Enhancing early math skills through digitally adapted social learning in the classroom

Nadir Díaz-Simón[a,b], Guillermo Trinidad[c], Dinorah de León [a], Elizabeth Spelke [d], Alejandro Maiche[a,e]

[a]Centro Interdisciplinario en Cognición para la Enseñanza y el Aprendizaje, Universidad de la República, Montevideo, Uruguay
[b]Numerical Cognition Laboratory, Department of Psychology and Brain and Mind Institute, Western University, London, Ontario, Canada
[c] Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay
[d] Department of Psychology, Harvard University, Boston, United States
[e]Instituto de Fundamentos y Métodos en Psicología, Facultad de Psicología, Universidad de la República. Montevideo, Uruguay

Abstract:

Many children worldwide fail to realize their potential for learning school mathematics. Diverse initiatives have been aimed at changing this situation, by using digital technologies to expand training possibilities and creating and disseminating new educational materials adapted to children's abilities. Most of these efforts focus on training that is adapted to individual children, however, and they draw the child's attention away from the teacher and their peers. Here we introduce a novel approach to digital learning, applicable to groups of children who learn together by playing with concrete materials in small social groups, and who receive feedback only at the group level, encouraging discussions to arrive at consensus responses to math problems. The social groups (typically composed of 4 students) work within the classroom under an adult's direct view. In a small-scale randomized experiment, we tested the effectiveness of such a program by comparing the math skills of children who played a set of math games in school, during part of the time reserved for math instruction, either in small groups or individually. When compared to a no-treatment control condition in which no games were played, no differences were found in children's mathematical gains, showing that the game play compensated for the shorter time of direct instruction that the children who played the math games had received. More importantly, the games played in small social groups with peer-focused interactive learning led to greater advances in children's math skills than the same games played individually on tablets. Gains were especially pronounced for the children whose math skills were least developed, contrary to the concern that cooperative group play will enhance learning disparities because the most advanced students are likely to guide the group activities. Our results show that digitally controlled peer interactions enhance learning of pre-school and primary school mathematics for children at all levels and especially for those who started the intervention with the least mathematical knowledge. Digitally controlled games, played by children in small groups, therefore, promise to enhance children's mastery of the mathematical skills taught in primary school above and beyond the effects of the regular math curriculum and of digitally controlled games targeted to individual children.

Keywords: mathematical cognition, early math abilities, cognitive training, peer interaction

Introduction

In modern societies, mathematical competence is important for the participation of citizens in daily decision-making, and it is a strong predictor of success in diverse fields (Claessens & Engel, 2013). Evidence from meta-analytic and longitudinal studies shows that early math skills are the strongest predictor of later school achievement (Duncan et al., 2007; Habermann et al., 2020). Furthermore, numerical literacy is associated with positive adult life outcomes and socioeconomic status (Ritchie & Bates, 2013).

It is widely believed that children's intuitive knowledge of mathematics positively impacts on their learning of math in school, and that children's intuitive knowledge and their motivation for developing math skills are enhanced by games that encourage playful math activities in families or groups of peers (Fisher, Hirsh-Pasek & Golinkoff, 2012; Ramani and Siegler; 2008). Laboratory studies provide evidence that games that exercise children's intuitive capacities for non-symbolic, approximate numerical comparison and addition produce a short-term boost in children's performance of symbolic arithmetic (Hyde et al., 2013; Khanum et al., 2016; Park and Brannon, 2013). The intuitions on which these games operate do not seem to depend on the sociocultural context, because the games are equally engaging and understandable to middle-class and disadvantaged children (Gilmore et al, 2007). In addition, infants and preschool children have a natural interest in and understanding of number and geometry (Dehaene, 2011; Jara-Ettinger et al., 2016; Newcombe & Huttenlocher, 2006), as well as a natural motivation to interact with friends and families in groups (Brownell et al., 2009). School interventions to enhance children's learning of mathematics may be most effective, therefore, when they build on these abilities and proclivities, in settings that involve peer interaction.

Although some digital games have been shown to improve children's school learning (Cheung & Slavin, 2013), most successful educational interventions have small effects, especially in the domain of mathematics: a recent meta-analysis of 252 published and successful international educational interventions found that the median effect size was just .07 for studies of mathematics (Evans & Yuan, 2022). Also, it often is not clear how to incorporate a successful intervention into a class setting (Clark et al., 2016; McTigue et al., 2019). Time consumed by games likely would replace some of the time taken by the school curriculum and therefore might have a negative impact on children's learning. Recent studies reveal that in eight weeks of school closure during the COVID-19 pandemic, children lost the equivalent of 20% of what would be achieved during a typical school year (Engzell et al., 2021). Moreover, the losses were larger for low-achieving students and for students in schools with less socio-cultural capital. Thus, games that take time away from the regular school math curriculum could have a negative impact on children's overall math learning, especially for the children who have the most to learn.

In the last decades, cognitive scientists have developed cognitive interventions specifically addressed to improve learning. Some of the most appealing interventions, based on findings from developmental cognitive science, have shown both overall improvements in school math learning and also reductions in the achievement gap between more and less advantaged children (Valle Lisboa et al., 2017). These results suggest that introducing games that train cognitive skills during some of the time allotted to the teaching of school instruction can be beneficial.

These results, together with the development of digital technologies that provide continuous and individualized feedback to children, has led to the introduction of a myriad of digital platforms and games inside and outside of schools (Bulman & Fairlie, 2016; Linden, 2008). However, many of the tools that have been developed for use in interventions do not show benefits (Reynvoet et al., 2021; Szűcs & Myers, 2017), and almost all have drawbacks when used in schools, because they focus the child's attention away from the teachers who are the most important source of children's instruction in mathematics, and from peers who face the same learning challenges in the math domain. Because the experiences children encounter in playing math games are very different from those involved in the teacher-led math curriculum, moreover, the skills that children develop in playing games may fail to transfer directly to instruction. Consistent with this possibility, experiences out of school that strengthen children's intuitive math abilities, such as selling in markets, fail to translate into any benefit for mathematics instruction in school, even for children who are full-time students while also working in markets (Banerjee et al., 2022).

Recent experiments have found that games using concrete materials, played by groups of children who communicate and cooperate or compete with one another, also can enhance children's math learning prior to formal instruction (Dillon et al., 2017). Classic research by Vygotsky (Vasileba & Balyasnikova, 2019), as well as contemporary research in social neuroscience (Clark & Dumas, 2015), reveal the important role of peer interaction and joint attention in children's learning. Consistent with those findings, children's learning is facilitated by joint attention and reciprocal interactions with social partners, as children showed greater improvement in groups that included social interaction activities in their sessions (Verga & Kotz, 2017; Conboy et al., 2015). Thus, cognitive interventions may benefit from paying attention to the social component of learning. Nevertheless, little research has addressed the role of social interaction in math learning, especially at the onset of formal schooling (Gersten, 2009). Learning in social contexts has numerous potential drawbacks, including the possibility that the most knowledgeable students will contribute the most to group activities and gain the most by doing so.

