Research Reports

A One-Year Classroom-Randomized Trial of Mental Abacus Instruction for First- and Second-Grade Students

David Barner*[a], Angeliki Athanasopoulou[a], Junyi Chu[a], Molly Lewis[b], Elisabeth Marchand[a], Rose Schneider[c], Michael Frank[b]

[a] Department of Psychology, University of California, San Diego, CA, USA. [b] Department of Psychology, Stanford University, Stanford, CA, USA. [c] Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA.

Abstract

Mental Abacus (MA) is a popular arithmetic technique in which students learn to solve math problems by visualizing a physical abacus structure. Prior studies conducted in Asia have found that MA can lead to exceptional mathematics achievement in highly motivated individuals, and that extensive training over multiple years can also benefit students in standard classroom settings. Here we explored the benefits of shorter-term MA training to typical students in a US school. Specifically, we tested whether MA (1) improves arithmetic performance relative to a standard math curriculum, and (2) leads to changes in spatial working memory, as claimed by several recent reports. To address these questions, we conducted a one-year, classroom-randomized trial of MA instruction. We found that first-graders students struggled to achieve abacus expertise over the course of the year, while second-graders were more successful. Neither age group showed a significant advantage in cognitive abilities or mathematical computation relative to controls, although older children showed some hints of an advantage in learning place-value concepts. Overall, our results suggest caution in the adoption of MA as a short-term educational intervention.

Keywords: Math education, mental arithmetic, mental abacus, spatial working memory, place value

Mental Abacus (MA) is a popular technique for supplementing elementary school mathematics education, especially in Asian countries, where there is a tradition of abacus use in schools and businesses (Li, 1959). Students of MA begin by learning to make rapid arithmetic computations on a physical abacus, a device which likely descended from early Greek and Roman counting boards and which has been used in Asia since at least 1200 AD (Ifrah, 2000; Menninger, 1969). With practice, students then learn to construct a mental representation of the device, allowing them to move the beads "in their mind," without reference to the physical device (Figure 1; Frank & Barner, 2012; Hatano, Miyake, & Binks, 1977; Miller & Stigler, 1991; Stigler, 1984). Skilled MA users can accomplish astonishing feats, and have used the technique in recent years to win international competitions, like the World Cup of Mental Arithmetic. Young children trained in MA are able to rapidly add, subtract, multiply, and divide large numbers using MA, and can even learn to compute square and cube roots.
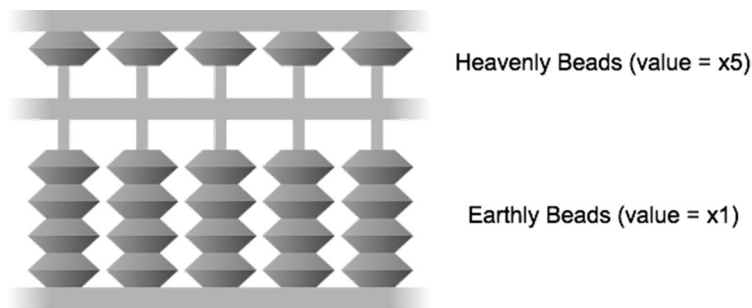
Figure 1. A typical abacus consists of several vertical columns upon which beads move freely. A horizontal beam divides each column into an upper and a lower deck. Each column has 4 or 5 beads on the lower deck ("earthly beads") and 1 or 2 beads on the upper deck ("heavenly beads"). On a column chosen by the user, each bead in the lower deck represents a cardinality of one, and each bead in the upper deck represents five. Lower deck beads on the next column to the left represent multiples of 10 (and 50 on top), and beads on the next column represent multiples of 100 (and 500 on top), etc. Beads are "in play" (contributing to a numerical representation) when moved towards the central beam.

Previous studies have argued that the unusual abilities of MA users stem from the format of MA computations, which are thought to be supported by gestural and visuospatial representations. Following early studies by Hatano and colleagues (Hatano, Miyake, & Binks, 1977), in our previous work we examined effects of verbal and motor interference on MA calculation (Brooks, Barner, Frank, & Goldin-Meadow, 2017; Frank & Barner, 2012). While naive control participants were much more disrupted by verbal interference, MA users showed larger interference effects for manual interference. In addition, a growing neuroimaging literature suggests that MA computations are associated with regions that support spatial and visual working memory (Chen et al., 2006; Du et al., 2013; Hanakawa, Honda, Okada, Fukuyama, & Shibasaki, 2003; Hu et al., 2011; Li et al., 2013; Li, Chen, & Huang, 2016; 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 other attested abacus systems found in the human historical record, the Soroban abacus represents number by chunking beads into small sets of 4 or 5, 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 3 or 4 abacus columns at a time, suggesting that each abacus column is represented as a distinct "object" in visuospatial working memory (Frank & Barner, 2012; Stigler, 1984). These facts suggest that MA builds on the visuospatial capacities of learners by transforming a serial linguistic process into a potentially less constrained visual workspace, and thereby linking abstract symbolic mathematics to concrete representations of objects and sets – a strategy often used in math education outside of MA, in the form of simple manipulative systems (Ball, 1992; Hatano, Amaiwa, & Shimizu, 1987; Uttal, Scudder, & DeLoache, 1997).

The apparent reliance of MA expertise on visuospatial working memory raises questions about its potential utility as an educational tool for typical K-12 students (the main question of the present research). In most previous reports of MA expertise, students were self-selecting, highly motivated learners, who may have been especially predisposed to learning math in a visuo-spatial format (Frank & Barner, 2012; Hatano, Miyake, & Binks, 1977; Huang et al., 2015; Stigler, 1984; Stigler, Chalip, & Miller, 1986). Consequently, these studies left open the question of whether MA can be effectively implemented in the classroom environments that serve broader and more representative groups of elementary school children, where significant variability in visual

working memory abilities exist. Can MA be learned by typical groups of K-12 students, or is it best suited to highly motivated students with especially strong visual working memory abilities?

In one recent study (Barner et al., 2016), we began to explore this question by assessing the effects of MA in a three-year randomized controlled trial. In that study, we tested 190 elementary school students in Gujarat, India (aged 5 to 7 years at outset), and compared the effect of three years of weekly MA training (3 hours per week) to equivalent amounts of additional standard math curriculum. At the end of 3 years, we found significant advantages due to MA training on a number of mathematics measures, including the Woodcock-Johnson III and WIAT. However – consistent with the hypothesis that MA might be most beneficial to children with high visuo-spatial working memory – we also found that improvement was mediated by children's individual visual working memory capacities at the beginning of the study. We also tested for whether MA led to advantages by decreasing children's anxiety about mathematics, but found no support for this hypothesis.

