Chapter 2

The Effects of SES, Grade-Repeating, and IQ in a Game-Based Approximate Math Intervention

Josh Langfus*, Alejandro Maiche†, Dinorah De León†, Dahiana Fitipalde†, Álvaro Mailhos† and Justin Halberda*

**Johns Hopkins University, Baltimore, MD, United States*

†Centro de Investigación Básica en Psicología, Montevideo, Uruguay

INTRODUCTION

What tools can be used to close the gap between low- and high-achieving students? While there may be many ways to intervene—for example, teacher training, improving facilities, updating curricula—a great deal of attention has focused on the potential benefits of using technology as a teaching tool in the classroom. Nonprofit organizations like One Laptop Per Child (Trucano, 2011) and the World Computer Exchange (2016) have partnered with schools around the world to give students access to computers, tablets, and the internet. To the extent that digital literacy will be a valuable skill for tomorrow's workforce, providing early access to technology might be a useful investment. Less clear is the role that technology can play in bolstering instruction in traditional subject areas such as science, math, and reading. Some exploratory work has examined the effectiveness of technology-based classroom interventions in math education, with results suggesting benefits of training that focuses on specific skills, for example, manipulating decimals (Zhang, Trussell, Gallegos, & Asam, 2015) and multiplication and division (Pilli & Aksu, 2013). In contrast, other recent work has explored the effectiveness of nontechnology interventions in locations with restricted access to computers, though it is not clear that improvement on these intervention activities transfers to other math abilities (Dillon, Kannan, Dean, Spelke, & Duflo, 2017).

In the context of recent interest in technology's role in education, we present some initial findings from an ongoing classroom technology intervention in

Mathematical Cognition and Learning, Vol. 5. https://doi.org/10.1016/B978-0-12-815952-1.00002-5

elementary schools in and surrounding Montevideo, Uruguay—the country's capital and largest city. This effort is unique because Uruguay stands out as one of the first countries in the world to commit to providing public access to information and communications technology (ICT) on a nation-wide scale. Not only has the Uruguayan government invested in establishing free internet connections in schools and public places across the country, but also since 2008 it has partnered with the nonprofit organization One Laptop Per Child to provide every school-aged child with their own computer to be used in the classroom and at home. The joint initiative is called Plan Ceibal ("Conectividad Educativa de Informática Básica para el Aprendizaje en Línea" or "Educational Connectivity/Basic Computing for Online Learning"), and over the last decade, the program has expanded from its start in the nation's primary schools to include every student in Uruguay from preschool through elementary and middle school.

The dramatic increase in access to ICT has created a need for content on these platforms that educators can use in the classroom. Our team has partnered with a dedicated group of teachers at public elementary schools in the Montevideo area to create educational software for the tablets and evaluate its effectiveness. The present study is the latest from this ongoing project; a previous study engaged 503 first-graders in a number estimation task (Odic et al., 2016), and in the work presented here we introduced 386 s- and third-graders from a wide range of backgrounds to three tablet-based magnitude training games. We used a pretest, intervention, posttest design, along with a Business-As-Usual (BAU) Control group, to examine the effectiveness of playing these games on several measures, including a standardized test of school math ability (speeded arithmetic). The context of Uruguay also affords us the opportunity to look at whether such interventions helps some students more than others, e.g., as a function of Socioeconomic Status (SES) and Grade-Repeating status. As we will discuss in more detail later, our project includes students from schools across the SES spectrum, and therefore allows us to study the relationships between SES and measures of achievement and cognitive abilities, as well as how SES might moderate the impact of our intervention software.

In addition to SES, the large proportion of students in Uruguayan schools that repeat a grade allowed us to examine whether our intervention was particularly helpful for these students, and more generally to ask what kinds of students benefit most from our technology-based intervention. Our initial hope was that students who are most vulnerable or disadvantaged might stand to gain the most from our intervention games which focus on simple, intuitive, "core" magnitude discrimination. To explore how the current study may begin to answer these questions, it is important to understand the context of schools in Uruguay.

THE URUGUAYAN CONTEXT

One hope of the educational community in Uruguay is that providing equivalent technology to all children may help to bridge the gaps between schools in high- and low-SES communities. As mentioned previously, the schools in Uruguay provide a unique test case for exploring the potential of technological interventions, not only because of the willingness of teachers and students to engage with technology, but also because this technology is accessible to schools in communities across the socioeconomic spectrum. Table 1 summarizes the number of students per school per SES quintile included in the current study. The Uruguayan National Public Education Administration (ANEP) uses a variety of factors to classify each school's SES context: household education level, socioeconomic indicators (e.g., percentage of households with access to potable water), and social integration (e.g., head of household's employment status and percentage of household children attending school). Using an aggregate of these factors, ANEP classifies schools into quintiles (with level 1 as the lowest).