The present study aims to assess the costs and benefits of a games-based intervention that aims to enhance children's mathematical skills near the onset of formal schooling. The game used in both training conditions, inspired by the games developed by Dean et al. (in prep), is a card sorting game with five different categories of tasks, each focused on a specific math skill: approximate, non-symbolic comparison using arrays of dots, approximate numerical comparison using numerical symbols, geometry, approximate arithmetic using a combination of dot arrays and numerical symbols, and symbolic arithmetic (week 5). Each card of the game presented a math challenge with a two-alternative forced choice format, such that children had to select one of two possible answers (shown in blue or red ink at the bottom of each card) in response to the challenge (see Figure 1A).

By using this game, we attempt to synergize the benefits of digitally presented feedback with experiences of the traditional dynamics in the classroom, such as face-to-face contact and peer interaction, and to compare the effects of group learning both to a similar intervention presented to children individually and to the regular math curriculum. We aim for an intervention that will be beneficial for school children at all levels of ability, but we focus especially on students with lower math abilities. Early interventions are especially important for such students, in order to prevent them from falling further behind as more advanced math concepts are introduced in later grades (Nelson & McMaster, 2019). Nevertheless, there is a real possibility that such students will suffer both from the reduction of time devoted to the formal math curriculum and from a tendency to defer to more able students during group-based play. To the best of our knowledge, no research has compared individual games with peer-interaction concrete games for learning early math. Testing these approaches side by side, using the same game with identical content, is important in order to find the best way to train the early math skills of the least able children and to evaluate and implement interventions at scale.

In the present study, a new math game, playable by small groups of children, is introduced and evaluated in an exploratory study. The game is played by groups of four children, with cards and a *magic box* that provides partially informative feedback on children's performance to the group of children themselves, and potentially to their teacher as well. We compare the gains of children who play this social math game both to children who play the same game individually, on a tablet in which the same cards appear and similar feedback is given to the individual child only. Because both games are played in school, during the time allotted to math instruction, we first ask whether playing either game (social or individual) lessens children's math learning by reducing the normal math curriculum. After discovering that it does not, we then ask whether the game is more effective when it is played individually on tablets or during peer interaction.

The final innovative feature of this study concerns the feedback that each tablet presents to the individual player and that the magic boxes present to the groups of players. Most games-based interventions provide full feedback on children's performance, delivered either automatically by the tablet on which the game is played (Salminen et al., 2015) or by the teacher who supervises the game play. Other games-based interventions provide no feedback to children (Odic et al., 2014). Here we aim to provide children with partial feedback on their performance: feedback that challenges children to think harder about the problems where something has gone wrong. In the group play condition, each child plays a single card and after all four children's cards are played, the magic box emits one sound if all four cards were played correctly and a different sound if one or more cards were played incorrectly, without indicating how many errors had been made or which cards were played in error. On getting the error signal, the children in the group compare their answers to arrive at a new determination of which answers are correct, and then they play the same four cards again, getting a second round of feedback. The game does not advance to new cards until the magic box indicates that all four cards have been played correctly. In the individual play condition, each child plays two cards and after the second is played, the tablet produces the same two sounds: one indicating that both cards were played correctly and the other indicating that one or both cards were played incorrectly. In the latter case, children must reflect on their judgments and play both cards again; they cannot progress to new cards until they play both cards correctly. In both conditions, therefore, feedback aims to promote children's reflection on their performance. For the social games, feedback also encourages children to discuss their judgments with other children in their group.

To compare the two modes of play, we randomized half the children within each classroom that received games training to play each kind of game (individual vs. social), and we administered pretests and post-tests of their mathematical abilities. To address the possibility that the games would be detrimental to children in both modalities because they drew time away from the regular school math curriculum, we also compared the gains shown by children in the two games conditions combined (hereafter, the treatment condition) to those shown by children who received no games and therefore experienced a longer period of teacher-led math instruction (the no-treatment, or control, condition).

Methods

Participants

Two hundred twelve children attending Senior Kindergarten (KG; n = 128) and First Grade (G1; n = 84) from five different schools located in the urban area of Montevideo, Uruguay participated in this study. The mean age of the sample was 77 months (6 years and 5 months, S*D* = 6.38 months, *range* = 65 months - 91 months).

Children assigned to the no-treatment condition attended two high SES schools (n=100; 49% girls). Sixty control participants were in KG and 40 were in G1. They received only the traditional educational instruction and were unaware of the existence of the intervention or the games that children from the other three schools were playing. The mean age of the control participants was 77.9 months (6 years and 6 months, *SD* = 5.89 months, *range* = 67 months - 89 months).

Children assigned to a game condition attended three different schools (*n*=112; 58% girls). Sixty-eight experimental participants were in KG and forty-four were in G1. Most of the children who participated in the intervention came from middle- to high-income families with the exception of 14 children from one of the schools, who lived in low-SES families. The mean age of the experimental participants was 76.2 months (6 years and 4 months, *SD* = 7.32 months, *range* = 65 - 93 months). The children were randomly assigned to one of the two game conditions within each classroom and grade level: *peer interaction* or *individual* play. Randomization to the two game conditions occurred at the child-level rather than the classroom level, based on the number of participants and their performance on the pre-test math measure. Thus, the randomization procedure yielded 56 children in the *peer interaction* condition (*Mean age* = 76.13 months, *SD* = 6.96 , *n* = 33 girls) and 56 children in the *individual* condition (*Mean age* = 76.09 months, *SD* = 6.54, *n* = 32 girls) with no significant age differences between the groups ($F(2, 210) = 2.10$, $p = .13$, $\eta_{2p} = .02$). Within each classroom, roughly equal numbers of children played individually and in groups. Thus, the numbers of children playing at the individual condition and peer interaction condition were 33 and 35, respectively, for the KG children and 23 and 21, respectively, for the G1 children.

This research was approved by the Research Ethics Committee of the Department of Psychology of Universidad de la República. The Research Assistants (RA) tested all the children in their own classroom at their schools.

Materials

All children (*N*=212) were assessed twice in their early math skills using an online self-administered test screener (PUMa, Marconi et al, in prep) on Android tablets with headphones that guide them through the evaluation. The assessment took place in each child's own classroom and lasted between 30 to 45 minutes, with the entire class being evaluated simultaneously.

Children assigned to the two treatment conditions (total *n*=112) played the game with the same set of cards. Half of them played individually using a tablet application specifically designed for the study and presenting images of the cards. The other half played with physical cards that they placed on top of a box with RFID sensors (the *Magic box*). Each physical card contained an RFID tag (invisible to the children) that enabled the box to interpret the cards (the "magic"). Each Magic box was equipped with an Arduino-powered smart box with two radio-frequency identification (RFID) readers: one for each side of the box (see figure 2A). Children who played individually using the tablet saw the same set of cards, but in a digital format on the screen; they did not manipulate any physical objects other than the tablet itself.

