CHILD DEVELOPMENT

Child Development, xxxx 2016, Volume 00, Number 0, Pages 1–13

Learning Mathematics in a Visuospatial Format: A Randomized, Controlled Trial of Mental Abacus Instruction

David Barner
University of California, San Diego

George Alvarez
Harvard University

Jessica Sullivan
Skidmore College

Neon Brooks
Harvard University

Mahesh Srinivasan
University of California, Berkeley

Michael C. Frank
Stanford University

Mental abacus (MA) is a technique of performing fast, accurate arithmetic using a mental image of an abacus; experts exhibit astonishing calculation abilities. Over 3 years, 204 elementary school students (age range at outset: 5–7 years old) participated in a randomized, controlled trial to test whether MA expertise (a) can be acquired in standard classroom settings, (b) improves students' mathematical abilities (beyond standard math curricula), and (c) is related to changes in basic cognitive capacities like working memory. MA students outperformed controls on arithmetic tasks, suggesting that MA expertise can be achieved by children in standard classrooms. MA training did not alter basic cognitive abilities; instead, differences in spatial working memory at the beginning of the study mediated MA learning.

Mathematics instruction typically begins by introducing children to a system of numerals and a set of arithmetic routines that operate on these numerals. For many children around the world, early math lessons are supplemented by the use of an abacus, a physical manipulative designed for representing exact quantities via the positions of counters, whose historical origins date to 1200 AD or earlier (Menninger, 1969). Extending the use of the physical abacus, children in countries such as India, China, Japan, and Singapore also learn a technique known as mental abacus (MA). Through MA, users create and manipulate a mental image of the physical device to perform arithmetic operations (see Figure 1 for details of how MA represents number). MA training results in remarkable abilities in expert users: It compares favorably to electronic calculators in speed and accuracy (Kojima, 1954), it enables rapid calculation even when users are speaking concurrently (Hatano, Miyake, & Binks, 1977), and it allows users as young as 10 years of age to win international calculation competitions like the Mental Calculation World Cup. MA training also appears to train numerical processing efficiency in children, as measured by a numerical Stroop paradigm (Du, Yao, Zhang, & Chen, 2014; Wang, Geng, Hu, Du, & Chen, 2013; Yao et al., 2015).

In the present study, we explored the nature of MA expertise. Specifically, we asked whether the extraordinary levels of achievement witnessed in experts can be attained by students in large K-12 classroom settings, and whether MA leads to gains in mathematics ability relative to more conventional curricula. In doing so, we also asked a more general question regarding the nature of expertise, and

We thank the children, families, and staff members of the Zenith School in Vadodara, India, for their patience and generosity over the past several years. We also thank Abbasi Barodawala and Mary Joseph of Zenith School, and Snehal Karia and Anand Karia of UCMAS India, for their invaluable contributions. Thank you also to Sean Barner, Eleanor Chestnut, Jonathan Gill, Ali Horowitz, Talia Konkle, Ally Kraus, Molly Lewis, Bria Long, Ann Nordmeyer, Viola Störmer, Jordan Suchow, and Katharine Tillman for their help with data collection. This work was funded by a grant to David Barner and George Alvarez from the NSF REESE (Grant 0910206) and by an NSF GRFP to Jessica Sullivan.

Corresponding concerning this article should be addressed to David Barner, Department of Psychology, UCSD, 9500 Gilman Drive, La Jolla, CA, 92093-0109. Electronic mail may be sent to barner@ucsd.edu.

 0009-3920/2016/xxxx-xxxx
DOI: 10.1111/cdev.12515

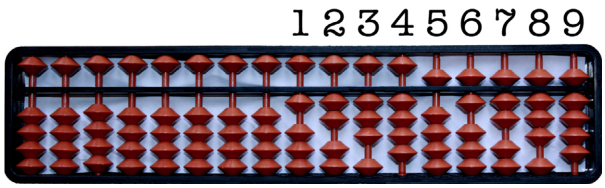

Figure 1. The Japanese soroban style abacus used by participants in this study, shown here representing the value 123,456,789. A physical abacus represents number via the arrangement of beads into columns, each of which represents a place value (e.g., ones, tens, hundreds, thousands, etc.), with values increasing from right to left. To become proficient at mental abacus, users of the physical abacus learn to create a mental image of the device and to manipulate this image to perform computations.

whether attaining unusual levels of performance requires changes to basic cognitive capacities, or instead arises via the exploitation of existing cognitive resources (see Ericsson & Smith, 1991, for review).

MA abilities appear to rely primarily on nonlinguistic representations, especially visuospatial working memory, as well as motor procedures that are learned during initial physical abacus training. Although the arithmetic computations of untrained college students are highly disrupted by verbal interference (e.g., concurrent speaking), MA users are less affected by concurrent linguistic tasks and much more affected by motor interference (Frank & Barner, 2011; Hatano et al., 1977). Consistent with this, while standard arithmetic routines recruit brain regions related to verbal processing and verbal working memory (VWM), MA computations selectively activate regions associated with vision and spatial working memory (SWM; Chen et al., 2006; Hu et al., 2011; Li et al., 2013; Tanaka, Michimata, Kaminaga, Honda, & Sadato, 2002). Both the structure of the abacus itself and MA users' computational limits are consistent with known limits to visual working memory. Like all attested abacus systems found in the human historical record, the soroban represents number by chunking beads into small sets of four or five, which corresponds to the hypothesized capacity limits described in the visual attention literature (e.g., Alvarez & Cavanagh, 2004; Atkinson, Campbell, & Francis, 1976; Luck & Vogel, 1997; Todd & Marois, 2004). Furthermore, users of MA appear to be limited to representing three or four abacus columns at a time, suggesting that each abacus column is represented as a distinct "object" in visuospatial working memory (Frank & Barner, 2011; Stigler, 1984). Although previous work has documented impressive abilities in MA experts, it does not address whether MA training can produce benefits for a broad range of students in a standard classroom setting. One previous study attempted to answer this question by assessing effects of MA training on mathematics performance in a large group of elementary school children (Stigler, Chalip, & Miller, 1986). However, students were not randomly assigned to conditions, and instead were self-selected according to their interest in abacus training. This lack of random assignment complicates inferences about MA as an educational intervention: Self-selecting groups of MA students are likely to be interested in abacus training, and thus may be more likely to benefit from MA training than randomly assigned groups of MA students. The present study thus tested the efficacy of MA training via random assignment. In particular, we were able to randomly assign participants in our study to receive either MA or standard math training.