In addition to making claims about its efficacy as an educational intervention, some recent studies have claimed that MA training may result in changes to working memory, given its visuo-spatial format (Dong et al., 2016; Na et al., 2015; Wang et al., 2015). However, support for this claim has been mixed, owing to the relatively small samples tested and the presence of mixed results across studies. For example, in one recent study (Dong et al., 2016), 36 adult subjects were divided into two groups who either received no training, or were given 90 minutes of MA training per day for 20 consecutive days. At the end of the study, the authors found that MA-trained participants performed significantly better at a digit span task, but not a letter span task, relative to controls (the latter task exhibited a marginally significant effect, with no difference at the end of the study between groups). Also, they found a difference between groups on a 4-back task, but not a matched 2-back task. Another study, by Wang et al. (2015) used a slightly larger design, and divided 70 7-year-old children into two groups, one of which received three years of MA training, for 2 hours per week. After three years of training, the MA group showed relatively greater mathematics ability. However, the effects of training on other cognitive abilities was mixed. Although Wang et al. found main effects indicating differences between the experimental and control groups on cognitive tasks, as well as within-group improvements in the MA group, critically, there were no group by time interactions for cognitive tasks, as would be required to support the conclusion that group differences actually resulted from MA training. Thus, although these reports are intriguing, they fall short of the standard required to show strong evidence for "cognitive transfer" effects (Simons et al., 2016).

As MA grows in popularity internationally, it has begun to appear in the United States, and has been introduced to multiple public schools to supplement existing mathematics instruction. This turn of events raises the question of whether MA can be implemented effectively in the context of the US public school system, and whether it compares favorably to alternative enrichment programs. Our past work leaves this question open for a number of reasons. First, our previous study was conducted at a private charitable school, which was able to adjust its curriculum to accommodate intensive MA training and provide substantial and ongoing teacher training in how to accommodate MA in the curriculum. Thus, it remains unknown whether the schedule and resources of a typical US public school can accommodate a novel technique like MA. Second, although our previous study was conducted at a private school, it served a population of very low-income families, and featured very large class sizes by US standards – around 60 – 70 students per class. Therefore, while the school had more flexibility to introduce MA, the students may have been less prepared for training than typical US students and may have received relatively little teacher attention, raising the possibility that US students

could benefit more, or with smaller amounts of training. Relevant to this suggestion, in that previous study, we not only found that our sample's visual working memory capacity was predictive of their MA uptake, but also that overall their visual working capacity was quite low relative to comparison samples from higher socio-economic status Indian children. Third, because of the large class size, a small number of teachers were responsible for MA instruction. Thus, we were not able to test for the generality of intervention benefits across classrooms.

In the present study, we investigated the potential effectiveness of MA in the US context by conducting a one-year, classroom-randomized trial of MA instruction in a Northeastern US public school. Classes were randomly assigned to either maintain their standard math curriculum (Common Core Singapore Math; Control) or to receive a mixture of their standard math plus MA training. We assessed performance during the first week of classes, before MA was administered, and near the final week of classes, after children in the treatment group had received a full academic year of training.

Our study had two main goals. First, we asked whether a group of US school children (N = 180) would experience greater benefit from MA training than from equivalent hours of standard math curriculum. This first goal was applied. Our interest was in whether MA instruction – as implemented by a standard commercial provider – would be sufficient to lead to benefits for students, either through increases in calculation abilities or decreases in math-related anxiety. We were not attempting to provide a "proof of concept" that optimal instructional practices could lead to benefit for some populations, as prior work has already provided some evidence of that claim (Barner et al., 2016; Stigler et al., 1986). Second, we also explored a theoretical question, and tested the claim that even modest amounts of MA training might yield benefits to visuo-spatial working memory abilities. Although "far transfer" – i.e., general improvements to spatial working memory from focused practice on a particular task – is generally thought to be rare (Simons et al., 2016), the studies reviewed above often report differences in performance on classic tests of spatial working memory after even small amounts of MA training. We reasoned that if these effects are reliable and strong enough to be of significance to educational practices, then they should also emerge in a longer study with a larger number of participants, like ours.

Methods

We conducted a classroom-randomized evaluation of MA training in Grades 1 and 2 in a large elementary school. Classes were randomly assigned to a Control group who received their standard math curriculum (Common Core Singapore Math) or an experimental group who received the same number of hours of instruction but divided between their standard curriculum and mental abacus. We assessed student performance during the first week of school, prior to MA instruction (pretest) and then at the end of the school year (posttest). Students were assessed at both time periods on measures of mathematics knowledge and on general cognitive measures (spatial working memory, executive function, and general reasoning). At posttest all students completed a measure of math anxiety. In addition, as a test of MA curriculum uptake, those students in the MA group were also tested on their ability to read an abacus.

Randomized Design and Curriculum

We partnered with a large school in a Northeastern US state, located in a large metropolitan area. The school was ethnically and socioeconomically diverse, with 67.1% of students eligible for free or reduced lunch in 2012 and > 90% black or Hispanic students. Prior to study initiation, we received a list of classrooms (N = 24) in Grades 1 and 2 and randomly assigned the classes to Control or MA conditions. We included inclusion classrooms (which contained some high-functioning children with developmental disabilities, N = 5) and dual language/bilingual transition classrooms (Spanish/English, N = 8); the remaining classes were general education (N = 7). After consultation with teachers and school administration, we excluded self-contained special education classrooms (N = 4). We communicated classroom assignments to a for-profit company who had already contracted with the school to provide teacher training in the MA curriculum. This company provided training to the relevant teachers in the first week of the school year as well as curriculum materials and supervision during the school year (with relatively frequent visits to check on classrooms' progress). Because both teachers and the MA curriculum developers were aware of the study design, this study (as with many educational interventions) was not fully blinded: teachers' expectations could be related to student performance. However, in an effort to diminish bias by researchers administering outcome measures, all researchers remained blind to participant condition during testing. Student assignment to condition was revealed only prior to administering the MA posttest measure (the final measure administered, and only to the MA group, hence the need for unblinding).

