Scaling cognitive science with classroom games for learning mathematics

Joshua T. Dean[1], Moira R. Dillon[2], Esther Duflo[3,4]*, Harini Kannan[3], Elizabeth S. Spelke[5]*

[1] Booth School of Business, University of Chicago, Chicago, IL, USA
[2] Department of Psychology, New York University, New York, NY, USA
[3] Abdul Latif Jameel Poverty Action Lab (J-PAL) South Asia, New Delhi, India
[4] Department of Economics, Massachusetts Institute of Technology, Cambridge, MA, USA
[5] Department of Psychology, Harvard University, Cambridge, MA, USA

*Correspondence to: eduflo@mit.edu or spelke@wjh.harvard.edu

Abstract

We report two randomized evaluations of educational interventions targeting children in low-income neighborhoods of Delhi, India. Four math games, played in social groups, associated early emerging and universally intuitive concepts of number and geometry with the mathematical language, symbols, and operations taught in Delhi government schools. The first experiment, conducted in 231 NGO-run preschool classrooms (1,986 children) showed that the games enhanced children's mastery of mathematics both immediately after the intervention and a year later, after children's first year of primary school. The second experiment, conducted in 141 kindergarten and first-grade classrooms in government schools (2,828 children), with no extra personnel, showed that the games remained effective when played cooperatively by larger groups of children, led by their regular teachers, during time allotted for math instruction. These experiments show how insights from basic research on children's mathematical intuitions can be leveraged to design ready-to-scale interventions for the foundational years.

Main text

Despite nearly universal enrollment in primary school, learning outcomes for reading and mathematics remain poor in many low- and middle-income countries. In rural India, for example, despite some progress in the previous year, 33.7% of children attending third grade in 2024 could not perform the subtraction operation taught in second grade[1]. Part of the reason for this achievement gap is the vast heterogeneity in children's level of math knowledge at school entry, which, combined with a strict curriculum[2], leaves many children behind. Such results have stimulated a movement to create preschools that will prepare children for learning in school. Despite much discussion of the importance of learning in the preschool years, however, there is no standard preschool curriculum or agreement on how to design one. Research in developmental cognitive science has long focused on cognitive development during these early years, but it has not been systematically translated into curricula that have been tested or implemented at scale[3].

In this paper, we develop and test a suite of game-based math interventions that builds on 50 years of research in the developmental cognitive sciences. The curriculum centers on activities that can be implemented at scale in early primary school classrooms in resource-poor settings, with few additional resources and no additional teaching staff. We test the effectiveness

of this curriculum on children's learning of school mathematics in two randomized experiments spanning three grades: preschool, kindergarten, and first grade. To do so, we assembled an interdisciplinary team of cognitive psychologists, economists, and researchers with extensive experience with governments, and we worked in collaboration with Pratham, India's largest education NGO, and with the Delhi government.

Behavioral and brain-imaging experiments in developmental cognitive science reveal that an intuitive grasp of number, geometry, and language emerges in infancy and grows over the preschool years. For example, infants from birth to 3 months have a sense of the approximate number of objects in a spatial array or actions in a sequence of events[4–6], and they recognize objects based on their shapes[7–10]. Toddlers spontaneously navigate by the geometry of the surrounding spatial layout[11–13], hone their representations of object shapes[9], begin to learn words for small numbers[14], and engage in arithmetic and algebraic reasoning well before entering school[15,16]. These early emerging numerical and spatial abilities are attested in diverse cultures and circumstances, including societies with little-to-no formal education[17–22].

Nevertheless, many children struggle to connect their intuitive math concepts to the language, symbols, and operations of school mathematics, and treat counting as a rote-learned sequence, much like "eeny meeny miny mo"[23–26]. Children's symbolic math concepts are strengthened, however, when they play games that pair intuitive math concepts with mathematical language and symbols[27]. When children learn math in school, moreover, their intuitive math concepts grow stronger[28], and as adults, brain regions responsive to arrays of dots that change in number respond to changes in number symbols as well[29]. Under some circumstances, therefore, intuitive and symbolic mathematical concepts synergize.

This synergy is nevertheless not self-evident. In a prior field experiment[30], a game-based curriculum focused on intuitive math concepts durably improved preschool children's intuitive math both immediately and one year later, relative to both passive- and active-control groups, but the game training showed little evidence of an effect on children's mastery of school mathematics. Prior to India's adoption of the New Education Policy emphasizing active and meaningful learning[31], symbols were taught through exercises such as counting by ones, twos, and threes and memorizing arithmetic facts by rote repetition. Children whose family members had minimal education in mathematics, therefore, were apt to learn to manipulate mathematical symbols by rote, endowing the symbols and words with no meaning. Indeed, many children who sell in markets and demonstrate a solid understanding of arithmetic in that context fail to leverage that understanding for learning school mathematics[32].

By contrast, preschool children in middle-class families typically learn to count[14], name small numbers and familiar shapes[23,33], and interpret a variety of spatial symbols using a variety of pictures, objects, and maps[34–36]. By talking with mathematically knowledgeable parents and playing card or board games, children may gain the meaningful, extended practice that symbol mastery requires. As children's language and symbolic skills improve, they use these media to further learn from parents, preschool teachers, and peers[37,38]. Children's subsequent success at learning mathematics in school is related to the mathematical language and symbols that they encounter at home[39–42]. In accord with these findings, a recent field experiment in low-resource communities in Ghana reveals that children whose mothers received a scholarship for secondary school that increased their chance to graduate from secondary school by 25 percentage points show higher numerical and spatial reasoning abilities when tested at 5 and 7 years, at and after the onset of formal math instruction[43].

Our game-based curriculum is based on the premise that similar experiences can be created in preschool and primary school classrooms, including in resource-poor areas where informal, home-based math activities may not occur. School-based math games may be especially effective if they pair mathematical symbols and language with images that elicit intuitive math concepts, and if they are played by children who seek and provide help to one another. Accordingly, we created a suite of card and board games, exercising both intuitive and symbolic skills, for preschool, kindergarten, and first-grade children in poor neighborhoods of Delhi, India, informed both by the methods and findings of Dillon et al.[30] and by the extensive research on infants' and children's developing knowledge of the language and symbols of number and geometry[14,15,28,44–46].

To go from the lab to scale and test the feasibility and effectiveness of our game-based curriculum in real-world conditions, we proceeded in two steps. First, following the design of Dillon et al.[30], we conducted a field experiment in NGO-run preschools to test the impact of three math-games curricula on children's immediate and future learning of the mathematics taught in their schools. Then, we developed procedures to allow the most effective curriculum to be introduced and overseen by the regular teachers in kindergarten and first-grade classrooms in government schools, after minimal teacher training and with no external support. The second experiment evaluated the effectiveness of this teacher-led program to determine whether the intervention could be widely adopted, with minimal changes, if found to be successful.

Experiment 1

Experiment 1 evaluated three math-games interventions in preschool classrooms run by Pratham, an education NGO.

The interventions were led by a facilitator and focused on the subset of children who were expected to progress to primary school the following year. They began with brief, formal teaching of numerical and spatial language and symbols, followed by one of three math-games curricula.

Pretraining

The children received pretraining to ensure that they were exposed to the minimum level of symbolic representation required to engage with the symbolic content in the games. To support and maintain their learning during gameplay, pretraining was accompanied by posters that paired the symbols with intuitive images (Figure 1). These posters remained present throughout the intervention, and children were encouraged to consult them during gameplay, either to check their accuracy or to reinforce their understanding of the symbols used in play.

Math games

After pretraining with basic symbolic representations using the poster in Figure 1, schools were randomly assigned to one of the four conditions: a no-treatment control condition or one of the three math-games conditions. Each math-game condition presented four games, two focused on number and two focused on geometry (Figures 2 and 3).

In one condition (Intuitive, 58 schools, 506 children), children played the games with cards and boards displaying images that exercised universal intuitive concepts of number and geometry (e.g., an image comparing groups of 7 vs. 12 dots), a replication of Dillon et al.[30] with the addition of the symbolic pretraining. In a second condition (Symbolic, 58 schools, 478

children), children played the games with cards presenting mathematical symbols rather than intuitive images (e.g., an image displaying the numerals 7 vs. 12), accompanied by the language of the pretraining. In a third condition (Mixed, 58 schools, 505 children), children played with the intuitive and symbolic games on alternating days, using half of the materials from each of the other conditions. All the games were created with inexpensive, locally produced materials and administered by locally hired adults who received six days of training. Finally, a control group received the standard Pratham curriculum (No Treatment, 57 schools, 497 children).