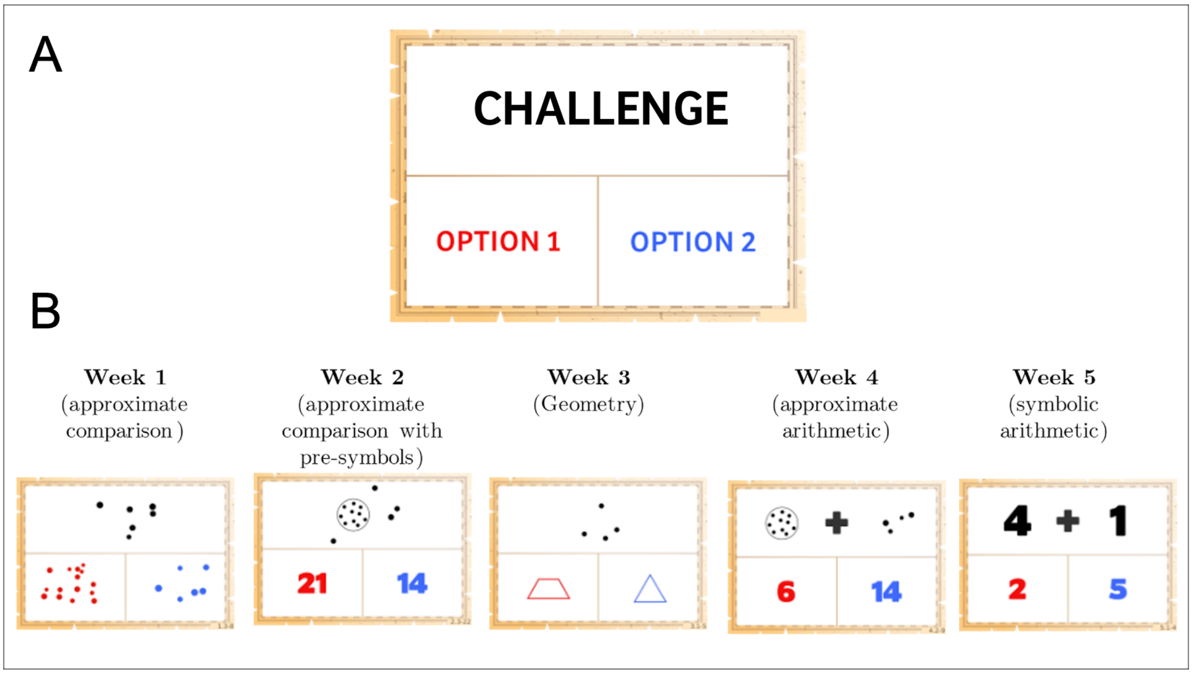
Procedure

The complete intervention lasted 5 weeks and consisted of three 30-minutes sessions per week. Therefore, children in the treatment conditions experienced 15 sessions of *peer-interaction* or *individual* games; one deck of cards was played per session. Each deck was composed of 36 cards, so the complete game offered 540 different challenges. In each deck, the correct response side was counterbalanced (red on half of the trials and blue on the other trials). The easiest trials were presented first, followed by progressively more difficult trials. Each week, children played with 3 decks devoted to each category of tasks (see figure 1B) and focused on a specific math skill: approximate, non-symbolic comparison (week 1), approximate comparison of numerical symbols (week 2), geometry (week 3), approximate arithmetic using both dot arrays and numerals (week 4) and symbolic arithmetic using numerals only (week 5).

On each day of the intervention, RAs prepared the rooms to be used for group play before the arrival of the children who were assigned to that condition. Children in the *peer-interaction* condition played in a room with four or five magic boxes (depending on the number of groups of four children), each with a deck of cards beside it. Once children arrived at the classroom, they sat around each magic box in groups of four and played with their peers (see figure 2A). Children received the game instructions orally by the RAs and the groups of four peers remained the same throughout all sessions of play. Children in the *individual condition* played in a room furnished with one tablet connected to one pair of headphones for each child. They played individually, each with a different tablet, usually sitting at shared tables for four children each. RAs prepared the 10-inch tablets from Plan Ceibal (Uruguayan National One Laptop per Child Program) with headphones for each child. (see figure 2B). When the children arrived at the classroom, they were instructed to put on the headphones and follow the audio instructions on how to play the game.

Figure 1

Examples of cards used in each category of the intervention. Panel A: the format of all cards. Panel B: examples of cards used for each math skill in each week of the training phase.



Both training conditions were embedded in a story: the children were told that their task was to help a pirate, “Brave Eye Patch”, to solve a set of challenges that would lead to the treasure. We used the same cover story and reward system to motivate children for both conditions. At the end of each day’s play, each child was given a sticker with a pirate theme with different designs.

Figure 2

Dynamics of the two training conditions. In the left panel (a), children are playing in the peer-interaction setting with a magic box that provides group feedback by audible beeps. In the right panel (b), children are playing individually with tablets that provide individual feedback through the player’s headphones .

(a) (b)

Experimental conditions

The procedure for both game conditions (individual and peer interaction) was similar except for the fact that the four children in the *peer-interaction* condition worked together to correct their problems, whereas each child in the *individual* condition corrected only their own problems. Because children in the *individual* condition played alone, we doubled the number of rounds and reduced the number of cards per round from four to two. Thus, the level of ambiguity in the feedback signal differed across the two conditions: the same error sound was given for rounds with 1, 2, 3 or 4 errors in the peer-interaction condition but was given for rounds with just 1 or 2 errors in the individual condition. By reducing the ambiguity in the individual condition, we aimed to equalize the burdens that the two conditions placed on each child, since the peer condition had greater uncertainty but more children available to resolve the uncertainty. For this reason, feedback frequency also differed: it occurred after 2 cards in the *individual* condition and after 4 cards in the *peer-interaction* condition. Both conditions, however, exposed children to the same number of cards presenting exactly the same task, and both conditions introduced uncertainty into the correction of errors, because children knew when errors had occurred but not how many errors occurred or which judgments were in error. They had to discover the error(s) by themselves, individually or as a group of four.

Individual

Children in the digital individual condition played the card sorting game on a tablet that displayed 2 cards on each turn: the child had to solve one card at a time indicating blue or red before the feedback signals either that both cards were played correctly or that one or two errors occurred (see Figure 2B). To continue to the next pair of cards in this tablet based game, both responses had to be correct. In this case, the tablet made a positive sound. If at least one of the answers were wrong, the tablet made a negative sound and both cards had to be answered again by the child. The sounds played by the tablets were exactly the same as those produced by the magic box in the peer interaction condition. The task was implemented in PsychoJS v3.1.5 software and presented from Pavlovia in the Google Chrome browser. A real version of the game can be played here.