Uruguay's national government tracks a wide range of data about the nation's public education system, including enrollment and completion rates across many demographics, and these data highlight differences across socioeconomic contexts. A 2017 report from the National Institute of Education Assessment (INEEd) showed that, while access to education has improved over the last 10 years, striking disparities still exist across levels of SES in terms of enrollment, retention, and grade-repetition. These disparities can be seen from the beginning of enrollment in school through graduation. For example, between 2013 and 2015, preschoolers (age 3) in the least advantaged schools were less likely to meet minimum standards for required attendance (at least 141 days in the year) compared to students at the most advantaged schools; 28% of students in SES Quintile 5 did not meet this standard, compared to 46% in Quintile 1 (INEEd, 2017). Though overall attendance increases, and the gap narrows as students progress to higher grades, the data show a consistent 10% difference in meeting the standard for minimum attendance between the highest and lowest SES schools from first grade through sixth grade (INEEd, 2017). Socioeconomic differences are also apparent in the number of students who repeat a grade, which can occur for several reasons including lack of attendance or poor performance. Already in the earliest grades there are differences between higher and lower SES Quintiles with respect to children being on-target to progress through school at the appropriate age. For example, at 7 years of age only 4% of students in Quintile 5 schools had repeated or were repeating a grade, while the figure jumps to 12% for children in Quintile 1 schools, according to country-wide data (INEEd, 2017). And, most dramatically, by the time students graduated from secondary school, 83% of students in the least advantaged schools had repeated at least one grade, compared to 24% in the most advantaged programs (INEEd, 2017).

Arguably the most striking differences across SES manifest in terms of the number of students who leave education at either the primary-, middle-, or high-school level. While Uruguay has made significant strides in ensuring that nearly every student, regardless of SES, completes primary education, only

TABLE 1 Condition Assignment by School With SES Quintile Information (1 = Low, 5 = High)

School		A		B	C		D		E		F		G		H		
SES Quintile		*1*		*1*	*2*		*2*		*3*		*4*		*5*		*5*		
Classrooms		***a***	***b***	***a***	***a***	***b***	***a***	***b***	***a***	***b***	***a***	***b***	***a***	***b***	***a***	***b***	***c***
$n \approx$		17	18	20	27	31	20	22	26	27	24	27	29	28	26	19	27
Grade	2				X	X	X	X	X	X			X	X	X	X	
	3	X	X	X							X	X					X
Condition	BAU control	X		X	X		X		X		X		X		X		
	Game intervention		X			X		X		X		X		X		X	X

Note that for all schools except for B, every school had 2 year-matched classrooms—one for each condition (Intervention and Control).

50% of least advantaged students graduate from middle school, compared to 95% of students from Quintile 5 schools, and the trend continues through the end of high school, where the gap is 15% vs. 71% (INEEd, 2017).

In terms of the effect of SES on access to education, and foreshadowing some of our own result, data from our study show the same patterns as the INEEd report. On both the pre- and postintervention evaluation days of the current study, more students in the lower SES schools were absent compared to students at higher SES schools (see Table 2). Students at poorer schools are also more likely to have repeated a grade; within our sample of 386 children, the percentage of students whose birthdays fell after the legal cutoff date for their grade level was 5.4% at Quintile 5 schools compared to 41.8% at Quintile 1 schools (see Table 3)—a rate that is even more dramatic than what INEEd has reported for the national percentages in Quintile 1 (INEEd, 2017). While these disparities are disheartening, the fact that the schools who have partnered with us for this project seem to represent a microcosm

TABLE 2 Attendance Data by SES Quintile for the Testing Days

Quintile	1	2	3	4	5	Overall
N	55	100	53	49	129	386
% Present both	53	40	55	90	77	62
% Absent pre	18	26	15	8	12	16
% Absent post	22	26	25	2	9	16
% Absent both	7	8	6	0	3	5
Total %	100	100	100	100	100	100

Number of students enrolled in each Quintile is given at the top, and percent of students who were present for both, one, or neither test are given below. Students in lower SES classrooms (Quintiles 1, 2, and 3) were overall less likely to be in school testing days.

TABLE 3 Number of Students per SES Quintile Counted as Repeaters

Quintile	Repeaters	Total Students	Percent Repeating
1	23	55	41.8
2	27	100	27.0
3	9	53	17.0
4	4	49	8.2
5	7	129	5.4

of the issues affecting the education system as a whole in Uruguay allows this and future studies to identify interventions that stand a greater chance of scaling up to meet those broader challenges.

BACKGROUND FOR THE CURRENT STUDY

Socioeconomic Status in Education

Children from low SES households typically underperform their middle- and high-SES peers in school achievement (Sirin, 2005; Valle-Lisboa et al., 2016). These differences may be the result of a range of factors associated with SES (Wilkinson & Pickett, 2010). Performance differences have been observed in many specific domains, including language development (Hart & Risley, 1975; Hoff, 2003), IQ (Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003), cognitive ability (Larson, Russ, Nelson, Olson, & Halfon, 2015), spatial knowledge (Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005; Verdine, Irwin, Golinkoff, & Hirsh-Pasek, 2014), as well as mathematics (Klibanoff, Levine, Huttenlocher, Vasilyeva, & Hedges, 2006). Based on this literature, we expect to find differences in the initial cognitive and curricular abilities of children as a function of SES. Given these differences in baseline abilities, we might expect that SES differences will differentially affect the impact of our training games and test-retest improvement.