Assessments and analyses

We measure children's performance on a variety of intuitive and symbolic math tests before the intervention (baseline) and at three points after it ended: 0–3 months later, during the break before the next school year (Endline 1); 6–9 months later, in the middle of the first year of primary school (Endline 2); and 12–15 months later, at the end of the first year of primary school (Endline 3). Half of the assessments focus on children's intuitive math concepts, and half on their symbolic concepts (see SI). They were presented on a laptop computer by a separate team of local nonexperts who were unaware of the children's treatment assignments. Of the children surveyed at baseline, 94% were surveyed at Endline 1, 88% at Endline 2, and 84% at Endline 3. This small attrition is balanced by baseline characteristics (See Table S21).

Our primary regression specification is **(1)**, where an outcome for child i in preschool j is regressed on indicators for the treatment and controls for age, gender, and baseline score.

$$(\mathbf{1})\ y_{i,j} = intuitive_j + symbolic_j + mixed_j + age_{i,j} + gender_{i,j} + baseline_{i,j} + \varepsilon_{i,j}$$

Our main outcomes are composite indices of overall, intuitive, and symbolic math performance, computed as averages of Z-scores on the relevant tests; coefficients thus represent treatment effect sizes (see Methods for full analysis details). Distributions of the three primary outcome indices show that scores spanned the response range without notable floor or ceiling effects in either experiment (Extended Data Fig. 5).

We predicted: (a) greater improvement overall for the children in the mixed-games condition; (b) greater improvement on the intuitive measures for the children in both the mixed and the intuitive conditions; and (c) greater improvement on the symbolic measures for the children in both the mixed and the symbolic conditions. We made no predictions about the relative impacts of the mixed-games and pure-games conditions, as the greater demands placed on children in the mixed-games condition (i.e., to master two formats, intuitive and symbolic, with half the material available for engaging with each) might either have hampered their learning (because the two formats competed for their time) or enhanced it (if the two formats were mutually reinforcing). We also made no predictions regarding gender effects, either on children's initial knowledge of number and geometry or on their learning from the math games and from school instruction. We include exploratory analyses of treatment effect heterogeneity by gender in both experiments, motivated by policy concerns about gender gaps in math education. We do not find that gender moderates our effects in either experiment (Extended Data Table 1).

Results

Table 1 presents the characteristics of the sample at baseline, and Table 2 presents the results from each condition. Each entry in Table 2 shows the effect size, t-test statistic, confidence interval, and p-value for the treatment in the row on the outcome in the column. In the intuitive condition, we replicate the three principal findings of Dillon et al.[30]: (1) The intuitive math games have an immediate impact on children's intuitive math abilities (an improvement of 0.199 standard deviations compared with the no-treatment control, $t(229) = 3.86$, $p < 0.001$), on their symbolic math abilities (coefficient = 0.155, $t(229) = 3.52$, $p = 0.001$), and overall (coefficient = 0.171, $t(229) = 4.13$, $p < 0.001$) at Endline 1, before the onset of primary school. (2) Children show a durable benefit of the intuitive math treatment on their intuitive grasp of mathematics: At Endline 3, more than a year after the end of the intervention, children who played the intuitive games continue to outperform their control-group peers on the intuitive composite measure, although the effect sizes decline from the first endline (coefficient = 0.092, t(229) = 2.17, p = 0.031). At the second endline the effect size is similar to the third endline, but is not significant at conventional levels (coefficient = 0.098, $t(229) = 1.78$, $p = 0.077$). (3) Children's symbolic performance declines from 0.155 ($t(229) = 3.52$, $p = 0.001$) at Endline 1 to 0.095 ($t(229) = 2.17$, $p = 0.031$) at Endline 2, to a nonsignificant value of 0.047 (t(229) = 1.19, p = 0.237) at Endline 3. We do not find evidence that the intuitive condition durably enhanced children's mastery of the symbols used in the primary school math curriculum.

The effects on the intuitive composite are smaller in this experiment than in Dillon et al.[30], likely due to higher performance in the control group: pooling across items and controlling for item fixed effects, we find that the control in this experiment performs 3.7 percentage points higher at baseline compared with the Dillon et al.[30] control-group mean of 39% correct at baseline ($p = 0.001$) and 6.5 percentage points higher at the first endline compared with the Dillon et al.[30] mean of 51.5% correct ($p < 0.001$). These findings may reflect improvements in the educational environment, including increased enrollment in private preschools and anecdotal reports that some activities from Dillon et al.[30] were adopted and used by the Pratham preschool teachers in regular teaching. Since the present study was conducted two years after Dillon et al.[30], it is unlikely that this improvement is due to any overlap in participants.

The symbolic treatment, by contrast, shows little evidence of an effect on the intuitive composite at any endline, but it produces an enduring improvement in the symbolic composite and, as a result, on the overall composite (See Table 2). The lack of evidence for impact on children's intuitive abilities is surprising because meaningful learning of math symbols typically enhances children's and adults' intuitive math abilities[45,47,48]. Thus, the symbolic condition may have encouraged children to learn the mathematical words and symbols by rote, without mapping them to their intuitive numerical or geometric concepts.

Finally, the mixed-games condition succeeds on both intuitive and symbolic measures at the first endline. Children in the mixed-games condition have significantly higher scores on the intuitive composite (coefficient = 0.203, $t(229) = 3.83$, $p < 0.001$) and the symbolic composite (coefficient = 0.189, $t(229) = 3.91$, $p < 0.001$) compared with the no-treatment control. Notably, these children perform as well as children in the intuitive condition on the intuitive composite ($F(1, 229) = 0.01$, $p = 0.939$) and as well as children in the symbolic condition on the symbolic composite ($F(1, 229) = 0.01$, $p = 0.940$). At the second endline, the effects are lower and not significant on either the intuitive (coefficient = 0.082, $t(229) = 1.51$, $p = 0.132$) or symbolic (coefficient = 0.057, $t(229) = 1.16$, $p = 0.248$) composites. However, at the third endline, the effects are significant again on both the intuitive (coefficient = 0.102, t(229) = 2.24, p = 0.026)

and symbolic (coefficient = 0.084, t(229) = 1.99, p = 0.048) composites, at similar levels to those of children in the other two conditions (test for equality with intuitive condition: F(1, 229) = 0.06, p = 0.808; test for equality with symbolic condition: F(1, 229) = 0.12, p = 0.725).

Discussion

Experiment 1 provides evidence that all three math-games conditions have beneficial and enduring effects on children's math learning. Replicating Dillon et al.[30], the intuitive games have immediate but declining effects on symbolic math learning, even when preceded by training on math symbols. By contrast, symbolic games training has an enduring effect on the symbolic measure, but there is little evidence it affects children's intuitive mathematical abilities at any time point. Because meaningful symbolic activities enhance children's intuitions about number and geometry, these findings raise the possibility that the symbolic math games connect mathematical language to spatial symbols but not to children's intuitive grasp of mathematics. Finally, alternating training with the intuitive and symbolic games produces significant impacts on all measures at the first and final endlines.

When children who received the mixed games are compared with those who received the same games in a purely intuitive or symbolic format, the effect sizes are similar even though children who played the single-format games received twice as much time and material in that format (see SI). Thus, alternating exercise of intuitive and symbolic skills in one set of games can produce gains in both intuitive and symbolic math knowledge without requiring any additional time. These findings suggest strong complementarities for the paired intuitive and symbolic format of the games.

Consistent with that suggestion, the effect sizes generated by the mixed condition are fairly large, relative to what is typically found in similar interventions. A meta-analysis of education interventions in low- and middle-income countries identified a median effect size of 0.07 standard deviations for mathematics[49]. By their measures, the mixed condition's effect size at Endline 1 places it between the 80th and 90th percentiles of effects on math learning. Moreover, point estimates were larger at Endline 3, whereas for most educational interventions, there is rapid decay after the intervention stops. This persistence may be because the intervention inherently linked fundamental intuitions with the symbols that continued to be used regularly in the classroom after the intervention ended.

Experiment 2

Given the long-term positive impacts of the mixed-games condition in Experiment 1, we sought to develop a scalable version of that curriculum, suitable for implementation in government schools without additional resources. This development required several modifications to the game designs from Experiment 1. First, the games needed to be easy for teachers to understand, with limited documentation and training. Second, children needed to be able to play them in much larger groups, with minimal teacher input or individual supervision. Third, the games needed to complement the local curriculum so that teachers could work them in among their other activities without losing time needed for the curriculum or increasing their teaching load.