We also investigated how students achieve expertise in MA. On one hand, MA expertise could result from changes in the user's ability to create and manipulate structures in visual working memory that are caused by MA training (a hypothesis we refer to as "cognitive transfer"). On the other hand, MA training may not create cognitive changes, but MA may instead exploit preexisting abilities, such that expertise arises only in individuals with relatively strong SWM abilities, who may be better able to learn MA (a hypothesis we call "cognitive moderation," because these cognitive abilities would serve as moderators of the technique's efficacy; Baron & Kenny, 1986). We describe these two alternative accounts of MA expertise in more detail next.

On the cognitive transfer hypothesis, MA expertise may result from gains in basic cognitive abilities—like imagery, working memory, and attention—that appear to be important for MA computation. Specifically, repeated practice of MA procedures—tracking and manipulating beads in visual space—may result in general advantages in visuospatial working memory that in turn allow for further MA mastery. By some accounts, MA training may even have cognitive effects outside of SWM, and may affect basic skills like reading ability (see Stigler et al., 1986). While rare, examples of cognitive transfer have been found in a small set of interventions targeting executive function and spatial cognition, which are both hypothesized to moderate academic performance (Diamond & Lee, 2011; Holmes, Gathercole, & Dunning, 2009; Uttal et al., 2013). Other research has suggested that working memory in early childhood may be flexible and strongly influenced by formal schooling experience (Roberts et al., 2015). Because MA requires hours of intensive practice with mental imagery, attentional allocation, and cross-domain integration of information (e.g., from numerical symbols to beads and back again), it may offer an especially strong opportunity for cognitive transfer. Furthermore, on the hypothesis that MA expertise results from transfer, it may be a technique that can be learned by any student who is willing to undergo the needed training, making it potentially useful in a wide range of classroom settings.

If MA training does lead to improvement in working memory, this improvement could provide a second route for the training to impact classroom arithmetic learning. Mathematical skills have been shown to rely not only on VWM resources, but also on the kind of visuospatial working memory that is used in MA training (Hubber, Gilmore, & Cragg, 2014; Simmons, Willis, & Adams, 2012). In addition, there are some suggestions that comprehensive working memory training can transfer to classroom mathematics learning, improving mathematics reasoning outcomes 6 months after training (Holmes et al., 2009).

On the cognitive moderation hypothesis (Ericsson & Smith, 1991), in contrast, MA may be most beneficial to a particular subset of students. Rather than stemming from changes to an individual's basic cognitive capacities, MA expertise may result when MA is learned and practiced by individuals who are particularly able to perform complex computations in working memory, and to manage the attentional demands required by the method (Frank & Barner, 2011). Specifically, individuals who exhibit unusual MA expertise (e.g., Frank & Barner, 2011; Hatano et al., 1977; Stigler, 1984) may begin training with unusual abilities to store and manipulate information in visual working memory. Such individuals may become experts not because the training affords expertise to anyone who pursues it, but because the training exploits users' existing cognitive resources. On this hypothesis, MA outcomes would be predicted by performance on cognitive tasks at the beginning of MA training because mastery of the technique requires some baseline level of cognitive abilities. But there would be no change in these abilities due to MA practice, unlike under the cognitive transfer hypothesis. Accordingly, the implementation of MA training in individual classroom settings would require care so as to ensure that the technique was appropriate for that particular group of students.

To explore the nature of MA and its utility in large classroom settings, we conducted a large longitudinal study at a school located in Vadodara, India. This school had previously adopted a short (1 hr) weekly MA training in addition to the standard curriculum for students in the second grade and above. Thus, instructors and appropriate training infrastructure were already in place. For our study, the school agreed to alter their curriculum so that starting at the beginning of the second grade, half of the children in our study could be randomly assigned to study MA for 3 hr/week (MA group). The remaining half of students were assigned to a control group who received no abacus training but instead performed 3 hr of supplementary practice using a state-approved K–12 mathematics curriculum (resulting in an identical amount of supplemental training). The supplemental curriculum was selected because it reflects the current standard in K–12 mathematics education in India, and thus represents the best supplemental training currently known to be available, and the most likely choice absent a stronger alternative.

We followed children over the course of 3 years and assessed outcomes using a battery of mathematical and cognitive assessments, including measures of mental rotation, approximate number, and spatial and VWM. These tasks were administered both prior to intervention and at the end of each school year so as to probe the extent of any possible cognitive mediation or transfer effects.

Method

Participants

We enrolled an entire cohort of English-medium students attending a charitable school for low-income children in Vadodara, India. Children are admitted to the school on a first-come, first-served basis for a fee of 630 rupees (US $10) per month, which is paid in full by the school trust in cases of need. At the initiation of the study, over 80% of children attending the school came from families who earned < US $2,000 per annum (~US $5.50/day). In our sample, 59% of children came from Hindu families and 41% from Muslim families. Most children were native speakers of Gujarati, the local dialect, and also spoke Hindi and English (the language of instruction in the English-medium program at the school). The total population of the school was approximately 2,100 students, ranging from pre-K to high school.

Data collection took place from June 2010 through March 2013. At the time of enrollment (which we refer to as Year 0), the participants were 204 children aged 5–7 years old who were beginning their second-grade year. We randomly assigned these students to two groups, MA and control. We then further randomly assigned children into three homeroom classrooms of approximately 65–70 children each (differing from their classroom assignments in the previous year), with one classroom comprising MA students, one control students, and one split half and half. Thus, children in each group took all of their classes together, with the exception of the split group, who were separated when receiving supplemental mathematics training (either MA or standard curriculum, depending on their group assignment). Because these differences could have affected students' uptake of the intervention—for example, due to differences in peer influence—we present analyses of possible classroom effects in the Supporting Information.

Of the 100 students in the MA group and the 104 students in the control group, 88 (88%) and 99 (95%) provided some data in every year of testing, respectively. Dropouts from the study were primarily due to changes of school. We analyzed data only from this group of 183 students (those who were present for the entire study). If a child was present for each study visit but had missing data for some individual measures, their data were included in the sample. The proportion of missing values for measures ranged from 0.1% (VWM) to 4.6% (number comparison). Missing values for individual measures were sometimes due to sickness, absences, or an inability to answer any items (especially in early years).