Classroom Geometric and Arithmetic Abilities

As with other content areas, math achievement varies as a function of SES in the United States (Sirin, 2005), Uruguay (Valle-Lisboa et al., 2016) and other South American countries (INEEd, 2017) for all ages. These differences can be tracked across a variety of domains, e.g., intuitive number sense, spatial reasoning. In the present work, we focused on assessing two domains of math achievement: arithmetic and geometry. Our measure of geometry was used primarily as a pilot and consisted of an assessment developed by the Spelke lab and, in a collaboration, translated into Spanish by one of our authors (A.M.) and used with permission (E. Spelke, personal communication, June 23, 2017). A typical question on this measure showed students several shapes and asked them, for example, to circle all that had sides of equal length. Another question in this pilot task assessed students' understanding of symmetry by asking whether a given shape was the same on either side of a dotted line. This measure will serve as a control to contrast with our main measure of interest: arithmetic ability.

To measure arithmetic ability, we used a version of the math fluency subtest of the Woodcock-Johnson Tests of Achievement (Woodcock, McGrew, & Mather, 2001/2007) adapted for use in Spanish-speaking contexts: the Batería III Woodcock-Muñoz (Woodcock et al., 2001/2007) (see Fig. 1). The test measures the number of addition, subtraction, and single-digit multiplication problems children can complete in 3 min. It is a reliable measure of arithmetic achievement and has been widely administered in many countries.

1 −1	0 +3	2 +2	4 −2	2 +1	3 −3	0 +0	3 −0	2 −1	2 +4
5 +0	3 −1	1 +6	4 +4	5 −0	1 +1	6 −1	3 +5	4 −1	5 −2
3 −2	5 +1	6 −3	2 −2	7 +1	4 −4	1 +8	4 −3	7 +2	4 +1
2 +5	8 −1	5 −4	3 +3	10 −2	3 +6	7 −2	2 +8	3 +1	9 −4
6 −2	4 +6	9 +3	8 −6	7 +5	10 −10	2 +6	5 −3	6 −6	3 +4
5 +5	8 −3	5 −1	8 +0	7 −4	1 +9	10 −6	8 +4	6 +8	9 −9
1 ×1	4 +5	7 +7	2 ×3	10 −5	1 ×2	3 +2	8 −8	9 +5	5 ×1
3 +7	1 ×4	9 −2	9 +8	3 ×3	6 −5	2 ×2	9 −3	3 ×1	6 +6

FIG. 1 An example page from the Woodcock-Muñoz speeded arithmetic assessment. Students had 3 min to complete as many problems as possible.

Because norming data for this test do not exist for Uruguay, here we report the raw scores (i.e., the number of problems completed correctly), age-normalized scores (normed to our sample), and regressions controlled for age, geometry, IQ, and vocabulary where appropriate. We expect that our children will differ in math ability (geometry and arithmetic) as a function of SES. We will also look at how this relates to IQ and ANS ability and Repeater status, and how these factors may influence the effectiveness of the intervention.

Approximate Number Abilities

Our magnitude training intervention approach was inspired by work in developmental psychology that has focused on the relationship between classroom mathematics and an intuitive number sense mediated by the approximate number system (ANS: Halberda & Feigenson, 2008). The ANS is a core knowledge system (Feigenson, Dehaene, & Spelke, 2004) that is present in newborns (Izard, Sann, Spelke, & Streri, 2009) and improves over the course of development (Halberda & Feigenson, 2008; Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Odic, Libertus, Feigenson, & Halberda, 2013). The system supports people's intuitive understanding of the approximate quantity of collections of items or events and supports the ability to add and subtract these quantities. Such a system might be specific to number, or the ANS may be part of a more general magnitude system that includes, e.g., surface area, time, and length (Lourenco, Bonny, Fernandez, & Rao, 2012; Sokolowski, Fias, Ononye, & Ansari, 2017; Walsh, 2003). Because we want to remain open to the possibility of a generalized magnitude system, we created mini-games for surface area, time, and number. However, in previous work we have found that each of these dimensions uniquely correlates with school math ability, suggesting some independence among these dimensions (Odic et al., 2016). Our own proposal is that the shared aspects of these representations may derive from the computations that allow one to compare magnitudes (e.g., to determine that one duration is longer than another, or that one size is larger than another), rather than from shared representations of quantity (Odic, Pietroski, Hunter, & Lidz, 2013). For the current paper, we are interested in improvements in this or any other parts of these systems, and in the possibility that training in magnitude games can transfer to speeded arithmetic performance.

A critical component of these ongoing debates is whether ANS precision plays a causal role in shaping formal math abilities. Intervention studies provide a tool to assess this hypothesis by intervening to improve ANS precision and observing if this leads to an improvement in formal math ability. A handful of research projects have looked at classroom interventions using computerized training programs to improve the precision of cognitive abilities, such as the ANS, that are associated with school math performance (DeWind & Brannon, 2012; Hyde, Khanum, & Spelke, 2014). There have been mixed successes, and more research is needed to understand what kinds of interventions are most successful and for which students. One possibility is that these types of ANS interventions may be particularly helpful in fostering foundational understandings of basic concepts or in situations where a child is struggling to attain proficiency with early skills.