To meet these conditions, we conducted two years of systematic piloting in government classrooms. First, we modified the content in the mixed-games condition to facilitate its integration with the existing kindergarten and first-grade curricula. Second, we reorganized the

gameplay so that classes of 20–40 children could play in parallel, overseen by the regular teacher. Children were divided into groups of four to five, and each group was given a single set of game materials. All groups played in parallel, cooperating within the group to maximize the speed and accuracy with which the group completed the game (Figure 4). The teacher was available to answer questions, resolve disputes, correct the group's performance, and determine and reward the group with the fastest and most accurate performance on each round (more details on the modifications of the games for Experiment 2 are provided in the SI).

These changes risked Experiment 2's intervention resulting in a lower impact than those of Experiment 1 and Dillon et al.[30] specifically because gameplay occupied some of the time that otherwise would have been devoted to the regular math curriculum and because small groups of children played with no facilitator to guide them. To reduce this latter disadvantage, gameplay was preceded by the pretraining and supported by the pretraining posters that were available throughout the course of the intervention (Figure 4a). Throughout gameplay, teachers encouraged the students to consult the posters or confer with their classmates if they were unsure of their next move in a game. Thus, students were empowered both to teach and learn from one another and to address the gaps in their own knowledge.

Experiment 2 was conducted in the kindergarten and first-grade classrooms of 141 public schools in Delhi during the second half of the academic year. Schools were randomized either to receive the modified version of the mixed games to be played in free periods or during times devoted to math instruction (71 schools) or to receive the school's regular math curriculum (70 schools). This resulted in a final sample of 1,411 children in kindergarten (90% between the ages of 4.5 and 5.4 years) and 1,417 children in first grade (90% between the ages of 5.5 and 6.5 years).

We measure children's performance at two time periods: before the intervention (baseline) and 0–3 months after the intervention ended (Endline). For kindergarten children, the tests are like those of Experiment 1. For first-grade children, both number and geometry tests contain harder problems (e.g., higher numbers), consistent with the children's school math curriculum. We test 1,357 of the kindergarten children (96.2%) and 1,344 of the first-grade children (94.8%) at the endline. This small attrition is balanced by treatment status (See Table S21).

Analyses

We use specification (2), run separately for kindergarten and first grade. Outcomes and analytic approach are otherwise identical to Experiment 1 (see Methods).

$$(2)\ y_{i,j} = math_j + age_{i,j} + gender_{i,j} + baseline_{i,j} + \varepsilon_{i,j}$$

Results

Table 3 shows descriptive statistics at baseline and confirms baseline balance. Table 4 presents the results at the endline on all three composite measures. Each entry in Table 4 shows the effect size, t-test statistic, confidence interval, and p-value for the treatment on the outcome in the column. For the kindergarten children, the games result in an improvement of 0.125 standard deviations on the symbolic measure ($t(140) = 3.49$, $p = 0.001$) and 0.112 standard deviations overall ($t(140) = 3.29$, $p = 0.001$). Although there is no significant effect of the games on the intuitive measure, the effect is positive (coefficient = 0.067, $t(140) = 1.35$, $p = 0.178$). For the first-grade children, the games result in improvements on the intuitive composite (coefficient

= 0.104, $t(140) = 2.54$, $p = 0.012$), the symbolic composite (coefficient = 0.107, $t(140) = 3.03$, $p = 0.003$), and overall (coefficient = 0.110, $t(140) = 3.43$, $p = 0.001$). Distributions of the three primary outcome indices show that scores spanned the response range without notable floor or ceiling effects in either experiment (Extended Data Fig. 6).

Because the games were played in groups, it is possible that the weaker students in a group could have been excluded from play and thereby not benefitted from the intervention. To assess this possibility, we pre-specified an additional analysis that split both grade-level samples into terciles of baseline performance on the overall-math composite. We then ran the regression in **(2)** for the overall-math composite separately for each tercile while combining the kindergarten and first-grade samples to increase power (Extended Data Table 2). The point estimates are smaller in the bottom tercile (coefficient = 0.067, s.e. 0.039) than in the middle (coefficient = 0.139, s.e. 0.036) or top terciles (coefficient = 0.132, s.e. 0.034). However, we are unable to reject equality of the bottom tercile effect to the middle tercile effect (p = 0.149) or to the top tercile effect (p = 0.207).

Discussion

Experiment 2 tests whether two curricula consisting of socially engaging math games, mixing intuitive and symbolic material and played in parallel by groups of children in kindergarten or first-grade classrooms in poor areas of Delhi, enhance children's learning of math in the first years of primary school. We found that both the kindergarten and the first-grade games curricula reliably enhanced children's learning of the mathematical language, concepts, and skills taught in their schools. The first-grade curriculum also had a significant impact on children's intuitive mathematical skills.

The effect sizes in Experiment 2 are smaller than those in Experiment 1, likely for several reasons. First, the children in Experiment 1 played the games under the direct supervision of an additional adult facilitator, whereas those in Experiment 2 played with no extra personnel and far less adult attention. Second, the children in Experiment 1 played the games individually with feedback on all their moves in the games, whereas those in Experiment 2 played in groups. Third, the facilitators in Experiment 1 were specifically hired and trained to engage children in playful activities, whereas the teachers in Experiment 2 were hired and trained to implement the government's school curricula, and they likely varied in their degree of interest in the math games and their skill at introducing and overseeing the gameplay. Experiment 2 thereby tested a version of the curricula that already integrated the implementation-fidelity challenges that typically attenuate effects at scale[2,50]. Fourth, although the games in both Experiments 1 and 2 took time away from the children's regular school activities, only in Experiment 2 did these regular activities specifically include formal instruction in mathematics. Kindergarten and first-grade children's improvement in Experiment 2 thus represents the additional effectiveness of the games over the existing curricula. In this respect as well, our experiment estimates the impact in any future full-scale implementation as a part of curriculum reform.

At the time of Experiment 2, the social climate in Indian government schools may have enhanced the impact of the games. In both grades, instruction was largely teacher-directed, with children expected to listen and respond individually to the teacher's prompts. In this context, Experiment 2's class-wide gameplay had potential advantages. By administering games to the full classroom, grouped into teams that cooperated internally and competed with one another, the games likely became "gamier" than those of Experiment 1, where an adult closely supervised and interacted with each child individually. Moreover, play with peers gave children

opportunities to teach and learn from each other, a practice that has been found to benefit children both as teachers and as learners[51]. In addition, the pretrained posters allowed children to address gaps in their own knowledge during gameplay, potentially increasing their sense of agency and independence in learning. Finally, by bringing the games into the setting in which children experienced the regular math curriculum, the skills developed during gameplay and formal instruction had a greater chance to synergize, and teachers had greater opportunities to observe their students' strengths and weaknesses.

These considerations may explain the consistent, positive impact of the games on students' math performance: the Experiment 2 effects lie around the 60th percentile of effect sizes on math learning found in educational interventions across 234 studies in low- and middle-income countries[49]. These impacts were achieved with regular teaching staff, trained for just four days, and no other changes to the curriculum or classroom organization. The only significant investment was the one-time printing costs for the materials. Socially engaging math games, building on children's early emerging, universal intuitions about number and geometry and linking the materials that elicit those intuitions to mathematical language, symbols, and operations, positively impact children's school math learning.

General discussion

As all countries attempt to improve educational outcomes for poor children, a natural place to look for new tools is in the basic-science literature on children's learning. Research in developmental cognitive science reveals that intuitive concepts of number and geometry emerge in infancy[4,6,8], are attested across cultures, and guide mathematical reasoning throughout life. Effective pedagogy can build on these intuitive concepts, just as effective pedagogy to promote literacy builds on the basic science of children's learning of language and of orthographic rules[52–56]. Experiments 1 and 2 demonstrate that it is possible to develop educational interventions that leverage them to promote learning of school mathematics and deploy them at scale.

While promising, there are three key limitations to the present study. First, we were unable to collect longer-term follow-up data for Experiment 2 due to extended school closures during the pandemic, and so we cannot assess whether the intervention's effects persist as they did in Experiment 1. Second, the point estimates for children in the bottom tercile of baseline performance in Experiment 2 are smaller than those for the middle and top terciles, although not significantly so. Future work should continue to build on these results with longer follow-ups and with a focus on students who might need more support. Finally, we are unable to reliably separate the effectiveness of the number training from the geometry training. Future work should attempt to separately identify whether our approach differs in its effectiveness by content domain.