Peer-Interaction play

Children in the peer interaction condition played the same games described above but in groups of four, generally sitting on the floor. Each group played the game with a Magic Box in the form that can be seen in Figure 2A. With these devices, games fostered peer interaction to find and correct errors in children's solutions to the challenges posed by the cards.

The group game started with each child solving one card by themselves and recording their answer by placing their card individually on the blue or red side of the Magic Box depending on their judgment of the side of the card (red or blue) that presented the correct answer. When all the answers of the 4 children in a group were correct, the Magic Box played the positive sound, indicating that children could advance to the next round and play the next four cards in the deck. When one or more answers was incorrect, the Magic Box played the negative sound, prompting the children to discuss their answers and collectively decide how to play each of the four cards again; play continued until the magic box indicated that all responses were correct, so that play with the next four cards could begin. Children were encouraged to help each other in order to obtain positive feedback from the Magic Box and advance to the next round of cards. This training condition therefore promoted peer interaction in order to advance rather than isolating each child with a different tablet.

Assessment

We used a between-subjects pre-test/post-test design. The pre-test and post-test assessments were each completed across one week before and after the game play.

For all children, the interval between the two assessments was 5-6 weeks. The pre-test and post-test assessments used the PUMa test (*Prueba Uruguaya de Matemática*): an online self-administered test screener composed of nine subtests for early math abilities. PUMa is a tool specifically designed to evaluate early math skills, particularly for kindergarten and first grade students. It was developed by our research team for quickly assessing early math skills of an entire classroom and identifying children who may be at risk for math learning difficulties. To maintain children's interest, different subtests are linked by an engaging story, presented through headphones, about two children who travel to different places in Uruguay where they find different problems to solve by using math. Currently, PUMa is being used in multiple public and private schools in Uruguay and Brazil, and has received high satisfaction ratings from teachers (for more information, see puma.cicea.uy).

PUMa consists of five subtests that assess symbolic math abilities and four subtests that assess intuitive math abilities. The five symbolic subtests assess the children’s abilities to represent and manipulate Arabic numerals through tasks of math fluency, verbal-numeral to Arabic transcoding, composition and decomposition of quantities and forward and backward ordering of numbers. The four non-symbolic assess approximate numerical comparison using arrays of dots, mental rotation of shapes, spatial pattern recognition task and detection of one-to-one correspondences. More details about all the subtests can be found in the supplementary materials.

Each of the nine subtests is composed of several trials in increasing order of difficulty, which allows the tool to be appropriate for both kindergarten and first grade students. In a validation study (N= 475; Marconi et al., in prep), PUMa showed good internal consistency (α = .86, ω = .87) and test-retest reliability (*ICC* = .90, r = .89, p < .01). Likewise, the PUMa showed adequate values of convergent validity with the third version of the Test of Early Mathematical Abilities (TEMA-3; r = .774, p < .01).

Data analysis

The data analysis was conducted using RStudio (RStudio Team, 2016). We transformed the pre-and post-test direct scores of PUMa into a standardized measure (Z-Scores) using the means and standard deviations of the pre-test assessment separately for first-grade and kindergarten children. This allowed us to perform analyses for all children together regardless of grade level.

To compare the pre-and post-test performance for the Control and the math games conditions[ES1] , paired Student t-tests were performed. Then, independent Student t-test analyses were performed to evaluate the effect of the treatment condition (Math games vs. Control) on pre-and post-test performance. We calculated the effect size using Cohen's d. Similar analyses were performed to compare the two math games conditions (peer interaction vs individual). Because the study involved randomization at the individual rather than the class level, consistent with our cross-class randomization procedures, we describe our effect sizes using the cut-offs that are recommended for small-scale experiments (Durlak, 2009), rather than those recommended for analyses of large-scale school-based interventions, which randomize students at the class level, cluster errors accordingly, and recommend more liberal guidelines for designating effect sizes as small, average, or large (Evans & Yuan, 2022).
For the children who received games (n=112), we performed these same analyses by dividing the total sample into three terciles based on the pre-test standardized scores of PUMa (see the section: Analysis by ability level).

Results

Analyses comparing math performance in the combined treatment conditions to that of the no-treatment control condition

In order to test whether the intervention interfered with the children's learning by reducing the time devoted to the normal math curriculum, we compared the progress in mathematical skills in the no-treatment control (*n*=100) and in the math games condition (*n*=112). Means and standard deviations of pretest and posttest scores of these two conditions are presented in Table 1.

Since the assignment to the control or experimental groups was not random (see participants section), there were pre-test differences, with a small effect size, between the two conditions ($t(210) = 2.47$, $p = .03$, $d = .31$). However, this difference is very similar to the post-test difference between the two conditions ($t(210) = 2.00$, $p < .05$, $d = .28$). There are no differences in the gain of math abilities after five weeks between the control and combined math games conditions ($t(210) = .09$, $p = .93$, $d = .01$), providing no evidence for learning losses caused by the reduction in math teaching that allowed for playing the math games.

Table 1

Descriptive statistics and group differences in the pre- and post-test assessment for children in the control and experimental conditions

Condition		Pre-test	Post-test	Gain	Group difference			
	n	*M (SD)*	*M (SD)*	*M (SD)*	*t*	*df*	*p*	*d*
Control	100	.16 (1.04)	.55 (1.11)	.39 (.64)	6.13	99	< .001	.36
Experimental	112	-0.14 (.94)	.23 (1.06)	.40 (.63)	6.69	111	< .001	.40

Note. Group differences show the comparison between pre- and post-test measures in math performance. *t:* Paired Student's t test. *df:* degree of freedom. *p: l*evel of significance. *d:* Cohen's d.

Based on the results of Table 1, we conclude that the intervention was not harmful to children, even though it reduced the time available for teacher-led math learning. Thus, we center our primary analyses on the main question of this study: What are the relative advantages and disadvantages of playing a math game in a group of four interacting peers, with concrete materials, compared to playing the same math game individually, on a tablet?