In the present work, our measure of ANS ability will be the number of problems solved in a timed paper-and-pencil dot comparison task (Fig. 2). This is a pilot version of a task that we are actively developing which aims to measure ANS ability in a group- or individual-administered, paper-and-pencil format (Mailhos et al., 2018). We expect to find differences in ANS ability across SES in pretest, and we also expect to replicate results showing a correlation between ANS ability and arithmetic ability (DeWind & Brannon, 2012; Hyde et al., 2014; Odic et al., 2016; Zosh, Verdine, Halberda, Hirsh-Pasek, & Golinkoff, 2018). Lastly, we will also look at whether SES,

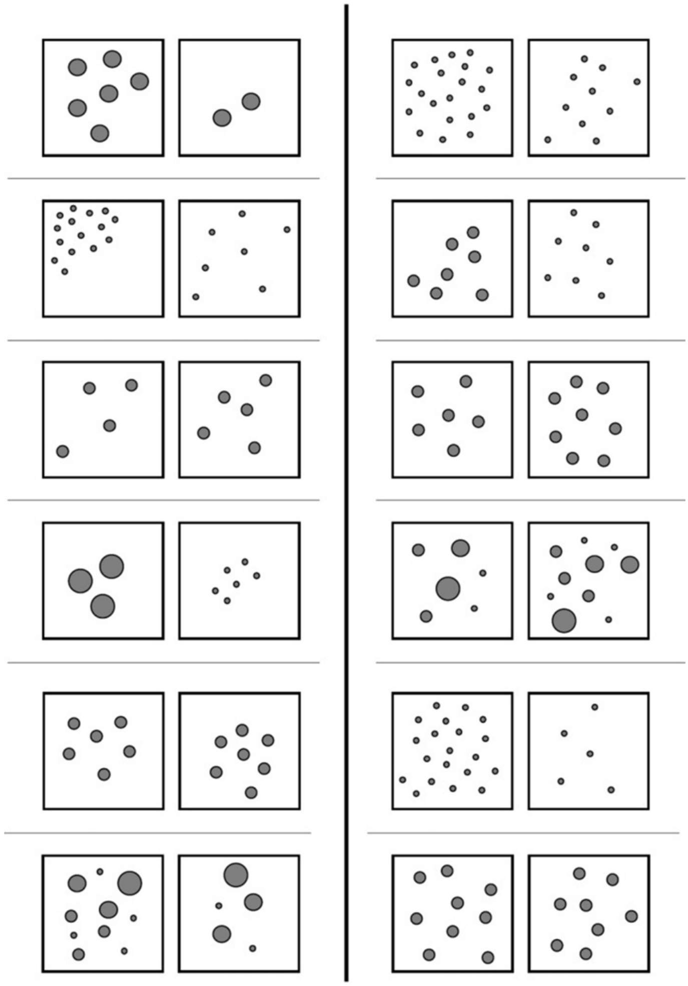

FIG. 2 Sample page from modified dots discrimination task. Students were instructed to consider each pair of squares and place an X in the box with more dots in it. They were given 3 min and told to begin at the top left and continue down each column until time ran out.

Grade-Repeating, and IQ affect the relationships between ANS ability, arithmetic ability, and intervention transfer.

THE PRESENT INTERVENTION STUDY

Our primary goal in partnering with communities in Montevideo was to create a resource that could improve students' math performance and that students could use in the classroom and at home. To do this, we aimed to develop an app that students would be motivated to engage with over an extended period of time and that trained approximate magnitude comparison over various domains.

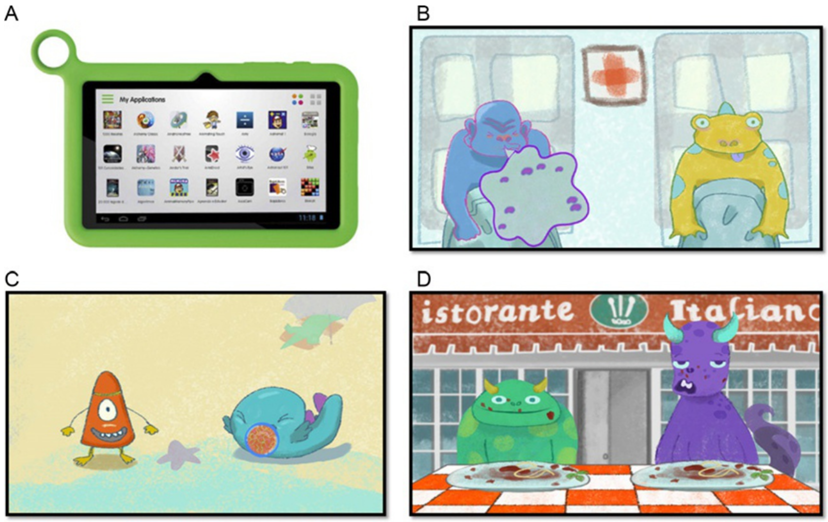

FIG. 3 (A) The tablet children used in the intervention and sample images from trials of each of the three training games: ANS (B), area (C), and time (D).

To make the intervention engaging for students, we developed a completely novel app that included extensive original artwork and characters (e.g., "monsters"), rewards for good performance (e.g., "stars"), and online tracking of gameplay. The app included three mini-games (see Fig. 3B–D), each focusing on one of the three abilities that have been discussed in the literature as being relevant to general magnitude representation: time, area, number. We sought to improve students' math abilities through training on these magnitude comparison tasks over the course of a 1-month intervention.