Beyond these practical implications, field testing at scale can also benefit basic research in cognitive science, neuroscience, and education by revealing the scope and limits of theories developed in the laboratory. Laboratory experiments allow the factors that influence learning to be tested individually, but they fail to reveal how the learning processes they uncover will operate in the rich, noisy environments in which children's knowledge grows. For example, experiments from several labs show short-term gains in children's performance on symbolic number tasks after brief training with intuitive number tasks[57], but neither Experiment 1 nor its predecessor[30] found enduring improvements in symbolic mathematical skills after long-term training on such tasks[30]. We hope these results will spur further theorizing in basic science labs

on the interplay between intuitive and symbolic math skill acquisition. Laboratory research in the cognitive sciences and field research in the economics of education therefore have complementary strengths and limitations, and each suggests insights to the other.

Ultimately, the impact of the present research will rely on the scaling and deployment of the effective Experiment 2 intervention. Since 2022, J-PAL has worked with Pratham and several state governments to help them adopt the games, pilot them locally, and integrate them with their curriculum and classroom activities at scale. To date, nine states have run pilots and two have begun the process of adopting the games in all their schools. The games have already reached 34,000 children, and one million are scheduled to be reached in the next few years through state-wide scale ups. This work shows the power of taking science from the lab to the field, and then from the field to scale, adapting such research for real conditions and working with governments to ensure that policy recommendations are implementable in the local context.

Methods

Ethics, consent, and preregistration

The research presented here was conducted in accordance with all relevant ethical regulations. Experiment 1 was reviewed and approved by the MIT Committee on the Use of Humans as Experimental Subjects (Protocol 1503006959) and the IFMR Human Subjects Committee. Experiment 2 was reviewed and approved by the MIT Committee on the Use of Humans as Experimental Subjects (Protocol 1805377780) and the IFMR Human Subjects Committee.

For Experiment 1, parents attended an informational session at the preschool center where research-team staff provided an overview of the project activities, and parents provided written consent for their children to participate; consent was again requested from either the classroom teacher or the parents before each endline assessment. For Experiment 2, oral consent was sought from parents at the time of study initiation, and parents received a Hindi-language circular outlining the voluntary nature of participation, data to be collected, and their option to withdraw at any time; parents who wished to opt out could return a written form declining participation, though no written refusals were received. Neither children, parents, nor schools received compensation for participation in either experiment.

Both experiments were preregistered at the American Economic Association Randomized Controlled Trial Registry. Experiment 1 was registered on 21 April 2016 as AEARCTR-0001195 (https://www.socialscienceregistry.org/trials/1195). Analysis plans for each endline were filed on August 22, 2016, March 25, 2017, and November 24, 2017, respectively. Experiment 2 was registered on 20 July 2018 as AEARCTR-0003143 (https://www.socialscienceregistry.org/trials/3143) and the endline analysis plan was filed on April 16, 2019. Sample sizes for both experiments were chosen to provide 80% power on a two-sided t-test to detect an effect of 0.20 standard deviations or larger on the composite test score outcomes. Randomization in both experiments was conducted at the school level. Participants were not blinded to treatment status, but the enumerators administering the outcome assessments were blind to treatment assignment.

Experiment 1

Games

Race to the Top (Figure 2) was played with cards, a spinner and a board. Following the procedures of Dillon et al.[30], children spun a spinner with either triangles or numerals and, based on the number where the arrow stopped, drew 1–3 cards that each displayed 0–3 triangles or the corresponding numeral. They then advanced a token on the game board by the number of spaces corresponding to the sum of the numbers on the cards.

Number Comparison (Figure 3, first two rows) was played with cards and presented two number games used by Dillon et al.[30] in a common, streamlined format. Children progressed from sorting cards by comparing the relative magnitudes of two numbers (here and below, presented as arrays of dots in the intuitive version or as numerals in the symbolic version) to sorting cards by mentally adding two numbers and then choosing which of two other numbers corresponded to the sum.

Find the Shape (Figure 3, third row) was played with cards and followed the logic of the geometry card game of Dillon et al.[30] but with streamlined materials. A shape category or property (conveyed by three example images or by a symbolic description) appeared on top of each card, and two shapes — one exhibiting the target category or property — appeared below; children sorted the cards by the color of the bottom shape that exhibited the target property.

Reading Maps (Figure 3, fourth row) followed the procedures of Dillon et al.'s[30] map reading game. While facing a large mat on the floor depicting large versions of different shapes appearing at different orientations, children drew a card depicting one shape (either iconically or symbolically) and a dot located on that shape. The card was described as a map of the mat on the floor, with the dot indicating the target location on the mat where children were tasked with placing a token.

A more detailed description of the games and of their progression over the four-month intervention appears in the SI.

Assessments

As in the experiment by Dillon et al.[30], each assessment was composed of six subtests: one test of sensitivity to relative numerical magnitudes in briefly presented arrays of dots; one test of sensitivity to geometric properties in six simultaneously presented geometric forms; two tests assessing children's mastery of words and symbols for numbers and shapes in accord with the vocabulary used by the regular preschool curriculum (e.g., object names like "samosa" for triangle and Hindi or English number words), and two tests presenting verbal problems eliciting explicit numerical or geometric reasoning. The tests were either the same as, or highly similar to those of Dillon et al.[30]. A more detailed description of the assessments appears in the SI.

Participants, setting, and procedure

Experiment 1

Experiment 1 was conducted two years after Dillon et al.[30] in the same system of private preschools operated by Pratham, a large nonprofit nongovernmental organization focused on improving and evaluating education throughout India. These preschools are typically run by Pratham volunteers out of their houses or other shared spaces. Each of the 231 participating preschools served up to 27 children ranging from 3–6 years old. One preschool experienced complete attrition before the first endline, and the analysis sample therefore comprises 230 clusters. We recruited children from each preschool who the regular teacher assessed would be graduating to primary school at the end of the academic year based on their ages. We then invited parents to an informational session where we explained the study and obtained permission for their children to participate. This resulted in a final sample of 1,986 participants with 90% children between 4.7 and 6.0 years-old. As intended, our sample progressed to primary school in the year following the intervention, with 96% reported enrolled by our third endline (12–15 months after the intervention).

The games were played in one small group per classroom during the last four months of the preschool year by a local facilitator hired for the experiment. Facilitators were trained on the games for about six days and assigned to two classrooms with the same treatment. Gameplay occurred during three one-hour sessions per week. To monitor children's progress in playing the games, a separate team of process monitors made unannounced visits to the classrooms in which games were being played, and they recorded whether the games' facilitator was present, whether

the games were played, how correctly the facilitator presented the games, how attentive the children were to the gameplay, and how accurately the children responded. This monitoring showed strong fidelity to the protocol during implementation. A more detailed description of the process monitoring appears in the SI.

Before starting regular gameplay, the games' facilitator engaged the children in pretraining sessions (Figure 1). The number pretraining introduced number words and the Arabic symbols 1–20; the geometry pretraining introduced the basic shape and geometric properties of congruence, similarity, and the sides and angles of simple polygons (Figure 1). Additional sessions were also interspersed between gameplay sessions, typically introducing new material before children encountered a new range of numbers or shape property in the games. Children were also exposed to pretraining on finding locations on shapes prior to the introduction of Reading Maps. (Figure 1). A more detailed description of the pretraining appears in the SI.

Experiment 2

Adapting the math games for use in primary schools

We conducted two years of systematic piloting in government classrooms, both to maximize the effectiveness of the mixed-games condition and to allow its integration with the kindergarten and first-grade school curricula. Because the kindergarten curriculum introduced words, numerals, and combinations of numerals from 1–20, and the first-grade curriculum introduced words and symbols from 1–99, we also reconfigured the mixed games both to help children understand the base-10 number system and to integrate the games' use of that system with the existing classroom teaching. Understanding base-10 relationships is particularly challenging for students who grow up speaking Hindi because the language obscures the base-10 structure (A. Srinivasan et al., manuscript in preparation). To aid children's learning, the mixed games paired each number symbol (i.e., each numeral or combination of two numerals) with an array of dots that were organized into circled groups of ten (Extended Data Fig. 3). In addition, while we used the same pretraining number poster as in Experiment 1 for children in kindergarten, two new posters with numbers from 1–40 and 40–100 were introduced for children in first grade (Extended Data Fig. 1).