Students in the Control group followed their standard, existing "Singapore Math" curriculum. Singapore Math focuses on a gradual transition between concrete and abstract representations, and often makes use of concrete visualizations of problems (e.g., using schematic diagrams). Standard math class was scheduled daily for one period per day (40 min). The MA training group received three periods per week (40 min each) of MA training from their own teachers, using external curriculum materials; the remaining two periods per week were used for standard Singapore Math instruction, using the same materials as the control group. The MA materials focused first on the introduction of the physical abacus, and then on the basics of the MA technique. Specifically, children in Grade 1 learned addition using the complement of 5 (to prepare them for MA addition, which requires this skill for the use of the top bead, which denotes 5). By the end of the year, some could do mental abacus computations using both bottom and top beads. Children in Grade 2 learned addition using the complement of 10 (to prepare for multi-column abacus problems) and some could do mental abacus addition using the complement of 5.

Participants

We sent consent forms home with all students in Grades 1 and 2 and followed up with several school-wide announcements. All children from participating classrooms were enrolled in the study if they had valid consent forms, though a small number were sick or otherwise absent and were not tested in particular tasks or at both time points. Participants were only included in analyses if they had been tested at both time points. We also received some consent forms from children in excluded classes. Table 1 and Table 2 give an overview of the number of participants in different groups in the study.

Table 1

Participants in the Study

Group	Grade	Children Enrolled	Consent Form Received	Data Collected
Control	1st	103	49 (48%)	26
	2nd	89	51 (57%)	44
Mental Abacus	1st	95	46 (48%)	35
	2nd	114	70 (61%)	59

Table 2

Demographic Description of Participant Groups

| Descriptive statistics | Control | | Mental Abacus | |
	Grade 1	Grade 2	Grade 1	Grade 2
N Participants	26	44	35	59
Mean age	6.52	7.56	6.37	7.46
Median Age	6.49	7.51	6.32	7.41
Std. Dev. Age	0.39	0.41	0.29	0.36
N Hispanic	15	20	14	26
N African-American	7	13	11	24
N Mixed Race/Other	4	11	10	9

Measures

Students were tested twice: Once at the beginning of the school year and once at the end. In each case, all students were tested during a 4 to 5 day period. Testing was performed in the school library or unused classrooms, using portable laptop computers for the cognitive assessments. Students completed tasks in groups of 2 – 4 over the course of one or two sessions, totaling about 45 – 60 minutes. Cognitive assessments were given first, then mathematics assessments.

Cognitive Assessment

We administered a short battery of cognitive assessments to test for cognitive transfer (see Supplementary Materials: Cognitive Assessment Tasks). The first was a continuous performance go-no-go task in which students had to press a key to indicate a target (roads) and not for a distractor (mountains) (Corkum & Siegel, 1993). Students completed 100 test trials lasting 1200ms each, with a set of slower initial practice trials to allow students to familiarize themselves to the task. Distractors appeared on 20% of trials. The second task was a spatial working memory change detection task, in which students had to track gradually longer sequences of locations in a spatial grid (see details in Barner et al., 2016). The third was a computerized forced-choice matrix reasoning task based on Raven's progressive matrices (Raven, Raven, & Court, 2003), which continued until students reached 36 questions or made three mistakes sequentially.

Mathematics Assessment

All students completed three short assessments. The first was the Woodcock-Johnson III standardized measure, a two-page battery ranging from simple arithmetic to much more advanced high-school topics (5

minutes). The second was an in-house assessment of conceptual understanding of place-value, a foundational math concept that is part of the Common Core curriculum for Grades 1 and 2. The task required filling in the missing quantity in a place-value decomposition (e.g., 400 + ___ + 1 = 451) (5 minutes, reported in Barner et al., 2016). The third was an arithmetic fluency measure, which included 48 basic arithmetic problems to be completed in a short period (5 minutes). This measure was designed to allow MA-trained children to demonstrate their fluency with simple arithmetic and was included because it showed the largest training effect in our previous study.

MA Intervention Fidelity

To test children's uptake of MA, we administered an in-house assessment of abacus translation (measuring the accuracy and speed with which children can translate from an abacus state to Arabic numerals). Children had 5 minutes to complete 23 problems.

Math Anxiety

We assessed math anxiety using a questionnaire adapted from Ramirez, Gunderson, Levine, and Beilock (2013). Consistent with our previous work (Barner et al., 2016), we included this measure both to test whether MA caused differences in anxiety and also as a control measure to ensure that differences between groups were not caused by MA students having reduced anxiety, rather than by changes in actual ability. This questionnaire presented a set of 16 mathematics questions and asked students to rate their anxiety regarding each problem using a 5-point Likert scale.

Pre-registration and Analytic Strategy

Confirmatory analyses specified below were pre-registered at the Open Science Framework (see Supplementary Materials: Pre-Registration). Because we were not certain what dependent measures would be feasible to administer during our pretest visit, we collected data and analyzed our pretest results prior to registration. We then registered our hypotheses regarding posttest analyses prior to posttest data collection. Throughout we denote planned (registered) analyses as "confirmatory" and post-hoc analyses as "exploratory." We also distinguish "primary" analyses, which are tests of the direct effects of the intervention, from "secondary" analyses, which are tests related to the mechanism of action for the intervention.

All measures were normalized into the unit interval for ease of comparison and interpretation. In general, we normalized by the total number of questions on a measure. Since no child was able to answer questions on the second page of the Woodcock-Johnson III (which includes fractions, negative numbers, etc.), we included only the first page, containing 25 questions. For the Go/No Go task, we used accuracy on "no go" trials. For spatial working memory, we normalized spans arbitrarily by dividing by 10; this decision does not affect any statistical analyses and was made to simplify visualization.

As in our previous work (Barner et al., 2016), we conducted separate mixed effects regression analyses with each of the mathematics and cognitive outcome measures as outcome variable. The form of these models was:

outcome ~ grade + year * intervention + (1 | subid) + (1 + year | class)

The key coefficient in these models was the year-by-intervention interaction term, capturing the possibility of a greater increase in the outcome measure as a function of random assignment to intervention group. The

random intercepts for each participant control for participant-level individual differences, and the random slopes and intercepts for classes capture different baseline levels and growth patterns across classes.