Teachers' Responses to the Software

Successfully translating from the lab to the classroom requires building bridges between researchers and educators, and our project cannot succeed without the work of teachers invested in partnering with us for the long term. Not only are teachers responsible for implementing the intervention in their classrooms, but their feedback is our primary means of understanding how students interact with the app. Nearly all the teachers whose classrooms participated in the current study were also with us for a separate pilot phase, and their feedback was particularly valuable in helping us to understand how to motivate the children to continue engaging with the games. For example, teachers suggested that students be able to personalize their game sessions by selecting an image as an avatar. They also suggested introducing a progress-tracking system into the game that would allow students to see how they were doing and compare their performance with their peers. We implemented these ideas in the version of the game for the current study, and we heard from teachers that these changes improved student engagement. At the end of the intervention, the teachers met with our team in Uruguay and offered final comments on the intervention, summarized by our team here:

> *El juego en términos generales, fue evaluado positivamente por parte de las maestras, la mayoría de ellas transmitieron el entusiasmo de los niños por jugar. La incorporación de las estrellas como premio fue de gran motivación para los niños, por ejemplo se veía que comparaban su progreso en los distintos mini-juegos con sus compañeros a la vez que los motivaba a seguir jugando. La temática del juego fue divertida y muy aceptada por los niños, con respecto a esto se observó que cada niño tenía un monstruo preferido, incluso nos compartieron dibujos de los monstruos. Para finalizar, luego de terminada la intervención los niños nos preguntaban si íbamos a volver con las tablets y el juego.*
>
> *The game, in general, was seen very positively by the teachers, the majority of whom communicated the children's enthusiasm for playing the game. The incorporation of the stars as a reward system was a great motivator for the kids, for example they saw students comparing progress in the different mini-games with their friends which motivated them to continue playing. The theme of the game was fun and very well accepted by the children, and with respect to this it was observed that each child had a favorite monster, and they shared with us drawings of these monsters. Finally, after the end of the intervention, the children asked us if we were going to use the tablets and play the games again*

These comments suggest that students enjoyed engaging with the training app and teachers were easily able to incorporate it into their normal lessons. These are key components of a classroom-based intervention.

Design of the Current Study

As mentioned previously, the study implemented a pretest, intervention, posttest design with a Business-As-Usual (BAU) Control group. We assessed formal Arithmetic Abilities (i.e., the number of single- and double-digit addition, subtraction and multiplication problems students could complete in 3 min), IQ [Raven's Progressive Matrices (Raven, Raven, & Court, 1998)], Approximate Number Abilities (the ability to rapidly determine the larger of two approximate quantities on our paper assessment), Vocabulary, and Geometry Ability. In our analyses, we focus on the diversity of our Uruguayan school sample in terms of SES, Grade-Repeating, IQ, and Arithmetic Ability, and we present improvement scores from pre- to posttest for both our Game Intervention group and our BAU Control group. Figure 4 illustrates the study's overall design. The boxes on the left and right represent the pre- and postintervention evaluations, respectively. All students participated in these evaluations. Classrooms at each of the participating schools were randomly assigned to either the Game

FIG. 4 Study design. Both groups (Intervention and Control) participated in pre- and postintervention evaluations of cognitive abilities. The intervention group was encouraged to play the three tablet games during the 5-week intervention period, during which the control group received Business-as-Usual instruction.

Intervention condition or the BAU Control condition for the five-week intervention period. During that time, students in the Game Intervention group interacted with the intervention software which consisted of three magnitude training games, each focusing on a different ability: area discrimination, time discrimination, and approximate number discrimination. Note that, consistent with Uruguayan school policy, students in the BAU Control condition also had access to tablets during this time as part of their normal curriculum, but they did not have access to the games we developed for the intervention.

Methods

Participants

Sixteen classrooms in eight public elementary schools in and around Montevideo participated in the study, including 386 s- and third-graders overall. Half of the classrooms were assigned to the intervention condition and the training game was installed on tablet computers provided by Plan Ceibal. During the 5-week intervention period, students in this group were encouraged approximately once per week by their teachers to play the game at school for a little while and at home as much as they liked. Otherwise, children in the Game Intervention group received Business-As-Usual (BAU) math instruction. The other eight classrooms comprised a BAU Control group and received normal math instruction. These students had access to tablets but not the training game and, as is usual for schools in Uruguay, use of the tablets was left to the discretion of the classroom teacher. For this study, it was important to us that the teachers in each group felt that they could instruct as normal, along with the small addition of access to the game for the Game Intervention group.

In Fig. 5 we see a map of Montevideo and the surrounding areas where our schools were located. Each school is lettered and the metrics for each school are summarized in Table 1. In this study, SES information was only available