No spatial symbols for geometric properties are part of the kindergarten or first-grade Indian curricula, but we adapted the spatial symbols used in the mixed games to capture relations that are included in school-like activities tasking children with analyzing and reproducing shapes (Extended Data Fig. 4). Geometry pretraining posters were considerably redesigned from Experiment 1, and both kindergarten and first-grade children were exposed to the same four posters (Figure 4 and Extended Data Fig. 2).

As for Experiment 1, pretraining occurred prior to gameplay as well as periodically throughout the course of gameplay, and posters were available for children to consult during gameplay (Figure 4).

We reorganized the gameplay so that classes of 20–40 children could play in parallel, overseen by the regular teacher. To this end, groups of 4–5 children were given a single set of game materials, and each child in a group was assigned a color (orange, green, purple, or yellow), which was associated with the games' color coding. For the two card games (adapted versions of Number Comparison and Find the Shape, Figure 4 and Extended Data Figs. 3, 4), children chose the cards whose borders matched their assigned color and sorted the cards in accord with the color of the correct response (red or blue). For the two board games (an adapted

version of Reading Maps and a simplified replacement for Race to the Top named Find and Move, Figures S8, S9), one card deck, two table-top boards at opposite orientations (one readable by the children on each side of their play space), and one set of colored markers were distributed such that each child who played with cards of a given color used the marker of the corresponding color to indicate the correct answer on a board. Thus, all children played in parallel, cooperating within their group to maximize the speed and accuracy with which the group completed the game. The teacher was available to answer questions and resolve disputes, to correct the group's performance (by scanning through the cards in the two boxes for the sorting games and by comparing the positions of the colored marks on the board in relation to an answer key), give corrective feedback to individual children who had sorted a card or marked the board incorrectly, and determine and reward the group with the fastest and most accurate performance on each round. Teachers could not, however, give feedback to every child after every round.

A more detailed description of the games and of their progression over the four-month intervention appears in the SI.

Assessments and timing

The assessments were similar in format to those of Dillon et al.[30] and of Experiment 1, but the symbols they used and the problem difficulty were aligned with the children's respective school curricula. A more detailed description of the assessments appears in the SI.

We originally planned to evaluate students in later grades, but in-person testing was precluded by the pandemic, during which extended school closures deprived many children of any math instruction for almost two years. Accordingly, we report only the findings from one endline.

Participants, setting, and procedure

We identified a set of government primary schools which had students enrolled in kindergarten and first grade in collaboration with the government. We sent home notices to the parents of the students enrolled in these grades in the local language, Hindi, describing the study and informing them that they could opt to have their child removed from the study by returning a signed form. We did not receive any returned forms opting out. This resulted in a final sample of 1,411 children in kindergarten (90% between 4.5- and 5.4-years-old) and 1,417 children in first grade (90% between 5.5- and 6.5-years-old).

Analyses

For Experiment 1, our primary regression specification is (1), with each outcome regressed on indicators for treatment condition and controls for age, gender, and baseline score. Main outcomes are composite indices of overall, intuitive, and symbolic math performance, computed as averages of Z-scores on the constituent tests standardized to baseline control values: tests of numerical and geometric sensitivity for the intuitive composite, and tests of mastery of mathematical language and notation (numerals, operations, and spatial symbols) for the symbolic composite. Coefficients thus represent treatment effect sizes. We cluster standard errors at the preschool level (the level of randomization), follow an intention-to-treat design (including children regardless of their school attendance), and test for significance with two-sided t-tests.

For Experiment 2, our primary regression specification is (2), run separately for the kindergarten and first-grade samples. The outcome for each child is regressed on an indicator for receiving the modified mixed games, with controls for age, gender, and the baseline value of the outcome. Composite indices are constructed as in Experiment 1, with Z-scores standardized to baseline control values. We cluster standard errors at the school level (the level of randomization), follow an intention-to-treat design, and test for significance with two-sided t-tests.

Data Availability Statement

The datasets generated and analyzed in this article are available in a Code Ocean capsule (https://codeocean.com/capsule/3743638/), currently access-restricted for peer review and to be made publicly available before publication.

Code Availability Statement

The code used to clean the data and produce the analyses presented in this article is available in a Code Ocean capsule (https://codeocean.com/capsule/3743638/), currently access-restricted for peer review and to be made publicly available before publication.

Acknowledgements

This research was supported by the UBS Optimus Foundation, the Douglas B. Marshall Jr. Family Foundation, King Philanthropies, and the Siegel Fund. The funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript. We would like to thank Olivia Fiske, Vrinda Kapur, Suvidhi Khurana, Deepika Kalra, Ruby Mittal and Madeline Duhon for excellent research assistance. We would also like to thank Rukmini Banerji, Shailendra Sharma and the staff at Pratham Delhi Education Initiative for their assistance with the first experiment, and the Education Department, Government of National Capital Territory of Delhi for assistance with the second.

Author Contributions

J.T.D., M.R.D., E.D., H.K., and E.S.S. contributed to all portions of the study.

Competing Interests

H.K. is employed by the Abdul Latif Jameel Poverty Action Lab (J-PAL), which is currently scaling implementation of the intervention based on the results of this study. The other authors declare no competing interests.

Tables

	Control	Intuitive	Symbolic	Mixed	P-Values
Demographics:					
Proportion Female	0.584	0.599	0.592	0.562	0.826
Age in Months at Baseline	62.36	62.66	62.46	62.51	0.982
Test Scores (Proportion Correct):					
Numerical Sensitivity	0.644	0.660	0.650	0.654	0.625
Geometric Sensitivity	0.274	0.256	0.274	0.277	0.452
Number Words and Symbols	0.501	0.503	0.491	0.520	0.815
Shape Names	0.472	0.506	0.485	0.473	0.230
Numerical Reasoning	0.634	0.610	0.621	0.628	0.306
Geometric Reasoning	0.530	0.521	0.518	0.526	0.916
Executive Function	0.640	0.632	0.625	0.641	0.438

Table 1. Experiment 1, characteristics at baseline*. The first four columns present means of the demographic information and baseline test scores for the children randomized to the four conditions in Experiment 1. The fifth column presents p-values for tests of joint equality of the no-treatment control, intuitive, symbolic, and mixed conditions (with standard errors clustered at the preschool level) for each measure. None of the tests are significant.*

Endline 1 (0–3 months)

	(1) All Math		(2) Symbolic		(3) Intuitive	
Intuitive Treatment (I)	0.171*** (0.089,0.252)	t(229)=4.13 [< 0.001]	0.155*** (0.068,0.242)	t(229)=3.52 [0.001]	0.199*** (0.097,0.300)	t(229)=3.86 [< 0.001]
Symbolic Treatment (S)	0.155*** (0.059,0.250)	t(229)=3.20 [0.002]	0.193*** (0.086,0.299)	t(229)=3.56 [< 0.001]	0.076 (-0.028,0.181)	t(229)=1.44 [0.151]
Mixed Treatment (M)	0.193*** (0.106,0.280)	t(229)=4.35 [< 0.001]	0.189*** (0.094,0.284)	t(229)=3.91 [< 0.001]	0.203*** (0.098,0.307)	t(229)=3.83 [< 0.001]
Hypothesis	I=S	I=S=M	S=M	I=S=M	I=M	I=S=M
F-test Stat	F(1,229)=0.13	F(2,229)=0.35	F(1,229)=0.01	F(2,229)=0.42	F(1,229)=0.01	F(2,229)=3.77
F-test P-value	P=0.718	P=0.708	P=0.940	P=0.656	P=0.939	P=0.024
Control Mean	0.484		0.531		0.390	
Observations	1,837		1,834		1,837	
Number of Clusters	230		230		230	

Endline 2 (6–9 months)