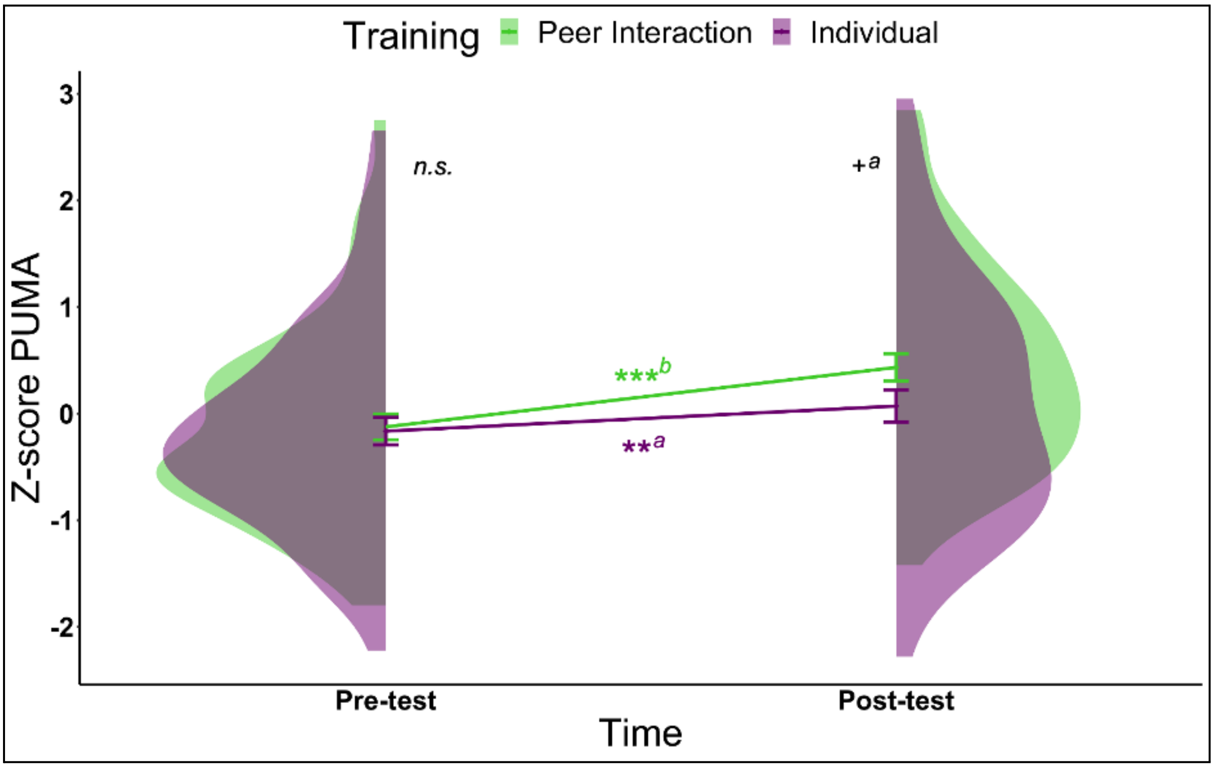
Analyses comparing math performance in the individual games condition to that of the peer-interaction games condition

To approach this question, we first compared the performance of the students in the two experimental conditions on the baseline math ability measure. There were no pre-intervention differences found in standardized PUMa scores between participants who received the peer-interaction training and those who underwent the intervention individually ($t(110) = .23$, $p<.82$, $d=.04$). We also found a high correlation between performance in the pretest and the post-test ($r = .81$, $p < .01$ for the training conditions; $r = .83$, $p < .01$ for the control condition), providing evidence for the temporal stability of our assessments.

To evaluate the effect of the intervention, we then compared the pretest and post-test scores on the PUMa for the two treatment conditions. Both experimental conditions show significant gains from pretest to post-test, with a medium effect size for children in the peer interaction condition ($t\ (55) = 7.14$, $p<.001$, $d=.59$) and a small effect size for participants in the individual condition ($t(55) = 2.79$, $p<.01$, $d=.22$). Comparison of post-test scores shows marginally significant differences, with small effect sizes, between the peer interaction and individual conditions ($t(110) = 1.83$, $p = .07$, $d = .35$; see Figure 3).

Figure 3

Mean pre- and post-training measures (± Standard Error), and density plot of math performance for Peer interaction and Individual Training conditions



Note. +: $p = .07$ **: $p < .01$. ***: $p < .001$. [a] : small size effect. [b]: medium size effect

Analyses by ability level

To analyze the effect of the intervention across different ability levels of mathematical performance, we performed a *post-hoc* analysis dividing the children who received math games into three groups, based on their performance on the PUMa at pretest: Beginner (lowest tercile), Intermediate (second tercile) and Proficient (highest tercile). Paired t-test analyses between pre- and post-training scores show significant improvements in math abilities for the beginners, with a large effect size for the peer interaction condition and a medium effect size for the individual condition (Table 2). For children in the intermediate and proficient terciles, only those who received the intervention in the peer interaction condition showed significant improvements, with a large effect size for the intermediate tercile and medium effect size for the most proficient tercile. Analyses comparing performance in the two training conditions to each other show a large advantage for the peer-interaction condition over the individual condition for the children at every tercile, but the effect sizes differed: large for the beginners, medium for the intermediates, and small for the proficient children (see Figure 4). It is possible, however, that these effects are inflated by noise in the test data, yielding regression to the mean.

Table 2

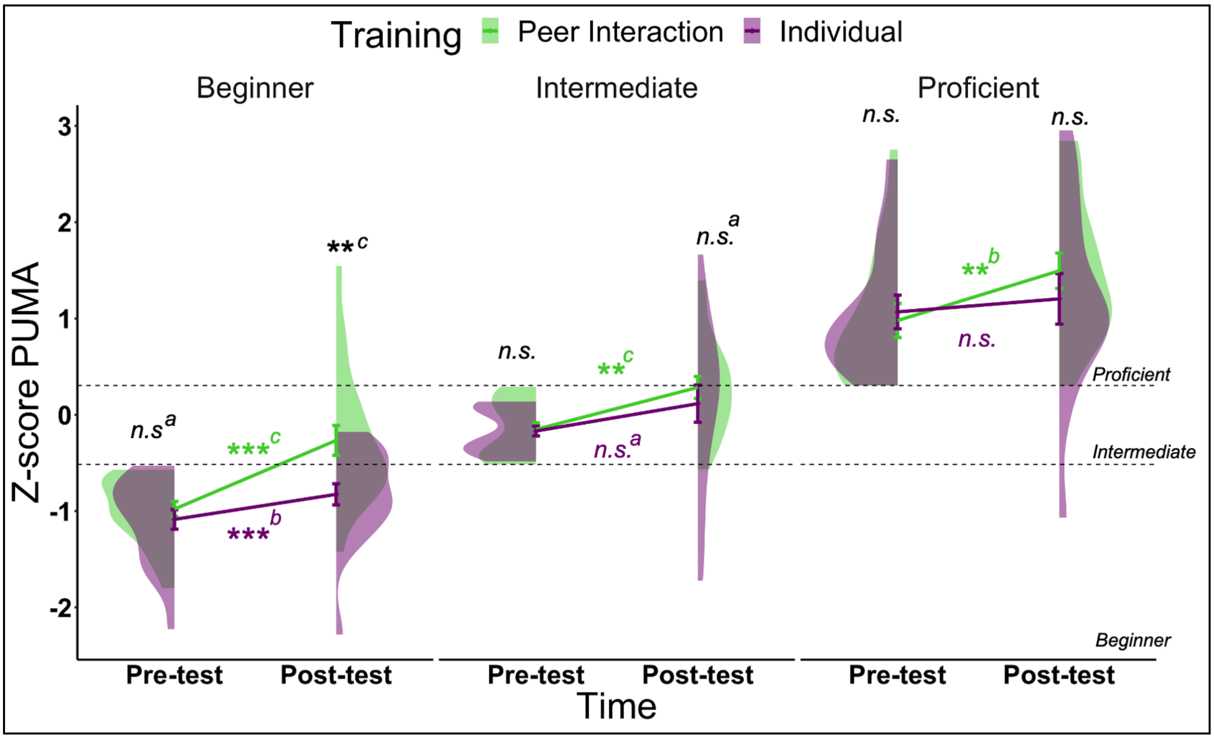
Descriptive statistics and group differences in the pre- and post-test assessment of math for both experimental conditions (Peer interaction vs Individual)