Intervention Procedure

Children in both the MA and control groups studied the school's standard (nonabacus) mathematics curriculum over the duration of the study in their regular home classrooms. Additionally, both groups received 3 hr/week of additional mathematics instruction as follows. In the MA group, children were given 3 hr/week of instruction in the use of the physical and MA by an experienced MA teacher outside the children's home classroom (such that control group children were not exposed to MA technique). The same teacher provided MA instruction to all MA children. Abacus instruction was broken into two 90-min sessions per week (3 hr/week total) and followed a common international curriculum that begins with use of the physical abacus for addition and subtraction, and then moves to MA computations. The 1st year of training focused primarily on the physical abacus, with greater emphasis placed on MA in subsequent years. Common activities in the MA training program included worksheet practice of addition and subtraction, practice translating abacus configurations into Arabic numerals, and practice doing speeded arithmetic using MA.

Control students were provided with two 90-min sessions per week of supplemental mathematics training using the Oxford University Press *New Enjoying Mathematics* series, designed in accordance with the Indian National Curriculum Framework (National Council for Education Research and Training, 2005). Texts in this series emphasize both conceptual mathematics and drills, including training of mental math with "worksheets focusing on special strategies followed by exercises for fast calculation" (see http://www.oup.co.in), and thus constitute a strong control to the MA manipulation.

Assessment Procedure

The study spanned 3 years of the participants' elementary education, and began with a baseline test before training began. In each of four annual assessments, children received a large battery of computerized and paper-based tasks. Year 0 assessments were given at the beginning of second grade; Years 1–3 assessments were given at the end of second, third, and fourth grades, respectively. All assessments included both measures of mathematics and more general cognitive measures. A small number of other tasks were included but are not discussed in the current study (see the Supporting Information for detailed descriptions of all measures and administration procedures).

Mathematics Measures

Children completed the Calculation subtest of the Woodcock–Johnson Tests of Achievement (WJ–IIIC) and the Math Fluency subtest of the Wechsler Individual Achievement Test (WIAT–III). We also administered two in-house assessments of mathematics skill that, unlike the standardized tests, were designed to specifically target arithmetic skills acquired between second and fourth grades (see the Supporting Information). The first measured the children's arithmetic abilities by testing performance in single- and multidigit addition, subtraction, division, and multiplication problems. The second measured the conceptual understanding of place value by asking children to complete number decomposition problems (e.g., 436 = 6 + ___ + 30).

Cognitive measures. At each assessment point, children completed one to two subsets of 10 problems from Raven's Progressive Matrices (Raven, 1998), as well as two paper-based tests of speeded mental rotation (one using letters and one using shapes). They also completed three computerized tasks: (a) an adaptive test of SWM, (b) an adaptive test of VWM, and (c) a number comparison task, in which children were asked to indicate the larger of two dot arrays (see the Supporting Information, Section 1 for detailed task descriptions and related citations). For the working memory tasks, we report estimates of span—that is, the average number of items on which a child was successful. For the number comparison task, we report Weber fractions (a measure of approximate number acuity, estimated from our task via the method of Halberda, Mazzocco, & Feigenson, 2008).

Grades. For each year, we obtained children's grades in English, math, science, and computer classes, as well as in music, art, and physical education.

Abacus-only measures. For each year after the intervention began, students in the MA group completed a set of three paper-and-pencil tasks to assess their ability to use an abacus. All three were administered after the end of all other testing. The first two, abacus sums (addition) and abacus arithmetic (addition and subtraction), tested the ability to do computations using a physical abacus, while the third, abacus decoding, tested the ability to decode abacus images into standard Arabic numerals. These tasks allowed us to verify that any differences between the MA and control groups were indeed mediated by changes to abacus skills.

Attitude measures. In Year 3, we administered two measures that explored whether the intervention had changed children's attitude toward mathematics and, thus, whether training effects might be mediated by differences in motivational level and engagement with mathematics. This allowed us not only to test whether MA caused changes in children's attitudes toward mathematics, but also to ask whether any differences we find between groups might be due, in part, to a placebo effect, whereby exposure to a new training paradigm improves performance by changing children's attitudes toward math. First, we administered a growth mindset questionnaire (adapted from Dweck, 1999), which probed children's attitudes regarding the malleability of their own intelligence. Second, we administered a math anxiety questionnaire (adapted from Ramirez, Gunderson, Levine, & Beilock, 2013), which measured the anxiety that children experienced when solving different kinds of math problems and participating in math class.

Data Analysis

Despite randomization, there were some baseline differences between the MA and control groups at initiation. The MA group performed significantly better on two of the four math assessments (though they did not differ on any of the cognitive measures): arithmetic, $t(183) = 2.65$, $p = .01$, and WIAT–III, $t(160) = 2.08$, $p = .04$. Because of these differences, we interpret simple comparisons between groups with caution. Instead, we used a longitudinal mixed models approach to quantify the statistical reliability of the effects of randomization to training group on our outcome measures. This approach controls for baseline effects for individual students and attempts to predict longitudinal gains (rather than absolute level of performance) as a function of intervention group (see the Supporting Information, Section 4.8 for more discussion and an alternative approach to this issue using propensity score matching). Critically, the baseline differences observed here (with a modest advantage for the MA group) render the longitudinal growth models *less* likely to find significant intervention effects. Consequently, in one interpretation, our study could be an overly conservative assessment of MA training. We revisit this issue and its implications for our findings in the Discussion.

For each outcome measure, we fit a baseline model that included a growth term for each student over time and an overall main effect of intervention group (to control for differences between groups at study initiation). We then tested whether the fit of this model was improved significantly by an interaction term capturing the effect of the intervention over time. All data and code for the analyses reported here are available at http://github.com/langcog/mentalabacus; further details of statistical models are available in the Supporting Information, Section 4.1.