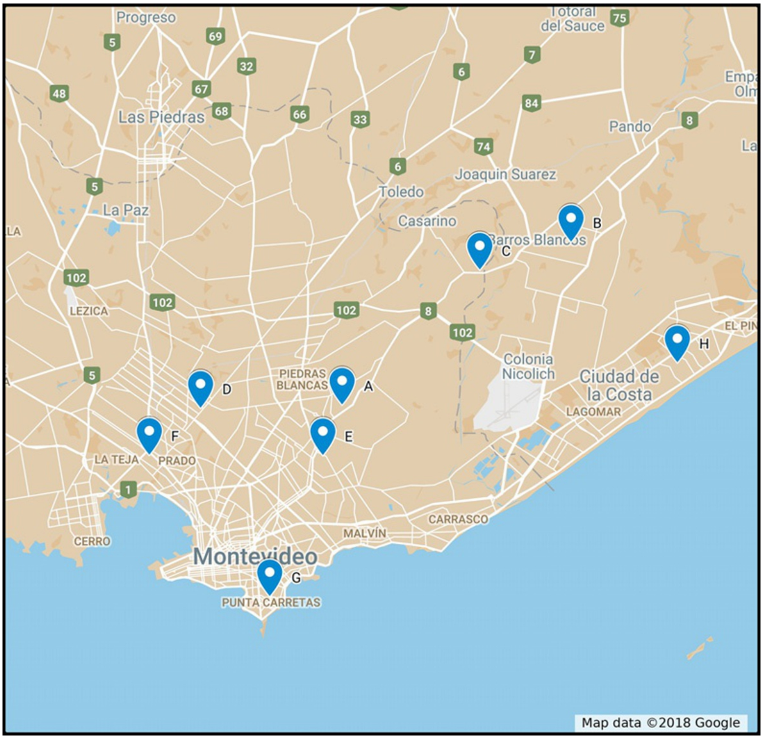

FIG. 5 Locations of our schools. *From Google Maps image used with in accordance with Google's fair-use guidelines: https://www.google.com/permissions/geoguidelines/attr-guide/.*

at the school level (as mentioned previously). In all but two schools, each intervention classroom was matched with a same-year (second- or third-grade) control classroom at the same school. All children spoke Spanish as their first language and all tasks were administered in Spanish. The posttest evaluations were administered by a researcher blind to the experimental conditions of the classrooms (J.L.).

Of the 386 students enrolled in classrooms that participated in the study, 19 students (5%) did not attend class on both pre- and postintervention testing days, and some students (overall 32%) were in attendance on one of these days but not both. As mentioned earlier, attendance at school depends to an extent on SES Quintile, and this is reflected in which students tended to be present for our pre- and posttest measures, as summarized in Table 2. Additionally, descriptive statistics about our sample (e.g., age ranges by condition and Grade) can be found in Table 4. Analyses were conducted using data from all students who completed the relevant measures for each statistical test.

Evaluation of repeater status. Our analyses considered the interaction between Repeater Status on cognitive abilities (such as IQ) and achievement (e.g., arithmetic ability). We considered Repeaters to be any children who had already repeated or were currently repeating a grade, and we determined

TABLE 4 Descriptive Statistics for Students Included in Analyses

Class	Condition	Mean Age (mo)	SD Age	Range (mo)	Females	Total
Second grade	Control	95.24	6.91	85.90–117.10	59	128
Second grade	Intervention	94.22	6.65	85.50–117.40	72	127
Third grade	Control	107.17	7.90	97.73–127.63	31	61
Third grade	Intervention	106.55	6.63	97.73–127.50	35	70
Overall	Control	99.13	9.14	85.90–127.63	90	189
	intervention	98.79	8.92	85.50–127.50	107	197

All ages are reported here as-of the pretest date before the 5-week intervention period.

this by comparing each student's birthday with the state-mandated cutoff date for their grade. By law, Uruguayan schools enforce strict cutoff dates for matriculation; it is illegal for parents to choose to hold a child back and start them later in school, a practice that occurs in some other countries like the United States. While our records did not indicate for each child the reason why they were a year older than their peers, the teachers reported that nearly all of these older children had been asked to repeat a grade because of poor performance. In our dataset, we had roughly the same percentage of Repeaters in Grades 2 and 3 (24% and 27%, respectively).

Materials

During the 5-week intervention period, students engaged with the training app using touchscreen tablets specifically designed for children, which they had access to throughout the training period. As mentioned previously, the training app consisted of three mini-games, each designed to probe a distinct modality of magnitude discrimination: time, area, and approximate number. The games could be played without a connection to the internet. Because use of the app was left to the discretion of the teachers in school, and the children themselves while at home, the number of times each child played each game was free to vary; however, in practice, children played each game multiple times over the course of the intervention; see *Engagement with the intervention* in "Results" section.

The cognitive and achievement tests were administered to all students in the study before and after the intervention period by trained researchers following a written protocol. As mentioned previously, we measured ANS Ability, Geometric Ability, Arithmetic Ability, Vocabulary size, and IQ (see Fig. 4). We also tracked SES at a school level using government-reported data (ANEP [Administración Nacional de Educación Primaria], 2012).

Intervention Games

Upon starting the intervention software, students saw the app's home page and could select one of the three mini-games by tapping on that game's icon. In each mini-game, students saw 12 trials in which the ratios of magnitudes (i.e., durations of time, numbers of items, and surface areas) were systematically varied, and they were asked to tap on the larger magnitude. Students received feedback after each trial indicating whether they made a correct response, and the games responded to three consecutive incorrect responses by repeating the game's instructions. After 12 trials the game ended, and a gold star appeared beside the number of correct answers made in that block. Tapping the screen returned the student to the app's home page, where the total number of correct responses across all blocks was displayed over the icon for each game.