Results

Descriptive Analyses

We had six primary outcome variables, corresponding to our six tasks: Three mathematics measures (Arithmetic, Place Value, and the standardized WJ III assessment) and three cognitive measures (Matrix Reasoning, Go/No Go, and Spatial Working Memory). The distribution of outcome variables for each task is shown in Figure 2. Performance on all measures was better for second graders than for first graders, and all measures showed positive growth over the course of the school year. Some showed larger changes than others due to features of the tasks themselves. For example, the place value measure was explicitly designed to capture content being learned during these two years of schooling and thus showed substantial movement (its distribution was also idiosyncratic because an understanding of two-place place-value would allow a student to complete a particular subset of questions). In contrast, the Go/No Go and Spatial WM tasks showed smaller changes relative to the amount of individual variation that we saw.
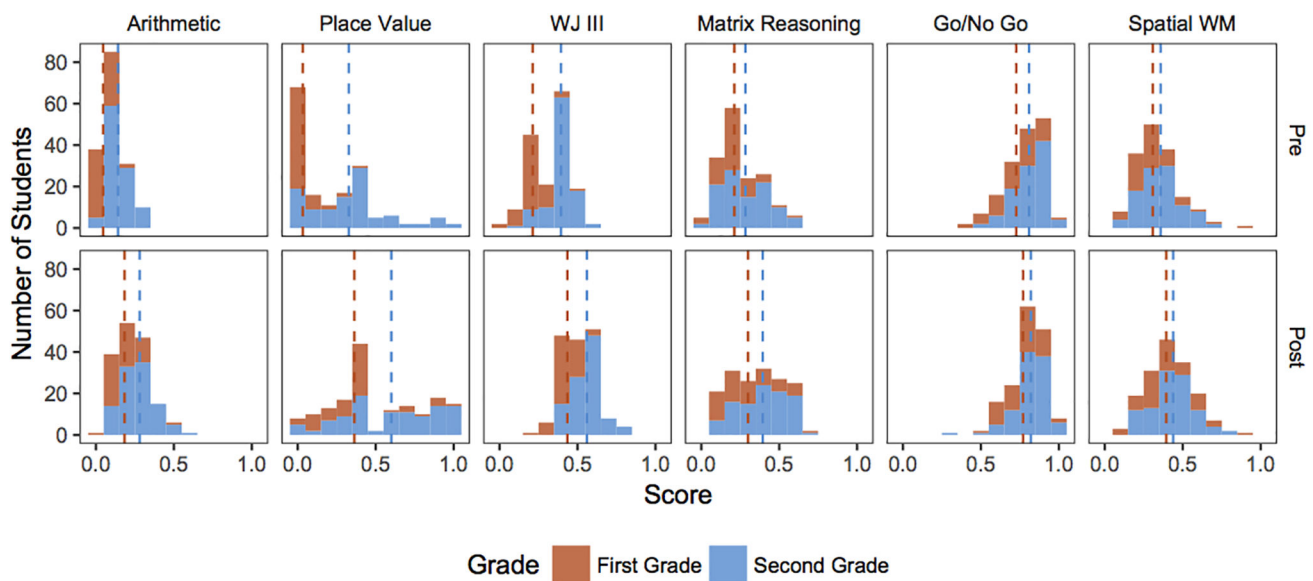

Figure 2. Histograms showing the distribution of scores from each task in our battery, split by grade level. Dashed lines show means. Upper panels show pre-test scores; lower panels show post-test scores.

All tasks showed evidence of modest test-retest reliability across the school year (range = .30 – .63, see Table 3), comparable to the reliability found in our previous work (Barner et al., 2016). Higher reliability would of course increase our power to detect condition effects, but would be difficult to achieve without substantially longer testing sessions. In addition, test-retest correlations were likely depressed because of real changes over the course of the study. For example, we would not expect place value scores to be highly correlated between pre- and post-tests given that many students learned new place value concepts over the course of the year.

Table 3

Test-Retest Correlations Within Task and Grade

Task	Grade	r	Lower 95% CI	Upper 95% CI	p
Arithmetic	1st	.52	0.30	0.68	<.0001
	2nd	.49	0.33	0.63	<.0001
Place Value	1st	.31	0.06	0.52	.0154
	2nd	.63	0.49	0.73	<.0001
WJ III	1st	.32	0.07	0.53	.0126
	2nd	.38	0.20	0.53	.0001
Matrix Reasoning	1st	.33	0.09	0.54	.0091
	2nd	.33	0.15	0.50	.0005
Go/No Go	1st	.56	0.35	0.71	<.0001
	2nd	.45	0.28	0.59	<.0001
Spatial WM	1st	.43	0.19	0.62	.0008
	2nd	.30	0.12	0.47	.0019

We also examined intervention uptake at the end of the study (Figure 3) by analyzing students' ability to decode abacus representations.
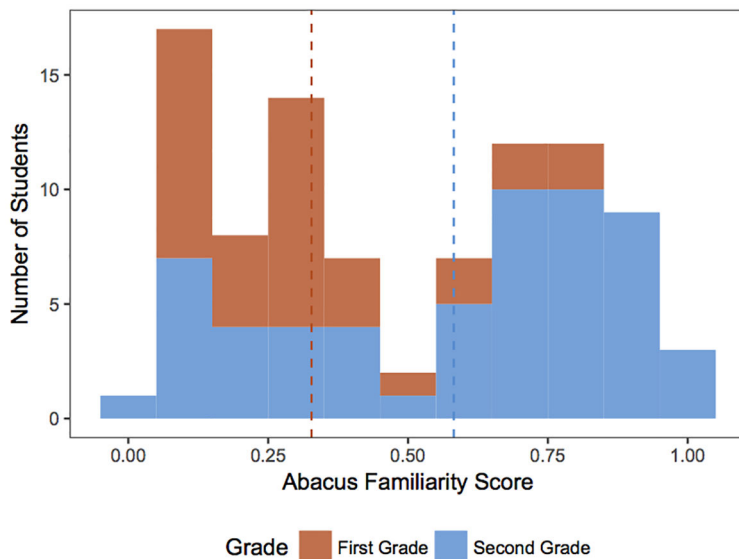

Figure 3. Histogram of abacus familiarity scores, broken down by grade. Dashed lines show means.

We found a roughly bimodal distribution of children, with some children relatively proficient at decoding abacus representations and others quite poor and only able to do so for 1 - 2 digit displays (a skill which they nevertheless did not possess prior to the intervention). The relative balance of children in the two modes was different across grades, however, with a much larger population of second-graders gaining proficiency in the technique. These uptake findings are an important metric of the appropriateness of MA instruction. A relatively

small proportion of first graders could accurately decode a multi-digit abacus by the end of one year of instruction (21%). Thus, MA may not have been an appropriate curriculum for these children, given their place value knowledge. We discuss this result in more depth below, but it qualifies the interpretation of all subsequent outcome measures for the intervention.