Because we did not have any a priori hypotheses about the shape of the dose–response function between the intervention and particular measures of interest, we fit models using three types of growth terms: (a) simple linear growth over time, (b) quadratic growth over time, and (c) independent growth for each year after baseline. This last type of model allows for the possibility of nonmonotonic growth patterns. We tested for interactions between group assignment and growth in each of these models. All *p* values are reported from likelihood ratio tests (see the Supporting Information, Section 4.1).

Results

Mathematics Outcomes

As seen in Figure 2, MA training produced significant gains in mathematical abilities relative to the control group. Consistent with this, in Year 3, we observed numerical differences between the two groups on three of the four mathematics tasks, with effect sizes of Cohen's d = .60, 95% CI [.30, .89] for arithmetic; .24, 95% CI [−.05, .52] for WJ–IIIC; and .28, 95% CI [.00, .57] for place value. We observed only a small numerical difference for WIAT–III, however, d = .13, 95% CI [−.15, .42].

Because the confidence intervals for the effect sizes described above represent confidence intervals on pairwise tests for Year 3 alone, they do not control for baseline differences at initiation. Thus, we used longitudinal growth models to assess whether advantages observed in the MA group were in fact driven by additional MA training (Figure 2). All three of these models (i.e., linear, quadratic, and nonmonotonic) showed strong Time × Condition interactions for both the arithmetic and WJ–IIIC measures, suggesting that performance on each of these tasks did improve with additional MA training. Likelihood ratio tests for adding the Time × Condition interaction term to the growth model were $\chi^2_{linear}(1) > 6.33$, $\chi^2_{quadratic}(2) > 11.56$, and $\chi^2_{independent}(3) > 12.51$, with $ps < .01$ in all cases.

Consistent with the small numerical difference observed between groups on the WIAT–III, this measure did not approach significance in any of the three growth models. The smaller effects observed on the standardized measures are perhaps not surprising, given the smaller number of arithmetic-focused items on these measures and hence the likelihood of them having lower sensitivity to indi-

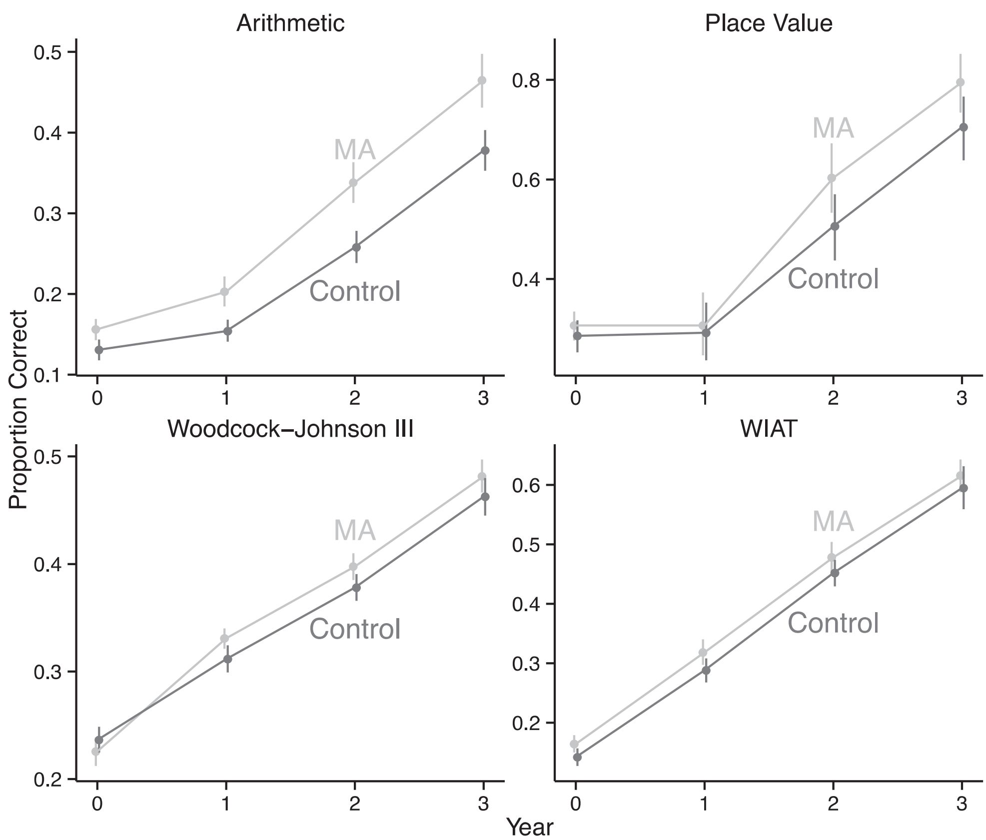

Figure 2. Mathematics outcome measures for the two intervention conditions, plotted by study year (with 0 being preintervention). Error bars show 95% confidence intervals computed by nonparametric bootstrap. MA = mental abacus; WIAT = Wechsler Individual Achievement Test.

vidual differences between children of this age (see the Supporting Information, Section 4.7 for further analysis). More surprising, however, was that the place value measure did not approach significance in the growth models, especially given that we observed substantial numerical differences between the groups (e.g., performance in Year 3 differed significantly between groups in a univariate analysis), $t(185) = 1.96$, $p = .05$. We speculate that we did not observe consistent growth with this measure due to its low reliability from Year 0 to Year 1 ($r = .22$).

Together, these analyses suggest that the MA intervention was more effective in building students' arithmetic skill than an equivalent amount of supplemental training in standard mathematics techniques. Effects of MA training on arithmetic ability were observed not only in our in-house measure, which included many arithmetic problems tailored to the level of elementary school students, but also on the WJ–III test of Calculation, a widely used standardized measure that includes a range of problem types and formats. While the evidence for differential gains in conceptual understanding of place value was more limited, MA students did not fall behind students in the control group, despite the fact that the MA curriculum primarily stresses rote calculation rather than conceptual understanding.

Cognitive Outcomes

Although MA training produced consistent gains in arithmetic ability, it did not produce consistent gains in the cognitive abilities we measured (Figure 3). Higher math performance in the MA group was therefore not the result of improved cognitive capacity due to MA training. For example, in Year 3, we observed between-group effect sizes of −.16, 95% CI [−.45, .12] on our number comparison measure (note that smaller Weber fractions indicate more accurate estimations). Also, we found an effect size of −.14, 95% CI [−.43, .15] for Raven's progressive matrices; of −.06, 95% CI [−.34, .23] for mental rotation; and of .05, 95% CI [−.24, .34] for SWM. Only one cognitive measure—VWM—showed an advantage in Year 3 for the MA group, .26, 95% CI [−.03, .55].