	(1) All Math		(2) Symbolic		(3) Intuitive	
Intuitive Treatment (I)	0.099** (0.015,0.183)	t(229)=2.33 [0.021]	0.095** (0.009,0.180)	t(229)=2.17 [0.031]	0.098* (-0.011,0.207)	t(229)=1.78 [0.077]
Symbolic Treatment (S)	0.058 (-0.029,0.146)	t(229)=1.31 [0.191]	0.087* (-0.006,0.181)	t(229)=1.84 [0.067]	-0.006 (-0.116,0.104)	t(229)=-0.11 [0.912]
Mixed Treatment (M)	0.067 (-0.021,0.155)	t(229)=1.50 [0.136]	0.057 (-0.040,0.154)	t(229)=1.16 [0.248]	0.082 (-0.025,0.189)	t(229)=1.51 [0.132]
Hypothesis	I=S	I=S=M	S=M	I=S=M	I=M	I=S=M
F-test Stat	F(1,229)=1.19	F(2,229)=0.69	F(1,229)=0.45	F(2,229)=0.43	F(1,229)=0.10	F(2,229)=2.23
F-test P-value	P=0.276	P=0.503	P=0.503	P=0.648	P=0.756	P=0.110
Control Mean	0.576		0.540		0.647	
Observations	1,720		1,718		1,720	
Number of Clusters	230		230		230	

Endline 3 (12–15 months)

	(1) All Math		(2) Symbolic		(3) Intuitive	
Intuitive Treatment (I)	0.066* (-0.001,0.134)	t(229)=1.95 [0.053]	0.048 (-0.031,0.127)	t(229)=1.20 [0.232]	0.095** (0.011,0.179)	t(229)=2.23 [0.027]
Symbolic Treatment (S)	0.074** (0.000,0.148)	t(229)=1.98 [0.049]	0.093** (0.008,0.178)	t(229)=2.17 [0.031]	0.040 (-0.059,0.139)	t(229)=0.79 [0.428]
Mixed Treatment (M)	0.089** (0.019,0.160)	t(229)=2.49 [0.013]	0.085** (0.001,0.169)	t(229)=1.99 [0.048]	0.100** (0.010,0.190)	t(229)=2.19 [0.030]
Hypothesis	I=S	I=S=M	S=M	I=S=M	I=M	I=S=M
F-test Stat	F(1,229)=0.05	F(2,229)=0.27	F(1,229)=0.04	F(2,229)=0.81	F(1,229)=0.01	F(2,229)=0.85
F-test P-value	P=0.815	P=0.766	P=0.842	P=0.444	P=0.909	P=0.430
Control Mean	-0.205		-0.357		0.099	
Observations	1,624		1,622		1,624	
Number of Clusters	230		230		230	

Table 2. Regressions of children's performance on the overall math composite measures in Experiment 1. *Treatment effects are estimated with (1) along with the relevant statistical tests. The indices are created by standardizing all test scores to the control condition at baseline and taking a simple average. For each outcome measure, the first column presents the treatment effects with 95% confidence intervals and stars indicating a p-value less than 0.1, 0.05, and 0.01. The second column presents a t-statistic for a test of whether the coefficient is equal to zero, with the p-value of the test in brackets. The bottom panel presents tests of equality of the different coefficients.*

The three rows respectively indicate the equality being tested, the F-statistic, and the associated p-value. All standard errors are clustered at the preschool level.

	Kindergarten			First Grade		
	Control	Treatment	P-Values	Control	Treatment	P-Values
Demographics:						
Proportion Female	0.471	0.493	0.734	0.448	0.495	0.455
Age in Months at Baseline	59.73	59.92	0.42	71.63	71.69	0.842
Test Scores (Proportion Correct):						
Numerical Sensitivity	0.639	0.640	0.953	0.732	0.714	0.066
Geometric Sensitivity	0.347	0.323	0.108	0.397	0.403	0.647
Number Words and Symbols	0.501	0.489	0.507	0.462	0.471	0.577
Shape Names	0.435	0.446	0.378	0.515	0.524	0.460
Numerical Reasoning	0.609	0.612	0.787	0.696	0.700	0.677
Geometric Reasoning	0.466	0.457	0.503	0.581	0.578	0.829
Letter Recognition	0.511	0.500	0.608	0.799	0.801	0.911
Executive Function	0.658	0.641	0.053	0.733	0.733	0.970

Table 3. Experiment 2, characteristics at baseline. *The first three columns present the results of the demographic information and baseline test scores for the kindergarten sample while the final three present those measures for the first-grade sample. The first two columns for each sample present the means and the third column presents p-values for a test of equality of the control and treatment means for each item. None of the differences are significant at the 0.05 level. Standard errors clustered at the school level.*

Kindergarten (0-3 months)

	(1) All Math		(2) Symbolic		(3) Intuitive	
Mixed Treatment	0.112***	t(140)=3.29	0.125***	t(140)=3.49	0.067	t(140)=1.35
	(0.045,0.180)	[0.001]	(0.054,0.196)	[0.001]	(-0.031,0.165)	[0.178]
Control Mean	0.391		0.262		0.711	
Observations	1,357		1,357		1,357	
Number of Clusters	141		141		141	

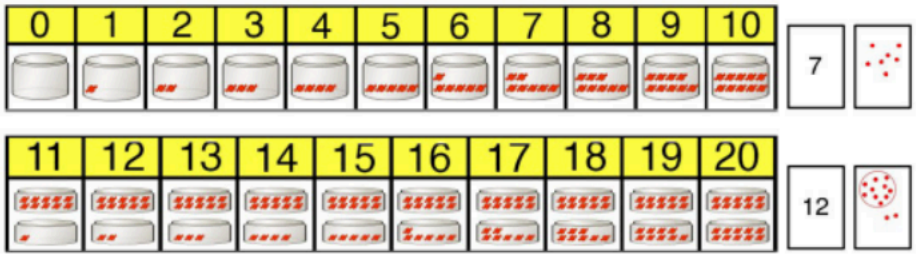
Grade 1 (0-3 months)

	(1) All Math		(2) Symbolic		(3) Intuitive	
Mixed Treatment	0.110***	t(140)=3.43	0.107***	t(140)=3.03	0.104**	t(140)=2.54
	(0.047,0.174)	[0.001]	(0.037,0.177)	[0.003]	(0.023,0.185)	[0.012]
Control Mean	0.369		0.294		0.557	
Observations	1,344		1,344		1,344	
Number of Clusters	141		141		141	

Table 4. *Regressions of kindergarten (top) and first-grade (bottom) children's performance on (1) the overall math composite, (2) the symbolic composite, and (3) the intuitive composite on an indicator for being in the mixed-games condition in Experiment 2. The treatment effects are estimated with (2) along with the relevant statistical tests. The indices were created by standardizing all test scores to the control condition at baseline and taking a simple average. For each outcome there are two columns. In the first column, the treatment effect is presented with stars indicating a p-value less than 0.1, 0.05, and 0.01. Below the point estimate in parentheses is the 95% confidence interval. The second column presents a t-statistic for a test of whether the coefficient is equal to zero. Below this statistic is the p-value of the test in brackets. All standard errors are clustered at the school level.*

Figures

a. Number Pretraining Posters



b. Geometry Pretraining Posters and Cards

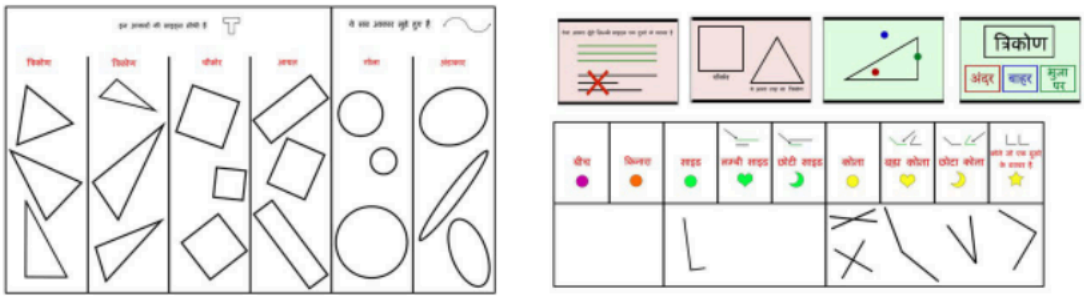


Figure 1. Pretraining materials for Experiment 1. Posters and cards served to connect intuitive math concepts to the language and symbols of mathematics. In (b), shape names (left) and geometric properties (right) are written in Hindi for the facilitators, who use these words while indicating the corresponding shapes (e.g., "triangle," "oval") and properties on the poster (e.g., "crossing lines," "small [i.e., acute] angle"). The front of the pink card asks children to identify the shapes on the poster that have equal-length sides, while the back provides the answers. The two sides of the green card convey different locations on a shape either pictorially (left) or in words (top: "triangle"; bottom-left: "inside"; bottom-middle: "outside" ; bottom-right: "on the side"). The poster shown below the sample cards conveys hints for: "center"; "side"; "long side"; "short side"; "corner"; "big corner"; "small corner"; and "special [i.e., right-angled] corner."