Primary Analyses

The primary question addressed by our confirmatory analyses was whether assignment to treatment condition (MA vs. Control) resulted in differential change in mathematical or cognitive measures. Due to model convergence issues, we deviated from our pre-registered plan by removing random slopes for individual classes (this decision is consistent with our standard operating procedures for how to deal with non-convergent mixed effects models). Table 4 shows all models, with *p*-values computed via the *t=z* method, which is appropriate here due to the relatively large amount of data (Barr, Levy, Scheepers, & Tily, 2013). Figures 4 and 5 show scores for mathematics and cognitive tasks, respectively.

Table 4

Coefficients and Statistics for Planned Linear Mixed Effects Models Fit to Each Task.

Task	Predictor	β	Std Err	t	p
Arithmetic					
	Intercept	0.040	0.015	2.63	.0085
	Second Grade	0.095	0.015	6.28	<.0001
	Post-Test	0.146	0.011	13.52	<.0001
	Mental Abacus	0.009	0.017	0.55	.5838
	Post-Test x Mental Abacus	-0.013	0.014	-0.90	.3665
Place Value					
	Intercept	0.038	0.039	0.98	.3257
	Second Grade	0.266	0.038	7.05	<.0001
	Post-Test	0.251	0.029	8.62	<.0001
	Mental Abacus	0.021	0.042	0.50	.6154
	Post-Test x Mental Abacus	0.074	0.038	1.94	.0527
WJ III					
	Intercept	0.230	0.013	17.93	<.0001
	Second Grade	0.154	0.012	12.98	<.0001
	Post-Test	0.186	0.012	15.67	<.0001
	Mental Abacus	-0.001	0.014	-0.10	.9218
	Post-Test x Mental Abacus	0.001	0.016	0.09	.9303
Matrix Reasoning					
	Intercept	0.201	0.026	7.75	<.0001
	Second Grade	0.081	0.026	3.17	.0015
	Post-Test	0.114	0.021	5.56	<.0001
	Mental Abacus	0.005	0.029	0.17	.8661
	Post-Test x Mental Abacus	-0.021	0.027	-0.78	.4353
Go/No Go					
	Intercept	0.730	0.017	42.03	<.0001
	Second Grade	0.066	0.017	3.98	.0001
	Post-Test	0.041	0.014	3.04	.0024
	Mental Abacus	0.012	0.019	0.64	.5233
	Post-Test x Mental Abacus	-0.032	0.018	-1.79	.0742

Task	Predictor	β	Std Err	t	p
Spatial WM					
	Intercept	0.296	0.020	14.44	<.0001
	Second Grade	0.049	0.019	2.59	.0095
	Post-Test	0.075	0.019	3.94	.0001
	Mental Abacus	0.023	0.022	1.02	.3098
	Post-Test x Mental Abacus	0.014	0.025	0.53	.5950

Beginning with the math measures, we did not see evidence of differential change in performance for either the in-house arithmetic or standardized WJ-III measures. This result differs from the findings of Barner et al. (2016), where MA-related improvements on these measures emerged numerically after a single year of training. We discuss possible reasons for this disparity below. We did see a numerical trend towards the predicted time x condition interaction for the place-value measure, however. Students in the MA condition tended to make a larger gain in place value scores over the course of the study than those in the control group. This result was marginal in the relevant mixed effects model (p = .052) so we interpret it with caution. Nevertheless, it is consistent with a similar trend in Barner et al. (2016), and appears to emerge in both Grades 1 and 2. Note that, in exploratory t-tests, we found a significant post-test difference between intervention groups (t(153.11) = -1.97, p = .05). This test was not significant for first-graders alone, (t(50.42) = -0.51, p = .61) but was for second-graders (t(95.61) = -2.24, p = .03).

In the cognitive measures, we did not see evidence of differential changes in performance for either matrix reasoning or spatial working memory. These results are consistent with our previous findings and suggest again that MA-related changes to spatial working memory do not result from even extensive MA training in typical groups of school children.

We did, however, find an unpredicted negative interaction of time and condition, such that students in the control group appeared to increase more in performance on the Go/No Go task. One possible explanation for this finding is a speed / accuracy tradeoff such that children in the MA group were less accurate but faster (perhaps stemming from the fact that MA training focuses on increasing computation speed, sometimes even at the expense of accuracy, such that computations are executed before the abacus "fades" from the mind's eye). This explanation appears plausible given a visual inspection of the reaction times (Figure 6). To further probe this possibility, we performed an exploratory analysis in which we re-ran our planned linear mixed effects model on Go/No Go accuracy scores but this time included a main effect of reaction time, to control for the different average timing of participants' responses on correct trials. Consistent with the existence of a speed / accuracy tradeoff, the magnitude of the time x condition interaction was now reduced by an order of magnitude and was no longer significant ($β$ = −0.005, p = .668). Thus, we believe the Go/No Go effect reflects a shift in a speed / accuracy tradeoff rather than a true change in cognitive functioning in the Control group.