For these cognitive measures, as with the arithmetic tasks, we used longitudinal growth models to assess whether advantages for the MA group were driven by training in MA. Because we used different sets of Raven's problems for each year, we could not fit growth models, but *t* tests showed no reliable effects of MA training for any year (all ts < 0.96, ps > .34). Similarly, longitudinal models (linear, quadratic, and independent) confirmed that none of the cognitive tasks (numerical comparison, mental rotation, VWM, or SWM), showed significant Time × Condition interactions, with one exception. For VWM, the nonindependent growth model showed a significant Time × Condition interaction ($p < .01$), though both linear and quadratic growth models showed no significant Time × Condition interaction (Supporting Information, Section 4.4). Thus, this result appears to have been driven by the fast growth in VWM span in Year 1 exhibited by the MA group, relative to the control group (see Figure 3).

The large effect of MA training on VWM in Year 1 is mirrored in a similar trend observed in SWM in Years 1 and 2, significant or close to significant in individual *t* tests, $t(185) = 2.36$, $p = .02$ and $t(184) = 1.84$, $p = .07$, but not in any longitudinal model. In both cases, the overall shape of the developmental curve is asymptotic, with working memory spans approaching approximately four items by Year 2 in the MA group. This pattern could be interpreted as evidence that differences in working memory between the MA and control groups do exist, but are expressed only in the rate of growth to asymptote rather than in the absolute level of the asymptote itself. Against this hypothesis, however, additional analyses (Supporting Information, Section 4.5) find that (a) our SWM task did not exhibit ceiling effects and (b) data from 20 American college undergraduates and 67 high-socioeconomic-status (SES) Indian children from the same region of India show that children in our study had overall lower SWM than higher SES children, and were far from being at adult levels of performance. Most important, these Year 1 effects surfaced before children began to receive training on the mental component of MA and were still learning the physical technique. We therefore do not believe that this result is likely to be related to the ultimate gains we see in MA across the study.

Academic Outcomes

MA did not produce large, consistent changes in students' grades across academic subjects, although we saw some small trends toward better math, science, and computer grades in the MA group in some models. These differences are subject to teacher bias, however, since teachers were of course knowledgeable about the intervention. Thus, we do not believe they should be weighted heavily in evaluating performance, especially since our own

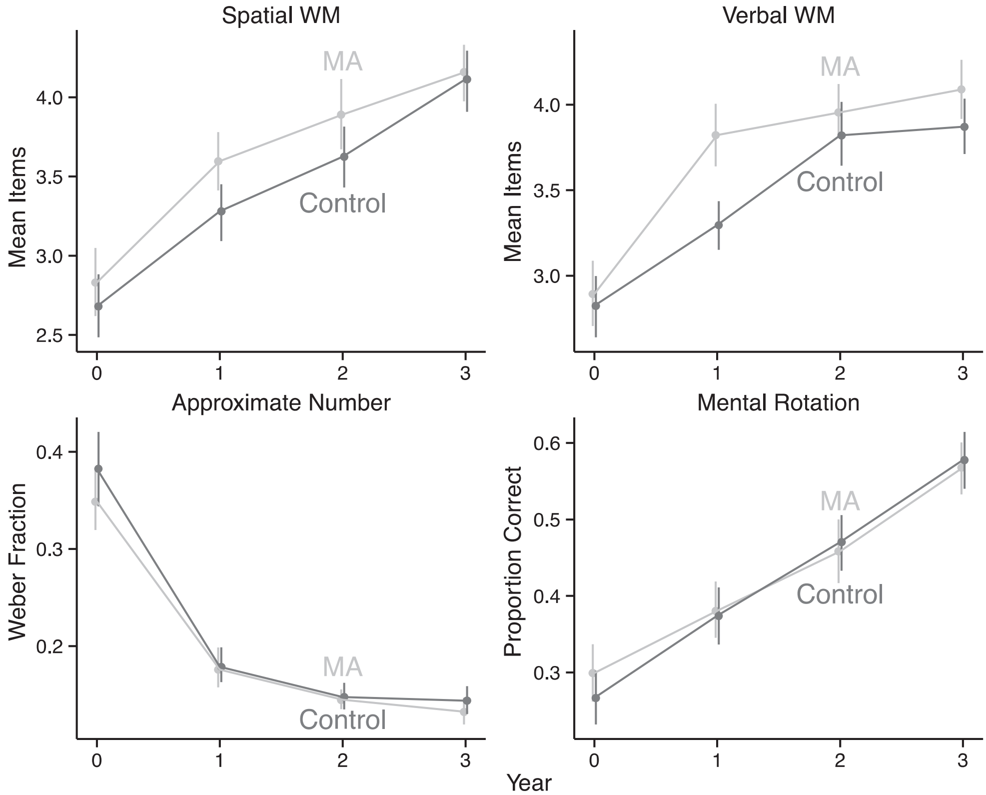

Figure 3. Cognitive outcome measures for the two intervention conditions, plotted by study year. Top axes show mean items correct in working memory span tasks, while bottom axes show proportion correct across trials in number comparison and mental rotation tasks. Error bars show 95% confidence intervals computed by nonparametric bootstrap. MA = mental abacus; WM = working memory.

standardized measures of mathematical competence were available for analysis (for additional analysis, see the Supporting Information and Figure S1).

Attitude Measures

There were no differences between groups on either children's self-reported mathematics anxiety, $t(184) = 1.05$, $p = .29$, or their endorsement of a growth mindset, $t(184) = -0.61$, $p = .54$. Thus, it is unlikely that differences we observed in mathematics measures were due to differential effects of our intervention on children's anxiety about mathematics or on their general mindset toward learning.

Mediators of Intervention Effects

Given that MA training produced gains in math outcomes, we next asked which factors mediated these gains, and thus whether individual differences between children at the beginning of the study predicted MA achievement. As already noted, MA training did not augment cognitive abilities, so the math advantages in the MA group could not have been driven by enhanced working memory, mental imagery, or approximate number acuity that resulted from MA training. However, it is possible that individual cognitive differences between children in the MA group (prior to their entry into the study) were responsible for how well they learned and benefited from MA. To explore this possibility, we conducted post hoc analyses using moderator variables.