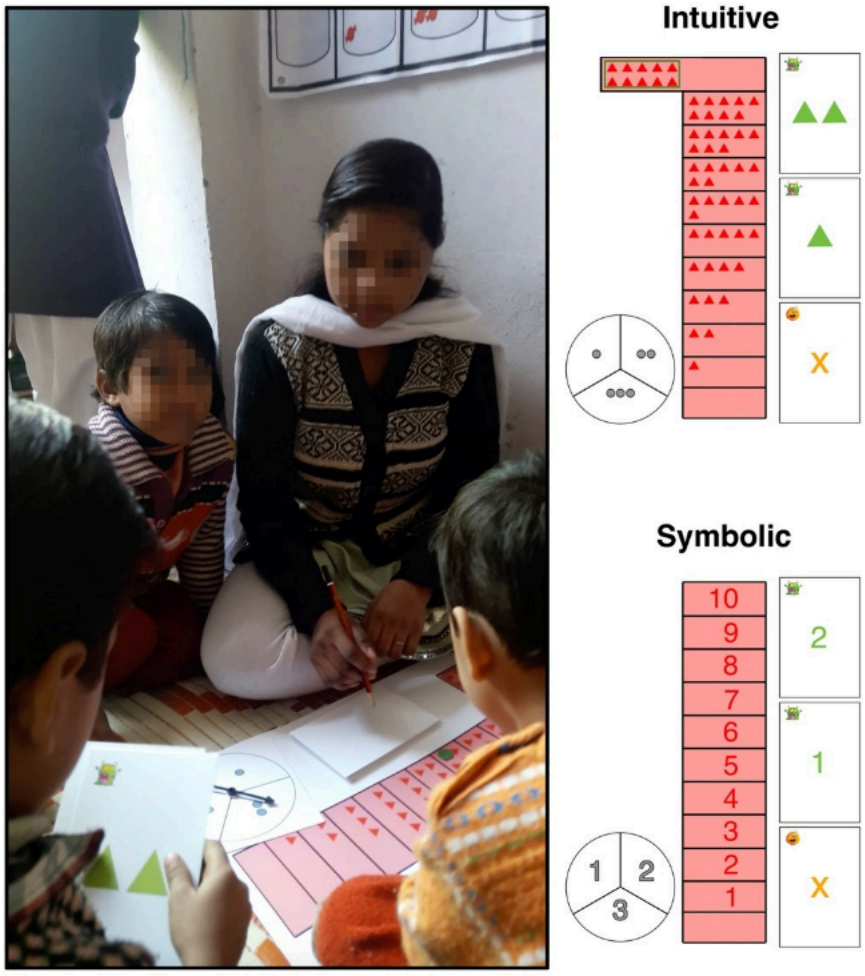


Figure 2. A math game from Experiment 1 focusing on one-to-one correspondence operations with small numbers, with materials eliciting intuitive and symbolic concepts. Preschool children are shown playing this game in a small group with the facilitator seated in front of a number training poster.

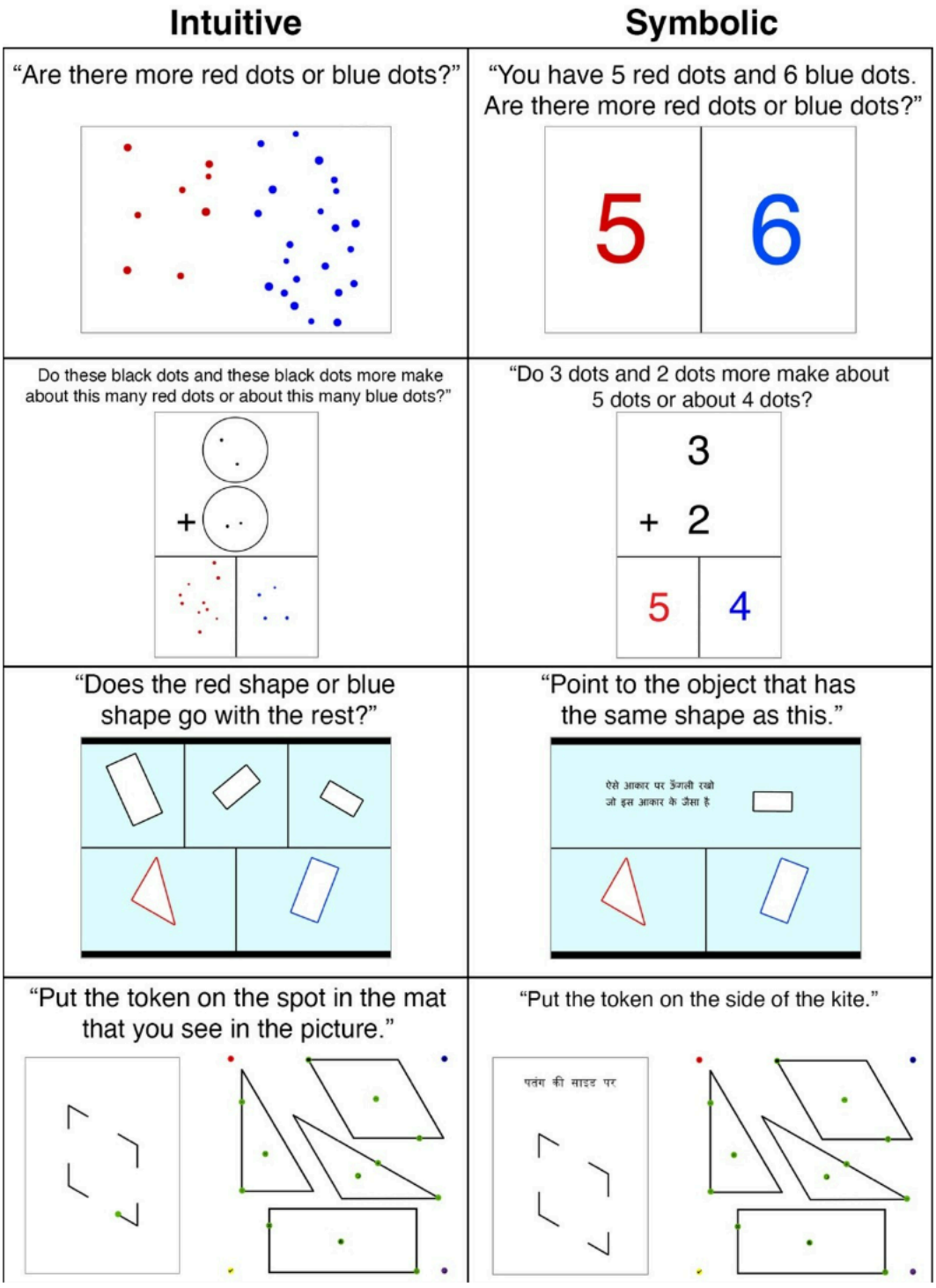


Figure 3. The three remaining math games used in Experiment 1. Number Comparison is designed to elicit intuitive concepts of large, approximate numbers (top-two rows); Find the Shape (third row) and Reading Maps (fourth row) are designed to enhance sensitivity to the properties of 2D geometric forms and 3D navigable layouts, respectively. The sample intuitive (left) and symbolic (right) materials are accompanied by instructions used by the facilitator, as indicated in the text accompanying each set of materials.

Figure 4. Pretraining and gameplay in Experiment 2. (a) Pretraining: a classroom teacher provides instruction with the pretraining materials, and a child consults the number pretraining poster. (b) Number game: example gameplay of an adapted version of Number Comparison in kindergarten and first-grade classrooms, shown from the perspectives of a child, a group of peers, and the full classroom. (c) Geometry game: example gameplay of an adapted version of Find the Shape across the same three perspectives. First-graders are seated at desks; kindergarten children are seated on the floor.