In sum, we saw limited evidence for the effectiveness of the MA intervention. In the math tasks, only the place value measure showed a hint of an intervention effect. And in the cognitive tasks, there were no intervention effects except for a possible shift in response criterion on the Go/No Go task.
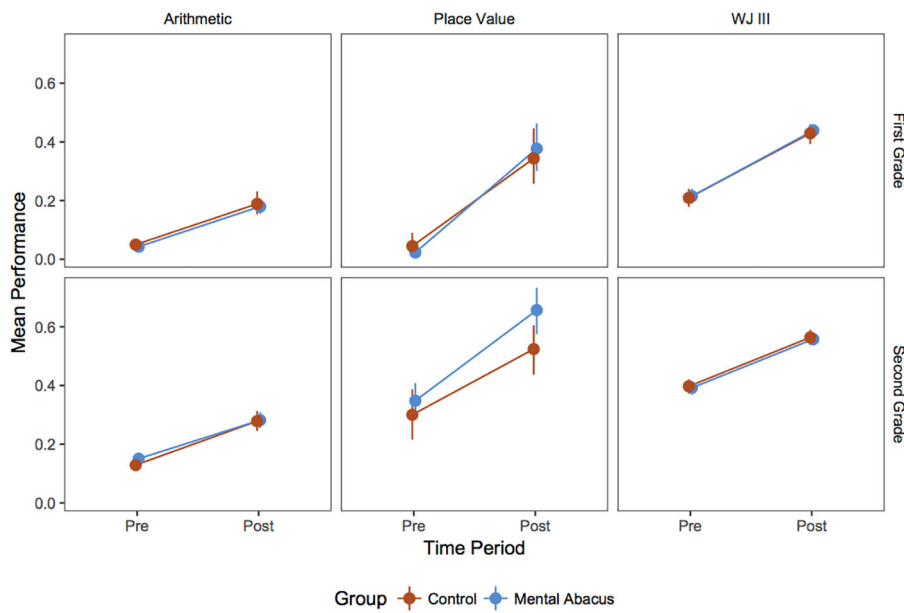
Figure 4. Performance on mathematics measures by time and grade. Error bars show 95% confidence intervals, computed by non-parametric bootstrap.
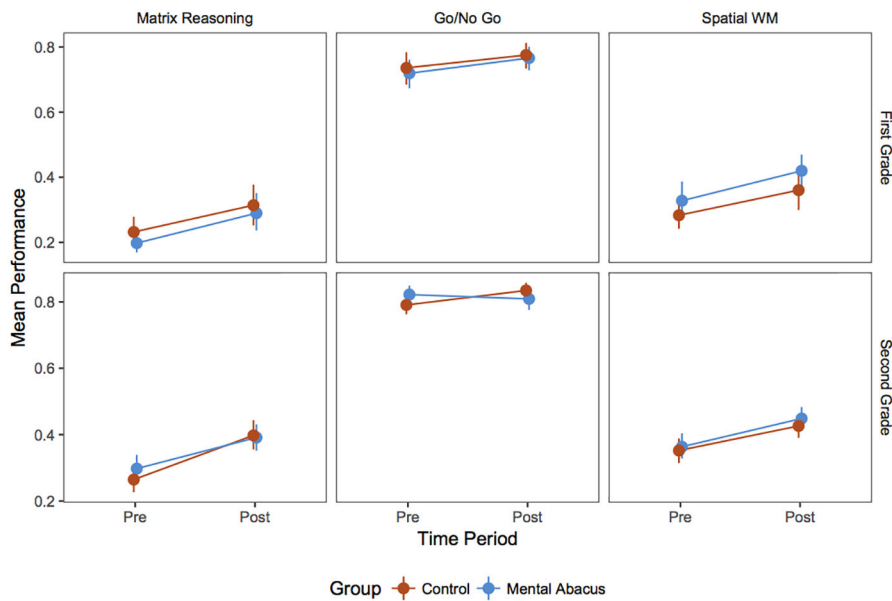

Figure 5. Performance on cognitive measures by time and grade. Error bars show 95% confidence intervals, computed by non-parametric bootstrap.
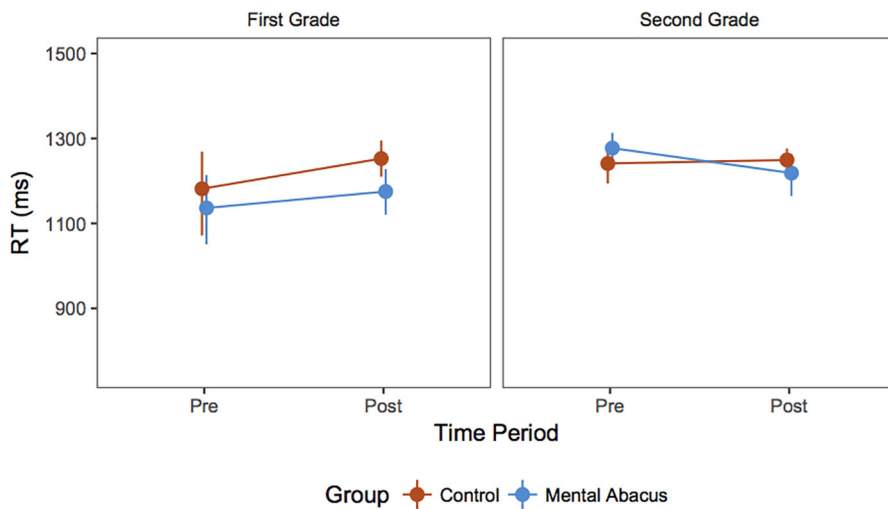
Figure 6. Reaction times for the go/no-go task. Plotting conventions are as above.

Secondary Analyses

Spatial Working Memory Analysis

In our previous study, we found that spatial working memory scores at study initiation moderated the effects of the intervention. Children who were above the median in spatial working memory tended to show the largest gains in arithmetic performance from studying MA. We pre-registered this confirmatory hypothesis and conducted it on all three of our math measures (Figure 7).
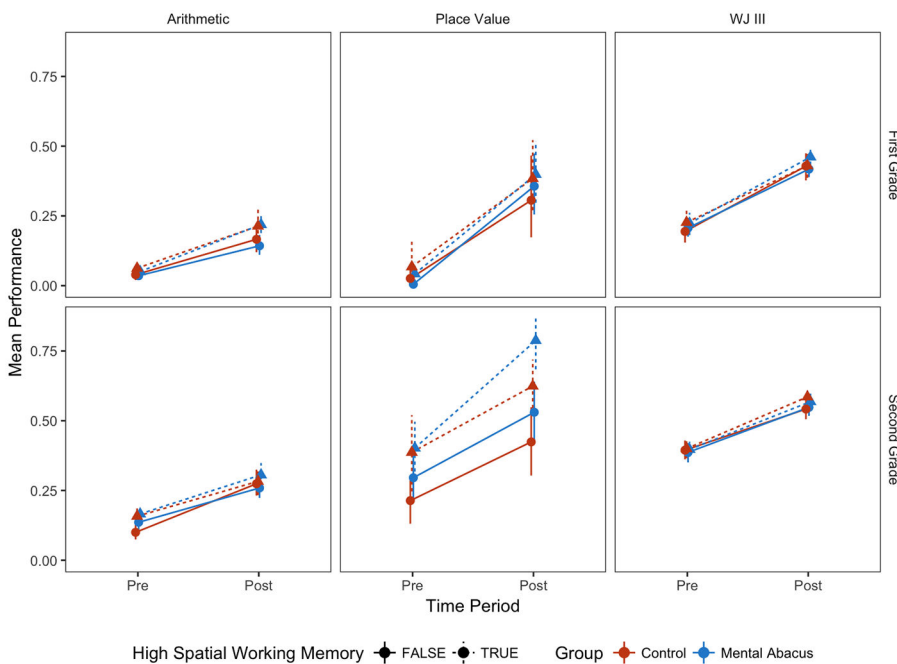

Figure 7. Median split results for outcome measures based on intervention group and spatial working memory scores. Plotting conventions are as above.