Our analytic approach relied on the same longitudinal modeling approach described earlier. For each math outcome variable, we fit models that included participants' Year 0 performance on each cognitive predictor (for simplicity and to avoid overparameterizing our models, we used linear and quadratic models only). The coefficient of interest was a three-way interaction of time, condition, and initial performance on the cognitive predictor of interest. This three-way interaction term captures the intuition that growth in performance on a task for MA participants is affected by their baseline abilities on a particular cognitive task. As stated earlier, we used likelihood ratio tests to assess whether these interaction terms improved model fit. Although with greater numbers of longitudinal measurements we could potentially have detected interactive growth patterns (e.g., gains in working memory driving later gains in mathematics), our current study did not have the temporal resolution for these analyses. We thus restrict our analysis to testing for mediation in mathematics outcomes on the basis of each of the cognitive variables measured at Year 0.

A median split of children according to SWM prior to MA training resulted in a low-SWM group with an average threshold of 1.9 items, and a high-SWM group with an average threshold of 3.7 items. Previous studies of SWM thresholds for middle- to high-SES 5- to 7-year-olds find thresholds of approximately 3.5 to 4 items (Logie & Pearson, 1997; Pickering, 2010). Thus, a subset of children in our study exhibited especially low SWM capacity (for comparison, slightly older high-SES participants from the same city had mean SWM thresholds of 4.7, and adult controls had a threshold of 6.4; see the Supporting Information). Related to this, SWM was a reliable moderator in both linear and quadratic models of arithmetic (Figure 4). Those children who began the study with relatively strong SWM skills and who were randomly assigned to MA training showed significantly stronger growth in our arithmetic assessment, $\chi^2_{\text{linear}}(1) = 4.63$, $p < .05$, and $\chi^2_{\text{quadratic}}(2) = 5.93$, $p = .05$. Relative to receiving equal amounts of standard math training, children with weaker SWM did not appear to benefit differentially from the MA intervention and instead performed at an equivalent level to the students in the control group. Since the MA technique relies on visuospatial resources for storage of the abacus image during computation (Frank & Barner, 2011), it seems likely that those children with relatively lower SWM spans struggled to learn to perform computations accurately using MA.

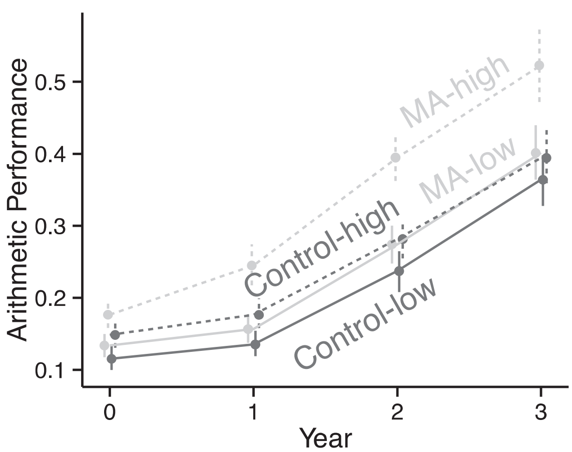

Figure 4. Performance on the arithmetic task, split by both intervention condition and median spatial working memory performance in Year 0. Error bars show 95% confidence intervals; lines show best fitting quadratic curves. MA = mental abacus.

There was no comparable mediation effect with VWM (additional analysis in the Supporting Information, Section 4.4), but a small number of additional moderation effects did approach significance in one of the two models. There was a trend toward an effect of SWM on WJ–IIIC, the other mathematics measure that showed strong MA effects, $\chi^2_{\text{quadratic}}(2) = 5.59$, $p = .06$, (Figure S3). In addition, there were trends toward effects of Year 0 mental rotation performance on arithmetic, $\chi^2_{\text{linear}}(1) = 2.71$, $p = .10$; place value, $\chi^2_{\text{linear}}(1) = 3.27$, $p = .07$; and an effect of number comparison acuity on WJ–IIIC, $\chi^2_{\text{quadratic}}(2) = 8.29$, $p = .02$ (Figure S4). These effects, though more tentatively supported, are nevertheless consistent with the hypothesis that the MA intervention was most effective for children with greater visuospatial abilities at the beginning of instruction.

Abacus-Only Measures

Confirming that children in the MA group learned to use an abacus, we found consistently high performance in abacus decoding for the MA group (> 80% correct for all years). Also, performance on the abacus arithmetic and sums tasks rose substantially from year to year, suggesting that children's abacus computation abilities improved over the course of their training (Figure 5). Performance on these tests of physical abacus arithmetic were significantly correlated with performance on our other math measures (in-house arithmetic: *r*s = .69, .74, and .81 for Years 1–3, respectively, all *p*s < .0001; WIAT: *r*s = .57, .54, .73, all *p*s < .0001; and WJ: *r*s = .45, .51, .64, all *p*s < .0001).

Critically, SWM span in Year 0 was also related to intervention uptake, as measured by the abacus-only tasks, which required MA students to use a physical abacus. Because we did not have abacus-only data for Year 0, we could not directly test whether SWM moderated growth, but we did find a main effect of SWM on all three measures of abacus uptake for both linear and quadratic growth models, all χ^2s (1) > 3.97, all *p*s < .05 (Figure 5).