Profile	Training Condition		Pre-test	Post-test	Gain	Group difference			
		n	*M (SD)*	*M (SD)*	*M (SD)*	*t*	*df*	*p*	*d*
All	Peer Interaction	56	-.12 (.91)	.43 (.96)	.56 (.59)	7.14	55	< .001	.59
	Individual	56	-.16 (.97)	.07 (1.13)	.24 (.63)	2.79	55	.007	.22
Beginner	Peer Interaction	20	-.98 (.35)	-.27 (.69)	.71 (.64)	5.01	19	< .001	1.29
	Individual	20	-1.09 (.46)	-.83 (.49)	.26 (.39)	2.96	19	.008	.55
Intermediate	Peer Interaction	20	-.15 (.31)	.28 (.51)	.44 (.54)	3.62	19	.002	1.03
	Individual	21	-.17 (.23)	.12 (.89)	.28 (.84)	1.55	20	.136	.44
Proficient	Peer Interaction	16	0.98 (.71)	1.50 (.74)	.52 (.56)	3.66	15	.002	.72
	Individual	15	1.07 (.68)	1.20 (1.02)	.14 (.59)	.89	14	.387	.16

Note. Group differences show the comparison between pre- and post-test measures in math performance. *t:* Paired Student's t test. *df:* degree of freedom*. p: l*evel of significance*. d:* Cohen's d.

Figure 4

Mean pre- and post-training measures (± Standard Error), and density plot of math performance divided by cognitive profile

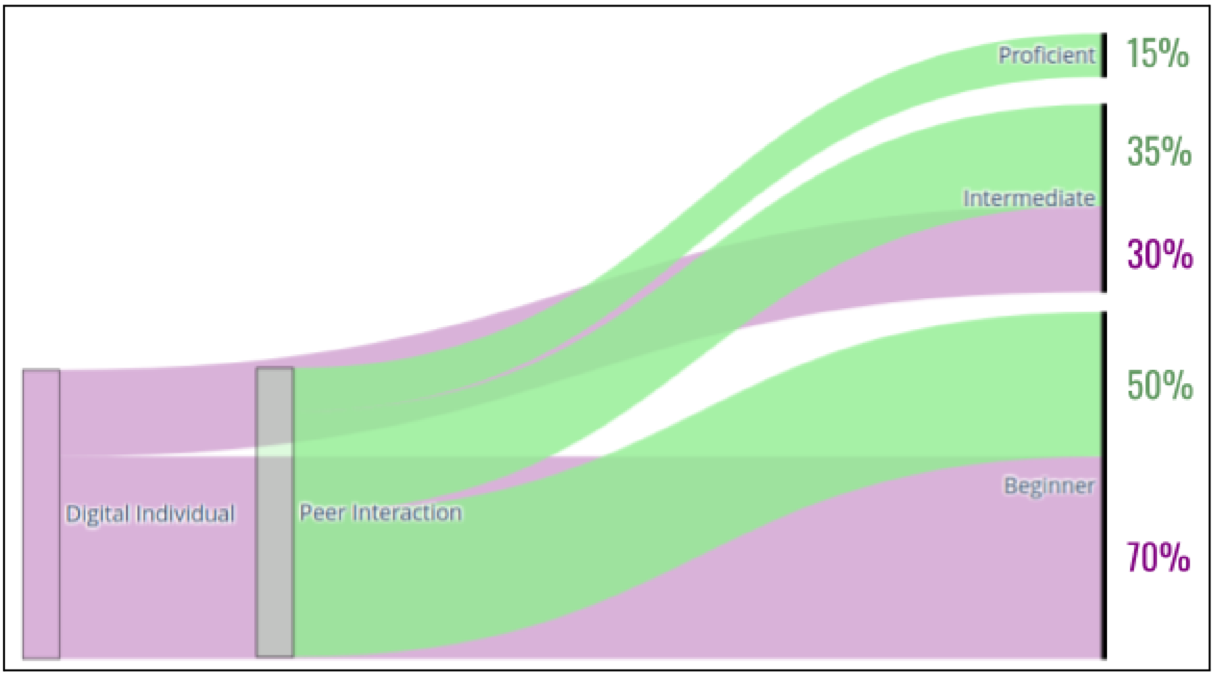


Note. **: $p < .01$. ***: $p < .001$. [a] : small size effect. [b]: medium size effect. [c]: large size effect. Dotted lines represent the cut-off points for the intermediate and proficient skill levels.

An analysis comparing the degree of improvement in children's performance across the two training conditions, performed on the sample size of 20 children for each training condition, is not subject to this regression effect. The analysis revealed a significant difference in the degree of improvement shown by children at the three ability levels, with a large effect size (.94). Fully 50% of the beginners who played the games with peer interaction moved to a higher tercile in the post-training assessment: either the intermediate level (35%) or the proficient level (15%). In contrast, only 30% of the beginners who played the games individually changed levels (see Figure 5). These findings run counter to the concern that peer interaction will disadvantage low-performing children, as higher performing children take the lead in group discussions and decision-making. Instead, group play of the present, intuitive and symbolic math games was most advantageous for the lowest performing children.

Figure 5

Cognitive mobility for beginner participants divided by training condition.



Discussion

In recent years, there has been growing interest among teachers and decision-makers in implementing institutional practices that aim to foster the development of mathematical skills by encouraging children to communicate with one another and solve problems in small group settings that motivate such interaction to solidify children's growing knowledge (*e.g.* Parrish, 2011). Group interaction carries with it a potential disadvantage, however: when a group of children must make collective decisions (here, how to correct the errors detected by the magic box), the most able children may participate more than other children do and thereby benefit more from the group activities, enhancing the disparities in children's skills of mathematical reasoning. There has also been wide interest in using digital technologies for individualized cognitive training in school contexts, because such training can be adapted to the performance level of each child, sparing high-performing children the boredom of classroom exercises they have already mastered, and sparing low-performing children the frustration of classroom exercises they are not ready to perform (Kiru et al., 2018). Individualized digital exercises also have disadvantages, however: they too may increase disparities across children by giving more proficient children more advanced material, and by diverting children's attention away from the teacher and classroom activities.