Time discrimination. In the time discrimination game, students had to decide which of two sounds had a longer duration. The screen showed two monsters: a green one on the left and a purple one on the right (see Fig. 3D). Students tapped the screen to start a trial in which each monster took its turn to make a singing-like sound for a specific amount of time. To provide additional visual cues, the monsters' mouths moved, and they moved their hands to cover their mouths for 1 s after completing their sound. Students tapped the monster that they thought sang for longer and received feedback based on whether they answered correctly. After correct answers, the monsters would laugh and smile; incorrect responses elicited a "No, no!" sound while the monsters shook their heads. The singing sound varied from trial to trial to make the game more interesting.

Area discrimination. In the area discrimination game, students had to decide which of two monsters blew the largest "bubble" (see Fig. 3C). As in the time discrimination game, students saw two monsters: a blue one on the right and a yellow one on the left. After tapping the screen to begin the trial, they saw a multicolor circle grow in front of one of the monsters. The circle expanded briefly and stayed at a constant size for a brief interval before appearing to "pop" and disappear. Next, a similar bubble appeared in front of the other monster. The expansion and disappearing of the bubbles was accompanied by "blowing" and "popping" sound effects. After both bubbles disappeared, students tapped the monster they thought blew the biggest bubble. After correct responses, the monsters would laugh and dance to music, while incorrect responses caused them to shake their heads and say "No, no!."

Approximate number discrimination. In the number discrimination game, students were asked to tap the monster that "sneezed" the most "germs" (see Fig. 3B). As in the other games, they saw two monsters on the screen—one on the left and the other on the right. At the start of the trial, the first monster scrunched up its face and made a sneezing sound as a large sneeze-cloud with a certain number of "germs" appeared under its nose. After less than a second, the cloud of germs corresponding to the first monster disappeared and the second monster took its turn to sneeze a cloud of germs. The germs consisted of a discrete number of bounded shapes with different patterns and textures across trials. On some of the trials, the shapes were of equal size, and on other trials the size of the germs varied. On half of the trials, the total area of the germ shapes was equal between monsters (viz., the total area was controlled across numbers), and in the other half of trials the size of each germ and the total area varied (viz., the total area did not consistently correlate with the number of germs). These controls ensured that both individual germ size and total germ area were poor predictors of number in our stimuli. As in the other two games, students received auditory and visual feedback for the responses in the form of the monsters laughing and dancing for correct responses or shaking their heads (i.e., for incorrect responses).

Results

Our most important measures were Arithmetic Ability (Batería III Woodcock-Muñoz), ANS Ability (paper/pencil ANS task), SES (as reported by ANEP), Repeater Status (determined by students' birthdates), and IQ (Raven's Progressive Matrices). We also include the brief geometry and vocabulary tests as pilot control measures. We first investigated pretest differences in Arithmetic Ability, ANS Ability, and IQ as a function of SES, Grade, and Repeater Status (i.e., Repeater vs. Non-Repeater). In our analysis of pretest abilities, we grouped both Game Intervention and BAU Control children together because, at this point, no children had received an intervention.

IQ and Repeater Status by SES Quintile

Independent from our interests in math intervention, factors like IQ, grade-repeating, and SES may have interesting interdependencies of their own. We first asked whether children's IQ might contribute to determining which children did and did not repeat a grade as a function of SES. In Fig. 6 each dot represents a child's age-normalized percentile score on the Raven's Progressive Matrices (IQ) assessment, arranged by SES Quintile and jittered along the *x*-axis to reveal all children. Filled-in and open dots correspond to Repeaters and Non-Repeaters, respectively. The filled and open gray dots indicate means for each group, with bars representing standard error. Because Raven's scores are age normalized to each student's birthdate, the scores in Fig. 6 can be compared across Grades 2 and 3, and across Repeaters and Non-Repeaters, that is, even though Repeaters were older than their classmates they are not favored on this measure.

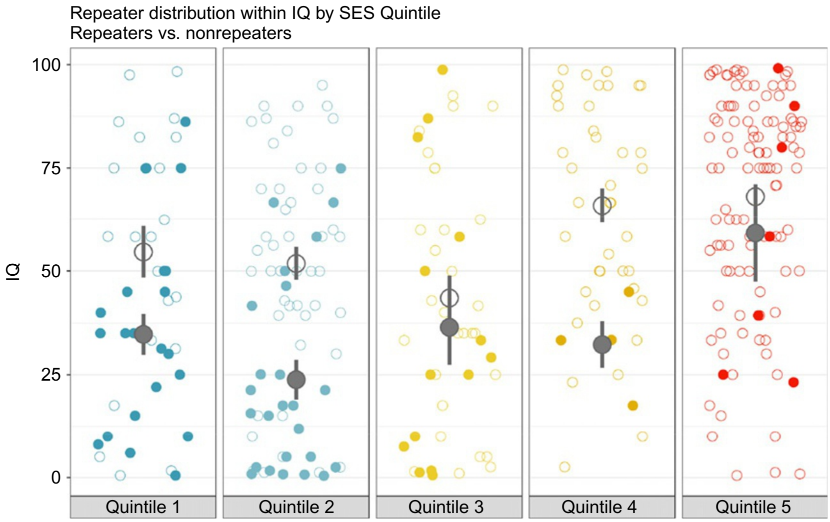

FIG. 6 Repeater Status compared with IQ within each SES Quintile. *Filled-in dots* indicate students who repeated a grade; *open circles* indicate students who never repeated a grade. *Dark gray dots* indicate group means for Repeaters *(filled)* and Non-Repeaters *(open)*, with bars indicating standard error of the mean.