Discussion

Our study investigated the nature of MA expertise, and whether MA is an effective tool for improving math outcomes in a standard classroom setting. To

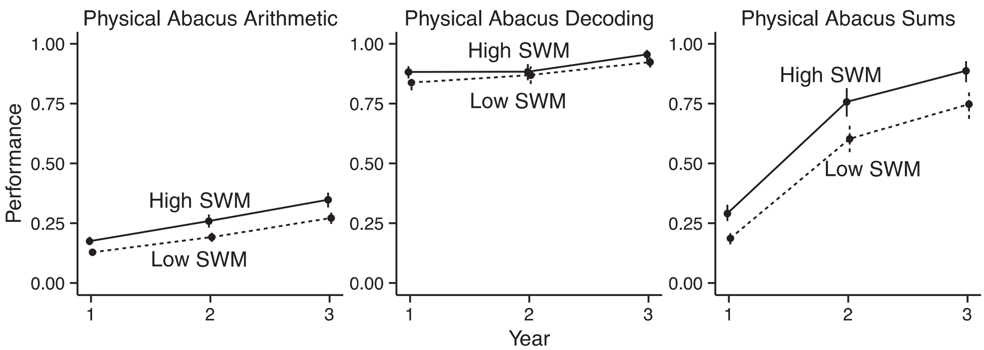

Figure 5. Performance on the Physical Abacus Sums, Decoding, and Arithmetic tasks (administered in Years 1–3), plotted by a median split on spatial working memory (SWM). Error bars show 95% confidence intervals computed by nonparametric bootstrap.

do this, we conducted a 3-year longitudinal study of MA training. We found that MA training led to measurable gains in students' ability to perform accurate arithmetic computations. These gains began to emerge after a single year of training—suggesting that simply learning to use the physical abacus had some effects on students' mathematics aptitude, even prior to learning the MA technique—and became more pronounced with time. Consistent with a role for abacus expertise in explaining the intervention effect, we found that physical abacus expertise at the end of the study was significantly correlated with arithmetic performance across all math measures within our experimental group (see Stigler et al., 1986, for a similar result). Also, these gains were not related to motivational or attitudinal differences toward mathematics or intellectual self-efficacy, suggesting that gains in the MA group are not easily attributable to a placebo effect. Finally, although there were signs of early gains in cognitive capacities like SWM in the MA group, such effects did not persist to the end of the study and could not explain gains in mathematics achievement. Instead, we found that gains were most pronounced among children who exhibited relatively higher SWM capacity at the beginning of the study.

One important limitation of our study was "unhappy randomization," which resulted in baseline differences between MA and control groups despite random assignment. The critical analyses we report here are longitudinal growth models that control for ability at initiation, and we report other statistical corrections in the Supporting Information. But we cannot rule out the possibility that these baseline differences affected our findings. In one interpretation, a baseline advantage for MA would make it *harder* for us to detect training advantages above baseline (especially if part of that initial advantage was due to random factors other than true mathematics ability). In another interpretation, however, if the baseline advantage of the MA group were due to true differences in ability, these baseline differences could cause a cascade of further positive learning outcomes that our models do not control for (see Siegler & Pyke, 2013, for an example). Either interpretation suggests that our findings—in particular, the quantitative results regarding the size of mathematical gains due to MA—should be interpreted cautiously until they are replicated with another sample.

Acknowledging the caveat above, these findings nevertheless support three main conclusions. First, compared to standard methods of mathematics training, MA may offer some benefits to students seeking supplemental instruction. Relative to additional training with techniques used in popular mathematics textbooks, MA instruction appeared to result in greater gains in arithmetic ability and equivalent effects on conceptual understanding. However, the data also provide reason to believe that this advantage may be limited to children who begin training with average or above-average SWM capacity: A median split of children based on their Year 0 SWM capacity revealed that children with relatively higher SWM capacity were especially likely to benefit from training. Because MA relies on visuospatial resources for the storage and maintenance of abacus images during computation (Frank & Barner, 2011), children with especially weak SWM may have attained only basic MA abilities—enough to reap benefits equal to additional hours of standard math, but not to acquire unusual expertise.

This difference related to SWM capacity suggests a second conclusion, which is that the development of MA expertise is mediated by children's preexisting cognitive abilities. Consequently, MA may not be suitable for all K–12 classroom environments, especially for groups of children who have especially low SWM or attentional capacities (a situation that may have been the case for some children in our study). Critically, this finding does not imply that MA training benefits depend on unusually strong cognitive abilities. Perhaps because we studied children from relatively disadvantaged backgrounds, few children in our sample had SWM capacities comparable to those seen among typical children in the United States. Studies currently underway are exploring this possibility.

Third, based on the discussion thus far, our findings are consistent with previous suggestions that "cognitive transfer" is rare. Although performance on basic measures of attention and memory can be improved via direct training on those measures (Diamond & Lee, 2011; Gathercole, Dunning, Holmes, & Wass, 2016; Melby-Lervåg & Hulme, 2012; Noack, Lövdén, Schmiedek, & Lindenberger, 2013), it may be difficult to achieve "far" transfer from training on unrelated tasks, even with hours of focused practice (Dunning, Holmes, & Gathercole, 2013; Owen et al., 2010; Redick et al., 2013). However, our findings suggest that although cognitive capacities are not importantly altered by MA, they may predict which children will benefit most from MA training. MA students who began our study with low SWM abilities did not differ in their math performance from control students, while those above the median made large gains on our arithmetic measure (similar effects were not seen for VWM).

Our study leaves open several questions about MA as an educational intervention. First, it remains uncertain how much training is necessary to benefit from MA. In our study, children received over 100 hr of MA instruction over 3 years. Future studies should investigate the efficacy of MA training in smaller, more focused sessions, and also whether the required number of training hours is smaller in different populations (e.g., in middle- or high-SES groups). Second, future studies should contrast MA with other training regimens that focus more exclusively on intensive arithmetic training. Our current focus was to assess the practical utility of MA as an alternative to current supplemental training practices, and our study suggests an advantage for MA, at least on some measures. However, it is possible that other forms of training can also yield the types of benefits observed for MA and that the levels of performance seen in our study are not unique to visuospatial techniques like MA.

Finally, our results raise questions regarding the efficacy of concrete manipulative systems in the classroom. Although we found positive effects of abacus training compared to other methods, this benefit emerged over 3 years of extensive weekly training. Previous studies have found mixed results regarding the effectiveness of manipulatives for teaching mathematics (Ball, 1992; Uttal, Scudder, & DeLoache, 1997). However, MA may be unlike other manipulative systems. Although the abacus is a concrete representation of numerosity that can be used to reinforce abstract concepts, the method is unique in requiring the use of highly routinized procedures for arithmetic calculation. Thus, additional research is needed to understand how MA differs from other manipulatives with respect to its educational benefits.

