of mathematics: Number, operations, and word problems. *The Arithmetic Teacher*, *35*, 14–19.

Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences*, *10*, 417–446. doi:10.1207/S15327809JLS1004new_2

Labinowicz, E. (1985). *Learning from children: New beginnings for teaching numerical thinking*. Menlo Park, CA: Addison-Wesley.

Lakoff, G., & Nunez, R. (2001). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York, NY: Basic Books.

Laski, E., Ermakova, A., & Vasilyeva, M. (2014). Early use of decomposition for addition and its relation to base-10 knowledge. *Journal of Applied Developmental Psychology*, *35*, 444–454. doi:10.1016/j.appdev.2014.07.002

Lillard, A. S. & Else-Quest, N. (2006). The early years: Evaluating Montessori education. *Science*, *313*, 1893–1894. doi:10.1126/science.1132362

Lillard, A. S. (2012). Preschool children's development in classic Montessori, supplemented Montessori, and conventional programs. *Journal of School Psychology*, *50*, 379–401. doi:10.1016/j.jsp.2012.01.001

Lillard, P. P. (1997). *Montessori in the classroom: A teacher's account of how children really learn*. New York, NY: Schocken.

Lopata, C., Wallace, N. V., & Finn, K. V. (2005). Comparison of academic achievement between Montessori and traditional education programs. *Journal of Research in Childhood Education*, *20*, 5–13. doi:10.1080/02568540509594546

McNeil, N. M., Uttal, D. H., Jarvin, L., & Sternberg, R. J. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learning and Instruction*, *19*, 171–184. doi:10.1016/j.learninstruc.2008.03.005

Miller, K. F., & Stigler, J. W. (1991). Meanings of skill: Effects of abacus expertise on number representation. *Cognition and Instruction*, *8*, 29–67. doi:10.1207/s1532690xci0801_2

Mix, K. S., Prather, R. W., Smith, L. B., & Stockton, J. D. (2014). Young children's interpretation of multidigit number names: From emerging competence to mastery. *Child Development*, *85*, 1306–1319. doi:10.1111/cdev.12197

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance—A longitudinal study on numerical development. *Research in Developmental Disabilities*, *32*, 1837–1851. doi:10.1016/j.ridd.2011.03.012

National Council of Teachers of Mathematics. (2000). *Principals and standards for school mathematics*. Reston, VA: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Authors.

Oakes, J. M., & Feldman, H. A. (2001). Statistical power for nonequivalent pretest–posttest designs: The impact of change-score versus ANCOVA models. *Evaluation Review*, *25*, 3–28. doi:10.1177/0193841X0102500101

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, *76*, 763–782.

Peterson, S. K., Mercer, C. D., & O'Shea, L. (1988). Teaching learning disabled students place value using the concrete to abstract sequence. *Learning Disabilities Research*, *4*, 52–56.

Piaget, J. (1951). *Play, dreams and imitation in childhood*. London, England: Heinemann.

Reimer, K., & Moyer, P. S. (2005). Third-graders learn about fractions using virtual manipulatives: A classroom study. *Journal of Computers in Mathematics and Science Teaching*, *24*, 5–25.

Resnick, L. B., & Omanson, S. F. (1987). Learning to understand arithmetic. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 3, pp. 41–95). Hillsdale, NJ: Erlbaum.

Richland, L. E., & Hansen, J. (2013). Reducing cognitive load in learning by analogy. *International Journal of Psychological Studies*, *5*, 69–80. doi:10.5539/ijps.v5n4p

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*, 428–444. doi:10.1111/cdev.2004.75.issue-2

Son, J. Y., Smith, L. B., & Goldstone, R. S. (2011). Connecting instances to promote children's relational reasoning. *Journal of Experimental Child Psychology*, *108*, 260–277. doi:10.1016/j.jecp.2010.08.011

Uttal, D. H., Amaya, M., Maita, M. R., Hand, L. L., Cohen, C. A., O'Doherty, K. & DeLoache, J. S. (2013). It works both ways: Transfer difficulties between manipulatives and written subtraction solutions. *Child Development Research*, Article ID 216367, doi:10.1155/2013/216367

Uttal, D. H., O'Doherty, K., Newland, R., Hand, L. L., & DeLoache, J. (2009). Dual representation and the linking of concrete and symbolic representations. *Child Development Perspectives*, *3*, 156–159. doi:10.1111/j.1750-8606.2009.00097.x

Uttal, D. H., Scudder, K. V., & DeLoache, J. S. (1997). Manipulatives as symbols: A new perspective on the use of concrete objects to teach mathematics. *Journal of Applied Developmental Psychology*, *18*, 37–54. doi:10.1016/S0193-3973(97)90013-7

Vance, J. H., & Kieren, T. E. (1971). Laboratory settings in mathematics: What does research say to the teacher? *The Arithmetic Teacher*, *18*, 585–589.