To the best of our knowledge, the group interaction condition that we introduce in this paper is the first intervention that aims to combine the features of peer-to-peer interaction with the features of having digital devices supporting the game, and the present experiment is the first to investigate the relative advantages of group and individual play, under conditions in which both types of play give adapted feedback to children.

Our results show that children who performed the intervention in group settings with peers (through magic boxes) improved their performance in early math skills more than those who performed the same training individually (through tablets). Moreover, contrary to concerns that group play and digital feedback each will favor the most advantaged children and enhance disparities in children's math learning within the classroom, we find that the benefits of group play are greatest for participants who started the intervention at the lowest level of performance in math. We conclude that peer-interaction is a key aspect of the learning process of math skills for those who usually stay behind in traditional formats of learning classes.

Why did the use of a group setting, mediated by educational technology (*i.e.,* the magic boxes), perform so much better than individual digital instruction, especially for the children who knew the least and needed help the most? Although the present experiment does not directly address this question, three features of the training condition with magic boxes may account for these effects. **First,** both training conditions used materials that are intuitively meaningful to all children. During teacher-led school math exercises, teachers necessarily must present challenges to the class that low-performing children fail to understand. The present games, however, used intuitive materials in both training conditions, pairing arrays of dots or geometric forms with mathematical symbols (Dean et al., in prep.). By using these materials, all the children in both training conditions were expected to find the math exercises meaningful. This experience would likely be most beneficial for the children who know the least math, because class exercises that are aimed at the middle or top of the class will not have been meaningful to them. **Second,** the magic box training condition allowed children with less knowledge of mathematics to learn from their more knowledgeable peers. Children who perform poorly in school may have less trust in their teachers, who reveal their errors publicly when they call in them and, by pitching material to the middle of the class, present them with challenges that they are not ready to meet. When a peer corrects or guides them, in contrast, they may pay greater attention and learn from this experience. **Third,** the use of partial feedback, rather than full feedback, in both conditions may have been especially advantageous for children in the magic box condition, because the feedback indicated when the group had made one or more errors but not which child or children were responsible for the errors. Thus, all the children could work constructively together to weed out the errors, through peer mathematical discussions. Because children are not informed of the exact number of errors they have made or which of the cards were played incorrectly, the peer interaction is a necessary step that, we believe, is a primary source of the benefits we found for children who participated in this condition. Furthermore, the use of partial feedback provided by the magic box on student responses enabled a wide range of behaviors that possibly impacted the development of mathematical skills. Some of the behaviors observed during the intervention aimed to build a shared understanding of mathematical ideas while encouraging children to build convincing arguments to express their mathematical ideas. This communication also provided children with opportunities to learn to see things from other children's perspectives.

Regardless of the roles played by the use intuitive materials, learning from peers, and building arguments to convince others, our findings show that peer interaction math games training can generate rapid improvements in the mathematical performance of children who begin with the least math knowledge, bringing them to intermediate levels or even, in some cases, to higher levels (Sorokin, 1959; figure 5). Our data show greater cognitive mobility for low-performing children in the peer-interaction condition than students in the individual condition. Furthermore, this technology facilitated teachers to simultaneously intervene with small groups engaged in discussing math challenges.

The present study has several limitations. First, with the current version of magic box, we are not able to track the evolution of students' responses during the training sessions, to assess their learning trajectories; further programming of the boxes should allow for such tracking in future studies. Second, the sample size was insufficient to achieve a high statistical power to the comparison of progress by the experimental groups relative to the control group, and the logistic limitations of the educational centers made it impossible to follow a Randomized Control Trial design for that comparison; children were randomized individually only across the two math games conditions, with half the children in each class assigned to individual and peer play. Although the primary function of the control group—to assess whether the interventions harmed children by reducing class time devoted to math—was met, future experiments are needed to evaluate the magnitude of the effects of the math games, relative to the no-treatment control. Third, the two training interventions were supervised by trained research assistants rather than by teachers. Further experiments, led by children's regular teachers and randomized at the class level rather than the child level, are needed to evaluate the scalability of digital peer intervention to enhance children's math learning in schools.

Despite these drawbacks, we believe that school-based interventions combining intuitive games, peer interaction, and partial digital feedback, focused on group rather than individual performance, have high promise for addressing the persistent gaps in children's primary school learning, especially in low-income populations and in low- and middle-income countries such as Uruguay. In principle, magic-box interventions could be used to foster children's learning in all domains, including reading, and at all class levels, including later grades of primary school and secondary education. We look forward to further developments and randomized evaluations of this technology.

References

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). The Naïve Utility Calculus: Computational Principles Underlying Commonsense Psychology. Trends in Cognitive Sciences, 13(7), 288-294. https://doi.org/10.1016/j.tics.2009.04.009

Banerjee, A., Fischer, G., Karlan, D., Lowe, M., & Roth, B. N. (2022). *Does the Invisible Hand Efficiently Guide Entry and Exit? Evidence from a Vegetable Market Experiment in India* (No. w30360). National Bureau of Economic Research.

Bulman, G. y R.W. Fairlie. 2016. "Technology and Education: Computers, Software and the Internet," en Eric Hanushek, Stephen Machin, and Ludger Woessmann, eds., Handbook of the Economics of Education, Elsevier, pp. 239-280.

Buschkuehl, M., & Jaeggi, S. M. (2010). Improving intelligence: a literature review. *Swiss Med Wkly*, *140*(19), 266–272.

Brownell, C. A., Svetlova, M., & Nichols, S. R. (2009). Toddlers' prosocial behavior: From instrumental to empathic to altruistic helping. Child Development, 80(4), 893-907. https://doi.org/10.1111/j.1467-8624.2009.01314.x

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, *9*, 88–113. https://doi.org/10.1016/j.edurev.2013.01.001

Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*, *115*(6), 1–29. https://doi.org/10.1177/016146811311500603

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. Review of Educational Research, 86(1), 79-122. https://doi.org/10.3102/0034654315582065

Clark I, Dumas G. Toward a neural basis for peer-interaction: what makes peer-learning tick? Front Psychol. 2015 Feb 10;6:28. doi: 10.3389/fpsyg.2015.00028.

Conboy BT, Brooks R, Meltzoff AN, Kuhl PK. (2015) Social Interaction in Infants' Learning of Second-Language Phonetics: An Exploration of Brain-Behavior Relations. Dev Neuropsychol. 2015; 40(4):216-29. doi: 10.1080/87565641.2015.1014487.

Day-Hess, C., & Clements, D. H. (2017). *The DREME Network: Research and Interventions in Early Childhood Mathematics*. *Advances in Child Development and Behavior* (1st ed., Vol. 53). Elsevier Inc. https://doi.org/10.1016/bs.acdb.2017.03.002

Dehaene, S. (2011). The number sense: How the mind creates mathematics. Oxford University Press.

Dean, J., Kannan, H., Dillon, M., Dufló, E. & Spelke, E. (in prep) From the lab to the field: Enhancing disadvantaged children's readiness for learning primary school mathematics.