In sum, we find evidence that MA—a system rooted in a centuries-old technology for arithmetic and accounting—is likely to afford some children a measurable advantage in arithmetic calculation compared to additional hours of standard math training. Our evidence also suggests that MA provides this benefit by building on children's preexisting cognitive capacities rather than by modifying their ability to visualize and manipulate objects in working memory. Future studies should explore the long-term benefits of enhanced arithmetic abilities using MA and the generalizability of this technique to other groups and cultural contexts.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111. doi:10.1111/j.0963-7214.2004.01502006.x

Atkinson, J., Campbell, F. W., & Francis, M. R. (1976). The magic number 4 ± 0: A new look at visual numerosity judgements. *Perception*, *5*, 327–334. doi:10.1068/p050327

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

Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. doi:10.1037/0022-3514.51.6.1173

Chen, F., Hu, Z., Zhao, X., Wang, R., Yang, Z., Wang, X., & Tang, X. (2006). Neural correlates of serial abacus mental calculation in children: A functional MRI study. *Neuroscience Letters*, *403*(1), 46–51. doi:10.1016/j.neulet.2006.04.041

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4–12 years old. *Science*, *333*, 959–964. doi:10.1126/science.1204529

Du, F., Yao, Y., Zhang, Q., & Chen, F. (2014). Long-term abacus training induces automatic processing of abacus numbers in children. *Perception*, *43*, 694–704. doi:10.1068/p7625

Dunning, D. L., Holmes, J., & Gathercole, S. (2013). Does working memory training lead to generalised improvements in children with low working memory? A randomised controlled trial *Developmental Science*, *16*, 915–926. doi:10.1111/desc.12068

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Ericsson, K. A., & Smith, J. (Eds.). (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge, UK: Cambridge University Press.

Frank, M. C., & Barner, D. (2011). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, *141*, 134–149. doi:10.1037/a0024427

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665–668. doi:10.1038/nature07246

Hatano, G., Miyake, Y., & Binks, M. G. (1977). Performance of expert abacus operators. *Cognition*, *5*, 57–71. doi:10.1016/0010-0277(77)90016-6

Holmes, J., Gathercole, S., & Dunning, D. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*, 9–15. doi:10.1111/j.1467-7687.2009.00848.x

Hu, Y., Geng, F., Tao, L., Hu, N., Du, F., Fu, K., & Chen, F. (2011). Enhanced white matter tract integrity in children with abacus training. *Human Brain Mapping*, *32*, 10–21. doi:10.1002/hbm20996

Hubber, P., Gilmore, C., & Cragg, L. (2014). The roles of the central executive and visuospatial storage in mental arithmetic: A comparison across strategies. *Quarterly Journal of Experimental Psychology*, *67*, 936–954. doi:10.1080/17470218.2013.838590

Kojima, T. (1954). *The Japanese abacus: Its use and theory*. Tokyo, Japan: Tuttle.

Li, Y., Hu, Y., Zhao, M., Wang, Y., Huang, J., & Chen, F. (2013). The neural pathway underlying a numerical working memory task in abacus-trained children and associated functional connectivity in the resting brain. *Brain research*, *1539*, 24–33.doi:10.1016/j.brainres.2013.09.030

Logie, R., & Pearson, D. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology*, *9*, 241–257. doi:10.1080/713752559

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. doi:10.1038/36846

Melby-Lervåg, M., & Hulme, C. (2012). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270–291. doi:10.1037/a0028228

Menninger, K. (1969). *Number words and number symbols: A cultural history of numbers*. Cambridge, MA: MIT Press.

National Council for Education Research and Training. (2005). *Indian National Curriculum Framework*. New Delhi, India: Author.

Noack, H., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2013). Age-related differences in temporal and spatial dimensions of episodic memory performance before and after hundred days of practice. *Psychology and Aging*, *28*, 467–480. doi: 10.1037/a0031489

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., & Ballard, C. G. (2010). Putting brain training to the test. *Nature*, *465*, 775–778. doi:10.1038/nature09042

Pickering, S. (2010). The development of visuo-spatial working memory. *Memory*, *9*, 423–432. doi:10.1080/09658210143000182

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development*, *14*, 187–202. doi:10.1080/15248372.2012.664593

Raven, J. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologist's Press.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., . . . Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*, 359–379. doi:10.1037/a0029082

Roberts, G., Quach, J., Mensah, F., Gathercole, S., Gold, L., Anderson, P., . . . Wake, M. (2015). Schooling duration rather than chronological age predicts working memory between 6 and 7 years: Memory Maestros study. *Journal of Developmental and Behavioral Pediatrics*, *36*, 68–74. doi:10.1097/DBP.0000000000000121

Siegler, R. S., & Pyke, A. A. (2013). Developmental and individual differences in understanding of fractions. *Developmental Psychology*, *49*, 1994–2004. doi:10.1037/a0031200

Simmons, F., Willis, C., & Adams, A. M. (2012). Different components of working memory have different relationships with different mathematical skills. *Journal of Experimental Child Psychology*, *111*, 139–155. doi:10.1016/j.jecp.2011.08.011

Stigler, J. W. (1984). "Mental abacus": The effect of abacus training on Chinese children's mental calculation. *Cognitive Psychology*, *16*, 145–176.

Stigler, J. W., Chalip, L., & Miller, K. (1986). Consequences of skill: The case of abacus training in Taiwan. *American Journal of Education*, *94*, 447–479. doi:10.1086/443862

Tanaka, S., Michimata, C., Kaminaga, T., Honda, M., & Sadato, N. (2002). Superior digit memory of abacus experts: An event-related functional MRI study. *NeuroReport*, *13*, 2187–2191. doi:10.1097/00001756-200212030-00005

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*, 751–754. doi:10.1038/nature02466

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*, 352–402. doi:10.1037/a0028446

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

Wang, Y., Geng, F., Hu, Y., Du, F., & Chen, F. (2013). Numerical processing efficiency improved in experienced mental abacus children. *Cognition*, *127*, 149–158. doi:10.1016/j.cognition.2012.12.004

Yao, Y., Du, F., Wang, C., Liu, Y., Weng, J., & Chen, F. (2015). Numerical processing efficiency improved in children using mental abacus: ERP evidence utilizing a numerical Stroop task. *Frontiers in Human Neuroscience*, *9*, 245. doi:10.3389/fnhum.2015.00245

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Data s1. Learning Mathematics in a Visuospatial Format.