Appendix. Content of Training Sessions (Experiment 1)

	Instructional Activities	
Lesson Topics	*Blocks*	*Symbols-Only*
LESSON 1 Place Value Concepts Objective: Understand multidigit number meanings	(1) Sort a pile of blocks into different block types onto place value mats. (2) Copy a written number using base-10 blocks. (3) Given a spoken number name and construct the equivalent representation using both base-10 blocks and number cards. (4) Identify the larger of two written numerals using blocks representations and explain in terms of the number of digits and their places.	(1) Sort a pile of number cards (70) and order them smallest to largest (0–9). (2) Copy a written number using the number cards. (3) Given a spoken number name and construct the equivalent representation using number cards. (4) Identify the larger of two written numerals and explain in terms of the number of digits and their places.
LESSON 2 Multidigit Addition Without Carrying Objective: Learn the procedures for multidigit addition without carrying.	(1) Construct representations of written addition problems using base-10 blocks and move the blocks to reach a solution. (2) Write the sums using number cards.	(1) Practice written multidigit addition problems on worksheets. (2) Write the solutions on the worksheets.

(*Continued*)

(*Continued*)

	Instructional Activities	
Lesson Topics	*Blocks*	*Symbols-Only*
LESSON 3 Addition With Carrying Objective: Learn the procedures for solving multidigit addition problems with carrying.	(1) Learn the written notation for addition with carrying using parallel representations of blocks. (2) Practice written problems using base-10 blocks as supports. (3) Make equivalent block patterns using different block combinations (e.g., 1 ten stick = 10 ones).	(1) Learn the written notation for addition with carrying. (2) Practice written addition problems with carrying on worksheets.
LESSON 4 Subtraction Without Borrowing Objective: Learn the procedures for solving multidigit subtraction problems.	(1) Represent and solve problems that require borrowing using base-10 blocks. (2) Write the solutions using number cards.	(1) Learn the notation for written subtraction problems. (2) Practice written subtraction problems.
LESSON 5 Subtraction With Borrowing Objective: Learn the procedures for solving multidigit subtraction problems with borrowing.	(1) Learn the notation for written subtraction problems using parallel representations of blocks. (2) Practice written problems using base-10 blocks as supports. (3) Make equivalent block patterns using different block combinations (e.g., 1 ten stick = 10 ones).	(1) Learn the notation for written subtraction problems with borrowing. (2) Practice written subtraction problems with borrowing.
LESSON 6 Review	Mixed review and written practice problems.	Mixed review and written practice problems.

NOTE: Roughly one third of the 101 children ($n = 20$ in the blocks condition and $n = 15$ in symbols-only condition) were held to a time limit of 30 min per lesson (or 180 total min of training) to ensure that all children had the same amount of exposure and completed all six lessons. These children completed all six lessons. The remaining children ($n = 32$ in the blocks condition and $n = 34$ in the symbols-only condition) were allowed to move through the lessons as slowly as they needed to reach mastery on each subtopic, with an upper limit of 6 weeks. Some, but not all, completed the six lessons in this timeframe. Also, some children progressed very quickly and finished in less time than others, who were given even more instruction in an attempt to get them through the whole training set. This approach resulted in a range of training amounts from 12 sessions to 20 sessions (mean = 14 sessions) or 360 to 600 instructional min (mean = 420 min). A preliminary analysis of variance indicated the pattern of performance on a composite posttest was parallel for children in the two groups (time-limited and self-paced), $F(1, 93) = 0.85$, $MSE = .05$, $p = .36$, $\eta_p^2 = .01$, so we have combined their data in the primary analyses. However, because not all children received the subtraction lessons, we analyzed performance on addition problems only.