References

1. ASER Centre. *Annual Status of Education Report (Rural) 2024* (Pratham, 2025).
2. Banerjee, A. V. *et al.* From proof of concept to scalable policies: challenges and solutions, with an application. *J. Econ. Perspect.* **31**, 73–102 (2017).
3. Bendini, M. & Devercelli, A. E. (eds) *Quality Early Learning: Nurturing Children's Potential* (World Bank, 2022).
4. Izard, V., Sann, C., Spelke, E. S. & Streri, A. Newborn infants perceive abstract numbers. *Proc. Natl Acad. Sci. USA* **106**, 10382–10385 (2009).
5. Hyde, D. C. & Spelke, E. S. Neural signatures of number processing in human infants: evidence for two core systems underlying numerical cognition. *Dev. Sci.* **14**, 360–371 (2011).
6. Gennari, G., Dehaene, S., Valera, C. & Dehaene-Lambertz, G. Spontaneous supra-modal encoding of number in the infant brain. *Curr. Biol.* **33**, 1906–1915.e6 (2023).
7. Ayzenberg, V. & Lourenco, S. F, Perception of an object's global shape is best described by a model of skeletal structure in human infants. *eLife* **11**, e74943 (2022).
8. Dillon, M. R., Izard, V. & Spelke, E. S. Infants' sensitivity to shape changes in 2D visual forms. *Infancy* **25**, 618–639 (2020).
9. Smith, L. B. Learning to recognize objects. *Psychol. Sci.* **14**, 244–250 (2003).
10. Landau, B., Smith, L. B. & Jones, S. S. The importance of shape in early lexical learning. *Cogn. Dev.* **3**, 299–321 (1988).
11. Hermer, L. & Spelke, E. S. A geometric process for spatial reorientation in young children. *Nature* **370**, 57–59 (1994).

12. Newcombe, N. S. & Huttenlocher, J. *Making Space: The Development of Spatial Representation and Reasoning* (MIT Press, 2000).
13. Dillon, M. R., Huang, Y. & Spelke, E. S. Core foundations of abstract geometry. *Proc. Natl Acad. Sci. USA* **110**, 14191–14195 (2013).
14. Carey, S. *The Origin of Concepts* (Oxford Univ. Press, 2009).
15. Gilmore, C. K., McCarthy, S. E. & Spelke, E. S. Symbolic arithmetic knowledge without instruction. *Nature* **447**, 589–591 (2007).
16. Kibbe, M. M. & Feigenson, L. Young children 'solve for x' using the Approximate Number System. *Dev. Sci.* **18**, 38–49 (2015).
17. Pica, P., Lemer, C., Izard, V. & Dehaene, S. Exact and approximate arithmetic in an Amazonian indigene group. *Science* **306**, 499–503 (2004).
18. Dehaene, S., Izard, V., Pica, P. & Spelke, E. S. Core knowledge of geometry in an Amazonian indigene group. *Science* **311**, 381–384 (2006).
19. Piazza, M., Pica, P., Izard, V., Spelke, E. S. & Dehaene, S. Education enhances the acuity of the nonverbal approximate number system. *Psychol. Sci.* **24**, 1037–1043 (2013).
20. Butterworth, B., Reeve, R., Reynolds, F. & Lloyd, D. Numerical thought with and without words: evidence from indigenous Australian children. *Proc. Natl Acad. Sci. USA* **105**, 13179–13184 (2008).
21. Biederman, I., Yue, X. & Davidoff, J. Representation of shape in individuals from a culture with minimal exposure to regular, simple artifacts: sensitivity to nonaccidental versus metric properties. *Psychol. Sci.* **20**, 1437–1442 (2009).
22. Sablé-Meyer, M., Ellis, K., Tenenbaum, J. & Dehaene, S. A language of thought for the mental representation of geometric shapes. *Cognit. Psychol.* **139**, 101527 (2022).

23. Wynn, K. Children's understanding of counting. *Cognition* **36**, 155–193 (1990).
24. Wynn, K. Children's acquisition of the number words and the counting system. *Cognit. Psychol.* **24**, 220–251 (1992).
25. Mix, K. S., Smith, L. B., Stockton, J. D., Cheng, Y.-L. & Barterian, J. A. Grounding the symbols for place value: evidence from training and long-term exposure to base-10 models. *J. Cogn. Dev.* **18**, 129–151 (2017).
26. Fuson, K. C. *Children's Counting and Concepts of Number* (Springer, 1988).
27. Siegler, R. S. & Opfer, J. E. The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* **14**, 237–243 (2003).
28. Halberda, J., Mazzocco, M. M. M. & Feigenson, L. Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* **455**, 665–668 (2008).
29. Piazza, M., Pinel, P., Le Bihan, D. & Dehaene, S. A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron* **53**, 293–305 (2007).
30. Dillon, M. R., Kannan, H., Dean, J. T., Spelke, E. S. & Duflo, E. Cognitive science in the field: a preschool intervention durably enhances intuitive but not formal mathematics. *Science* **357**, 47–55 (2017).
31. *National Education Policy 2020* (Ministry of Human Resource Development, Government of India, 2020).
32. Banerjee, A. V. *et al.* Children's arithmetic skills do not transfer between applied and academic mathematics. *Nature* **639**, 673–681 (2025).
33. Yang, C. Ontogeny and phylogeny of language. *Proc. Natl Acad. Sci. USA* **110**, 6324–6327 (2013).

34. DeLoache, J. S. Becoming symbol-minded. *Trends Cogn. Sci.* **8**, 66–70 (2004).
35. Dillon, M. R. & Spelke, E. S. Core geometry in perspective. *Dev. Sci.* **18**, 894–908 (2015).
36. Lee, S. A., Shusterman, A. & Spelke, E. S. Reorientation and landmark-guided search by young children: evidence for two systems. *Psychol. Sci.* **17**, 577–582 (2006).
37. Gweon, H. Inferential social learning: cognitive foundations of human social learning and teaching. *Trends Cogn. Sci.* **25**, 896–910 (2021).
38. Tomasello, M. *Origins of Human Communication* (MIT Press, 2008).
39. Negen, J. & Sarnecka, B. W. Number-concept acquisition and general vocabulary development. *Child Dev.* **83**, 2019–2027 (2012).
40. Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M. & Hedges, L. V. Preschool children's mathematical knowledge: the effect of teacher 'math talk.' *Dev. Psychol.* **42**, 59–69 (2006).
41. Gibson, D. J., Gunderson, E. A. & Levine, S. C. Causal effects of parent number talk on preschoolers' number knowledge. *Child Dev.* **91**, e1162–e1177 (2020).
42. Pruden, S. M., Levine, S. C. & Huttenlocher, J. Children's spatial thinking: does talk about the spatial world matter? *Dev. Sci.* **14**, 1417–1430 (2011).
43. Duflo, E., Dupas, P., Spelke, E. S. & Walsh, M. P. *Intergenerational Impacts of Secondary Education: Experimental Evidence from Ghana* NBER Working Paper No. 32742 (National Bureau of Economic Research, 2024).
44. Dehaene, S. *The Number Sense: How the Mind Creates Mathematics* (Oxford Univ. Press, 1997).

45. Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q. & Germine, L. Number sense across the lifespan as revealed by a massive Internet-based sample. *Proc. Natl Acad. Sci. USA* **109**, 11116–11120 (2012).
46. Spelke, E. S. Core knowledge, language, and number. *Lang. Learn. Dev.* **13**, 147–170 (2017).
47. Piazza, M. Neurocognitive start-up tools for symbolic number representations. *Trends Cogn. Sci.* **14**, 542–551 (2010).
48. Jara-Ettinger, J., Piantadosi, S., Spelke, E. S., Levy, R. & Gibson, E. Mastery of the logic of natural numbers is not the result of mastery of counting: evidence from late counters. *Dev. Sci.* **20**, 1–11 (2017).
49. Evans, D. K. & Yuan, F. How big are effect sizes in international education studies? *Educ. Eval. Policy Anal.* **44**, 532–540 (2022).
50. Angrist, N. & Meager, R. Implementation Matters: Generalizing Treatment Effects in Education. Preprint at *SSRN* https://dx.doi.org/10.2139/ssrn.4487496 (2023).
51. Calero, C. I., Zylberberg, A., Ais, J., Semelman, M. & Sigman, M. Young children are natural pedagogues. *Cogn. Dev.* **35**, 65–78 (2015).
52. Gibson, E. J. The ontogeny of reading. *Am. Psychol.* **25**, 136–143 (1970).
53. Treiman, R. Teaching and learning spelling. *Child Dev. Perspect.* **12**, 235–239 (2018).
54. Gibson, E. J. & Levin, H. *The Psychology of Reading* (MIT Press, 1975).
55. Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. Human-level concept learning through probabilistic program induction. *Science* **350**, 1332–1338 (2015).
56. Dehaene, S. *Reading in the Brain: The New Science of How We Read* (Penguin Books, 2010).

57. Hyde, D. C., Khanum, S. & Spelke, E. S. Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition* **131**, 92–107 (2014).