Donnelly, R., & Patrinos, H. A. (2021). Learning loss during Covid-19: An early systematic review. *Prospects*, (0123456789). https://doi.org/10.1007/s11125-021-09582-6

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020). *COVID-19 and learning loss — disparities grow and students need help*.

Dowker, A. (2017). *Interventions for Primary School Children With Difficulties in Mathematics*. *Advances in Child Development and Behavior* (1st ed., Vol. 53). Elsevier Inc. https://doi.org/10.1016/bs.acdb.2017.04.004

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of pediatric psychology*, *34*(9), 917-928.

Evans, D. K. & Yuan, F. (2022). How big are effect sizes in international education studies? Educational Evaluation and Policy Analysis, Volume 44, Issue 3. https://doi.org/10.3102/01623737221079646

Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007 May;39(2):175-91. doi: 10.3758/bf03193146.

Foster, C., Burkhardt, H., & Schoenfeld, A. (2022). Crisis- ready educational design: The case of mathematics. *The Curriculum Journal*, *00*, 1–17. doi: 10.1002/curj.159

Fisher, Kelly & Hirsh-Pasek, Kathy & Golinkoff, Roberta. (2012). Fostering mathematical thinking through playful learning. Contemporary Debates on Child Development and Education

Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11267–11271. doi: 10.1073/pnas.1220568110

Gilmore, CK, McCarthy, SE, Spelke, ES (2007) Symbolic arithmetic knowledge without instruction, NATURE, 447(7144), pp.589-+, ISSN: 0028-0836. DOI: 10.1038/nature05850.

Habermann S, Donlan C, Göbel SM, Hulme C. The critical role of Arabic numeral knowledge as a longitudinal predictor of arithmetic development. J Exp Child Psychol. 2020 May;193:104794. doi: 10.1016/j.jecp.2019.104794

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. Cognition, 131(1), 92-107. https://doi.org/10.1016/j.cognition.2013.12.007

Khanum, S., Hanif, R., Spelke, E.S., Berteletti, I. & Hyde, D.C. (2016) Effects of Non-Symbolic Approximate Number Practice on Symbolic Numerical Abilities in Pakistani Children. *PLoS ONE* 11(10): e0164436. https://doi.org/10.1371/journal.pone.0164436

Kiru, E. W., Doabler, C. T., Sorrells, A. M., & Cooc, N. A. (2018). A synthesis of technology-mediated mathematics interventions for students with or at risk for mathematics learning disabilities. *Journal of Special Education Technology*, *33*(2), 111-123.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the Potential Impact of COVID-19 School Closures on Academic Achievement. *Educational Researcher*, *49*(8), 549–565. https://doi.org/10.3102/0013189X20965918

LeFevre, J. A., Skwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home Numeracy Experiences and Children’s Math Performance in the Early School Years. *Canadian Journal of Behavioural Science*, *41*(2), 55–66. doi: 10.1037/aOOl4532

Lehrl, S., Kluczniok, K., & Rossbach, H. G. (2016). Longer-term associations of preschool education: The predictive role of preschool quality for the development of mathematical skills through elementary school. *Early Childhood Research Quarterly*, *36*, 475–488. https://doi.org/10.1016/j.ecresq.2016.01.013

Linden, Leigh L. 2008. “Complement or Substitute? The Effect of Technology on Student Achievement in India.” Edited by Michael Trucano. InfoDev Working Paper 17, World Bank, Washington, DC

Lintern, G., & Boot, W. R. (2021). Cognitive Training: Transfer Beyond the Laboratory? *Human Factor*, *63*(3), 531-547. https://doi.org/10.1177/0018720819879814

Maldonado, E. J., & De Witte, K. (2022). The effect of school closures on standardised student test outcomes. *British Educatioonal Research Journalonal Research Journal*, *48*(1), 49–94. https://doi.org/10.1002/berj.3754

Marconi, C.,de León, D, Luzardo, M, Maiche, A. (in prep). Self-administered digital math assessment test for early childhood. Universidad de La República, Uruguay.

McTigue, E. M., Solheim, O. J., Zimmer, W. K., and Uppstad, P. H. (2019). Critically reviewing GraphoGame across the world: recommendations and cautions for research and implementation of computer-assisted instruction for word-reading acquisition. Read. Res. Q. 55, 45–73. https://doi: 10.1002/rrq.256

National Research Council. (2009). Mathematics learning in early childhood: Paths toward excellence and equity. Washington, DC: National Academies Press.

National Council of Teachers of Mathematics [NCTM]. (1991). Professional standards for teaching mathematics. https://jenmascheck.files.wordpress.com/2015/08/professionalstandards-for-teaching-mathematics.pdf

Nelson, G., & McMaster, K. L. (2019). The effects of early numeracy interventions for students in preschool and early elementary: A meta-analysis. *Journal of Educational Psychology*, *111*(6), 1001–1022. https://doi.org/10.1037/edu0000334

Newcombe, N. S., & Huttenlocher, J. (2006). Development of spatial cognition. Handbook of Child Psychology, 6th Edition, Volume 2: Cognition, Perception, and Language, 734-776. https://doi.org/10.1002/9780470147658.chpsy0315

Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology: General*, *143*(1), 255.

Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. Psychological science, 24(10), 2013-2019.

Price, G. R., & Wilkey, E. D. (2017). Cognitive mechanisms underlying the relation between nonsymbolic and symbolic magnitude processing and their relation to math. *Cognitive Development*, *44*(August 2016), 139–149. https://doi.org/10.1016/j.cogdev.2017.09.003

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How Numeracy Influences Risk Comprehension and Medical Decision Making. *Psychological Bulletin*, *135*(6), 943–973. https://doi.org/10.1037/a0017327

Reynvoet, B., Vanbecelaere, S., Depaepe, F., & Sasanguie, D. (2021). *Intervention studies in math: A metareview*. *Heterogeneous Contributions to Numerical Cognition*. Elsevier Inc. https://doi.org/10.1016/b978-0-12-817414-2.00012-9

Ritchie, S. J., & Bates, T. C. (2013). Enduring Links From Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status. *Psychological Science*, *24*(7), 1301–1308. https://doi.org/10.1177/0956797612466268

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games.

Salminen, J. B., Koponen, T. K., Leskinen, M., Poikkeus, A. M., & Aro, M. T. (2015). Individual variance in responsiveness to early computerized mathematics intervention. *Learning and Individual Differences*, *43*, 124-131.

Team, R. C. (2020). R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Vasileva O, Balyasnikova N. (Re)Introducing Vygotsky's Thought: From Historical Overview to Contemporary Psychology. Front Psychol. 2019 Aug 7;10:1515. doi: 10.3389/fpsyg.2019.01515. PMID: 31447717; PMCID: PMC6692430.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, March). What Works Clearinghouse: Procedures and standards handbook (Version 4.0). http://whatworks.ed.gov.

Supplementary Materials: **here.**