Of the three, only place value showed the predicted pattern, and only for the second graders – as might be expected based on the limited uptake among first-graders. Numerically, the pattern for place value was similar to what we observed in the arithmetic measure in our first study: Greater growth for high spatial WM children in the MA group. Nevertheless, in exploratory mixed-effects models, the three-way interaction of spatial working memory, time, and condition was not significant. Likely our study would have required considerably more power to detect such an effect.

Math Anxiety

We further assessed whether the MA intervention led to changes in math anxiety at the end of the study via an exploratory analysis. As shown in Figure 8, though first-graders showed overall more math anxiety than second graders, there were no significant differences in math anxiety between groups.
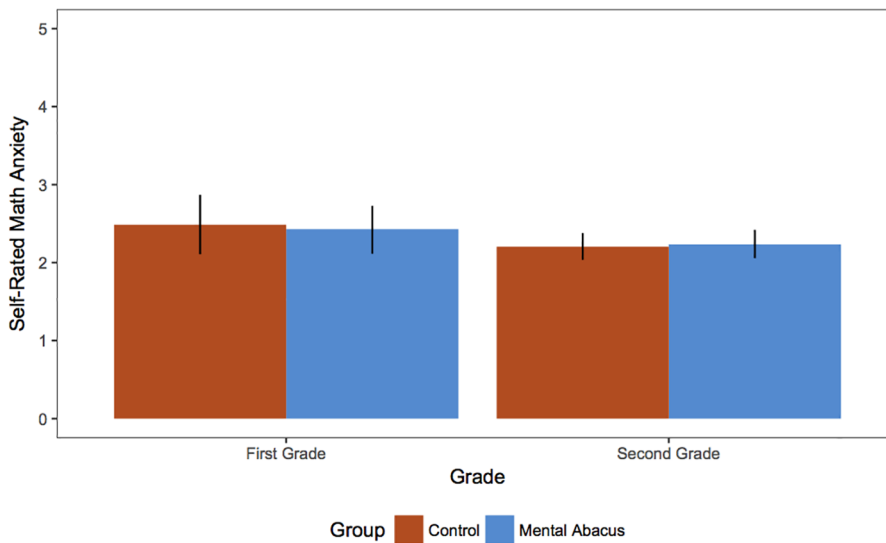

Figure 8. Average math anxiety ratings (on a five-point Likert scale), split by group and grade level. Error bars show 95% confidence intervals computed by non-parametric bootstrap.

General Discussion

We investigated the potential effectiveness of mental abacus (MA) instruction in the US context by conducting a one-year, classroom-randomized trial of MA. The study had two main goals. First, we asked whether a large group of US school children, distributed across a number of classrooms, would experience greater benefit from MA training than from equivalent hours of standard math curriculum. Second, we explored the claim, made by several recent studies, that even small amounts of MA training might yield benefits to visuo-spatial working memory abilities. Rather than being a "best case" implementation of MA, our study reports a relatively realistic trial, in that – with the exception of classroom randomization – its implementation was planned and executed by the school and a private MA curriculum vendor with little researcher input.

Our study yielded two main results. First, we found that the first-grade children in our study generally failed to master the abacus as a format for numerical representation: Even after a year of weekly training, they struggled to translate numbers between abacus and numeral formats. Perhaps this difficulty was due to their very limited

place-value understanding at study initiation (as shown by their floor-level performance on this measure). Regardless of its source, this limitation meant that they were missing one of the prerequisites for MA learning, thereby compromising their ability to experience other gains from the technique. In addition, they were not able to use the technique to construct an understanding of place value. Second, we found that, for the second-graders, one year was sufficient to acquire a basic understanding, but this level of mastery yielded only the beginnings of measurable benefits to mathematics achievement. We saw hints of improvement in place value understanding in the second grade MA group, but no effects on arithmetic performance more broadly. Overall, we did not find evidence that one year of training was sufficient to augment working memory or result in other measurable changes in cognitive tasks (e.g., reasoning or go/no-go).

These results have several implications. First, they suggest that there is prerequisite knowledge for students to learn MA. Although our study does not reveal precisely what these prerequisites are, basic numerical concepts, basic understanding of arithmetic operations, and some foundational understanding of place-value are all potential candidates. If these prerequisites are not satisfied, MA may be difficult for children to master in typical classroom environments. Thus, starting MA at an age before these prerequisites are in place may not be effective. Second, they suggest that although longer term interventions may benefit children, short-term MA interventions may be less likely to provide an advantage to children unless (1) they more intensive than the intervention pursued here, or (2) they are administered to better-prepared and/or older students. While this result may be interpreted as a negative finding, it might also be taken as evidence that alternative approaches like MA – which children may potentially find more enjoyable than existing techniques – can be deployed without sacrifice to learning (since the MA group was not numerically or statistically *worse* than control). Despite focusing heavily on procedural learning of basic arithmetic operations like addition and subtraction, MA resulted in similar gains in knowledge of place value and arithmetic, relative to our control curriculum, which was designed to augment conceptual understanding.

Finally, a third implication of this work relates to the potential non-mathematical benefits of MA training. One common claim made by developers of MA curricula is that the technique not only accelerates learning of mathematics, but also promotes the growth of domain-general cognitive capacities like memory and attention. Consistent with these claims, several recent studies have reported remarkable changes in spatial working memory capacity after only brief experience with MA technique (e.g., Dong et al., 2016; Na et al., 2015; Wang et al., 2015). Casting doubt on such findings, however, are studies that fail to find reliable transfer of working memory training to other tasks (Melby-Lervåg & Hulme, 2013; Redick et al., 2013). Consistent with this worry, our study suggests that, to the extent these findings are reliable, they may not extend beyond tightly controlled laboratory settings, or may be reserved to adult learners. Drawing on a larger sample size than in previous studies, we find that students fail to show growth in spatial working memory as a function of MA training. This result suggests that previously reported failures to find effects on spatial working memory in Indian students was not unique to that population (e.g., because of their unusually low spatial working memory at the onset of the study; Barner et al., 2016). Instead, in a standard US classroom setting, MA provides a useful tool for supplementing instruction – equal to other methods and perhaps with some benefits to older students or students with better spatial working memory – but is not a tool for making children "smarter."

One question that remains unanswered by this study is whether MA might yield a significant advantage over alternative techniques in a more protracted or in-depth intervention. Past research found that the largest benefits emerged after three years of training – the amount of time it takes typical children to complete most

existing MA curricula. Given that the first year of MA training focuses extensively on physical abacus training, and less so on mental computations, the possibility of greater gains after more training seems plausible. Data from the present study do not give us reason to doubt that US children would perform differently from Indian children given additional training. In fact, the strong spatial working memory of children in the US sample suggests that advantages might emerge earlier if training continued (at least for the second graders). Still, additional evidence is required before MA can be recommended as superior to existing techniques.

In addition, because MA requires specialized teacher training, future studies should explore instructional components. One open question in this area is the amount of teacher training necessary for effective instruction – since teachers in our school appear to have found the technique challenging and required unplanned mid-year visits from the curriculum provider. A second question is how easily MA technique can be transmitted across teachers, such that instruction can persist within schools without incurring additional training costs each new year. Because MA is an unfamiliar technique in US schools, teacher training is a major challenge to implementing the technique in the classroom. Previous studies leveraged contexts in which MA instruction was already relatively common (Barner et al., 2016; Stigler et al., 1986), making it possible that they understated the difficulties associated with implementing an MA intervention in new contexts.

In summary, we found that MA instruction led to results comparable to our control group after one year of training. Although MA instruction has led to mathematical performance gains in past work, it may only produce such results after a longer intervention or in populations with stronger prerequisite mathematical or cognitive abilities. Further, consistent with past results, we find that MA training does not augment children's pre-existing cognitive capacities.

Supplementary Materials

Cognitive Assessment Tasks

All tasks are visible at https://www.testmybrain.org/launch/mfrank.html (press control+z to initiate tasks).

Pre-Registration

Pre-registration information for our study can be found at https://osf.io/st2jn

Data and Code

All data and analytic code for the project can be found at https://github.com/langcog/majic

Competing Interests

The authors have declared that no competing interests exist.

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 judgments. *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.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. doi:10.1016/j.jml.2012.11.001

Barner, D., Alvarez, G., Sullivan, J., Brooks, N., Srinivasan, M., & Frank, M. C. (2016). Learning mathematics in a visuospatial format: A randomized, controlled trial of mental abacus instruction. *Child Development, 87*, 1146-1158. doi:10.1111/cdev.12515

Brooks, N. B., Barner, D., Frank, M., & Goldin-Meadow, S. (2017). The role of gesture in supporting mental representations: The case of mental abacus arithmetic. *Cognitive Science*. Advance online publication. doi:10.1111/cogs.12527

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*(1-2), 46-51. doi:10.1016/j.neulet.2006.04.041

Corkum, P. V., & Siegel, L. S. (1993). Is the Continuous Performance Task a valuable research tool for use with children with Attention-Deficit-Hyperactivity Disorder? *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 34*(7), 1217-1239. doi:10.1111/j.1469-7610.1993.tb01784.x

Dong, S., Wang, C., Xie, Y., Hu, Y., Weng, J., & Chen, F. (2016). The impact of abacus training on working memory and underlying neural correlates in young adults. *Neuroscience, 332*, 181-190. doi:10.1016/j.neuroscience.2016.06.051

Du, F., Chen, F., Li, Y., Hu, Y., Tian, M., & Zhang, H. (2013). Abacus training modulates the neural correlates of exact and approximate calculations in Chinese children: An fMRI study. *BioMed Research International, 2013*, Article 694075. doi:10.1155/2013/694075

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General, 141*, 134-149. doi:10.1037/a0024427

Hanakawa, T., Honda, M., Okada, T., Fukuyama, H., & Shibasaki, H. (2003). Neural correlates underlying mental calculation in abacus experts: A functional magnetic resonance imaging study. *NeuroImage, 19*(2), 296-307. doi:10.1016/S1053-8119(03)00050-8

Hatano, G., Amaiwa, S., & Shimizu, K. (1987). Formation of a mental abacus for computation and its use as a memory device for digits: A developmental study. *Developmental Psychology, 23*(6), 832-838. doi:10.1037/0012-1649.23.6.832

Hatano, G., Miyake, Y., & Binks, M. G. (1977). Performance of expert abacus operators. *Cognition, 5*, 47-55. doi:10.1016/0010-0277(77)90016-6

Huang, J., Du, F. L., Yao, Y., Wan, Q., Wang, X. S., & Chen, F. Y. (2015). Numerical magnitude processing in abacus-trained children with superior mathematical ability: An EEG study. *Journal of Zhejiang University SCIENCE B, 16*(8), 661-671. doi:10.1631/jzus.B1400287

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/hbm.20996

Ifrah, G. (2000). *The universal history of computing: From the abacus to quantum computing.* New York, NY, USA: John Wiley & Sons.

Li, S. (1959). Origin and development of the Chinese abacus. *Journal of the Association for Computing Machinery, 6*(1), 102-110. doi:10.1145/320954.320962

Li, Y., Chen, F., & Huang, W. (2016). Neural plasticity following abacus training in humans: A review and future directions. *Neural Plasticity, 2016*, Article 1213723. doi:10.1155/2016/1213723

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

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. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*(2), 270-291. doi:10.1037/a0028228

Menninger, K. (1969). *Number words and number symbols: A cultural history of numbers* (Paul Broneer, Trans.). New York, NY, USA: Dover Publications.

Miller, K. F., & Stigler, J. W. (1991). Meanings of skill: Effects of abacus expertise on number representation. *Cognition and Instruction, 8*(1), 29-67. doi:10.1207/s1532690xci0801_2

Na, K.-S., Lee, S. I., Park, J.-H., Jung, H.-Y., & Ryu, J.-H. (2015). Association between Abacus training and improvement in response inhibition: A case-control study. *Clinical Psychopharmacology and Neuroscience, 13*(2), 163-167. doi:10.9758/cpn.2015.13.2.163

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., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* San Antonio, TX, USA: Harcourt Assessment.

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*(2), 359-379. doi:10.1037/a0029082

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest, 17*(3), 103-186. doi:10.1177/1529100616661983

Stigler, J. W. (1984). "Mental abacus": The effect of abacus training on Chinese children's mental calculation. *Cognitive Psychology, 16*, 145-176. doi:10.1016/0010-0285(84)90006-9

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., 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, C., Geng, F., Yao, Y., Weng, J., Hu, Y., & Chen, F. (2015). Abacus training affects math and task switching abilities and modulates their relationships in Chinese children. *PLOS ONE, 10*(10), Article e0139930.

PsychOpen
publishing psychology