The first thing we notice is that every Quintile contains the full range of IQ scores (i.e., we see scores from approximately 0–100 in each Quintile). Second, there is a small effect of SES on IQ in both Repeaters and Non-Repeaters, with higher SES corresponding to somewhat higher IQ scores on average, though this trend is perhaps minor compared to the broad similarity in spread that we see in this measure (0–100). This difference in measured IQ as a function of SES replicates patterns previously demonstrated in the literature (Turkheimer et al., 2003).

One of the most dramatic patterns in Fig. 6 is that the proportion of Repeaters is much higher in the lower quintiles compared to the higher quintiles. The effect of SES on the percentage of students who were repeating a grade is so dramatic that the difference between SES Quintile 1 (disadvantaged) and SES Quintile 5 (advantaged) was nearly eightfold in our sample: 41.8% of students in SES Quintile 1 had or were repeating a grade while only 5.4% of students in SES Quintile 5 had repeated or were repeating a grade. This trend is noticeable in Table 4, but in Fig. 6 we can see the pattern at the level of the child by looking at the ratio of filled to open circles as a function of SES.

Finally, in Fig. 6, we see that the mean IQ tends to be lower for Repeaters compared to Non-Repeaters. This result is important for considering how to effectively help these struggling students. Further, it is noteworthy that Repeaters tend to cluster at the bottom of the IQ scale in Quintiles 1 and 2, whereas the mean IQ of Repeaters was similar to Non-Repeaters in two of the three upper Quintiles (and note that in the one upper Quintile where this was not true, Quintile 4, there were only four Repeater children and so a larger sample may be required before concluding that this Quintile would show a significant difference). Interestingly, while lower IQ is linked with repeating a grade in SES Quintiles 1 and 2, it is less linked in the more advantaged schools (SES Quintiles 3 and 5).

Thus while it is true that many factors may impact performance on IQ tests like the one we have used here, our data support the inference that SES serves as a protective factor against repeating a grade. One possible explanation is that lower IQ students who might otherwise repeat a grade may receive more support in higher SES schools compared to students in disadvantaged schools and therefore be able to follow the standard trajectory. Conversely, students in low SES schools appear to need a higher IQ to avoid having to repeat a grade. This interpretation would be consistent with recent work showing that SES may moderate the impact of environmental factors that affect school performance. More specifically, higher SES mitigates the potential negative effects of factors such as low IQ (Tucker-Drob & Bates, 2016; Turkheimer et al., 2003).

Pre-Intervention Arithmetic by Grade and Repeater Status

Next, we considered Pre-Intervention Arithmetic Ability. Fig. 7 shows the performance for four groups of children across SES and Grade (i.e., Grade 2 Non-Repeaters, Grade 2 Repeaters, Grade 3 Non-Repeaters, and Grade 3

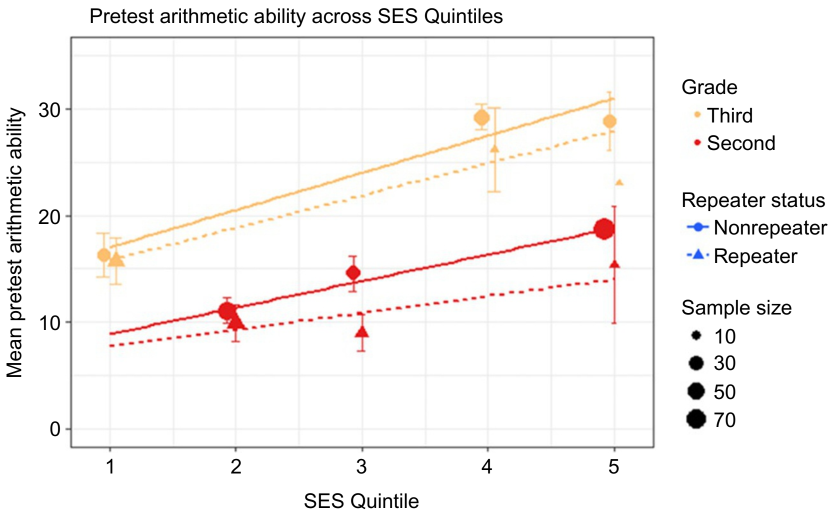

FIG. 7 Pretest Woodcock-Muñoz Arithmetic Ability score by SES quintile separated by Grade (second and third) and Repeater Status (Repeater vs. Non-Repeater). For this figure, four separate planned regressions were run (one for each group). The effect of SES on performance is apparent in each group (all $n \geq 35$), indicated by the upward slope of the lines as SES Quintile increases (all $P < .01$). The *top pair of lines* represents third-graders, while *bottom pair* represents second-graders (reflecting the positive difference that one additional year of school has on speeded arithmetic performance); *dashed lines* through *triangles* and *solid lines* through *circles* correspond to Repeaters and Non-Repeaters, respectively (these lines are regression lines based on the raw data from each child in the group). The size of each triangle and circle is proportional the sample size in that category. *Symbols* within each Quintile have been slightly jittered horizontally to prevent overlap.

Repeaters) in our assessment of Arithmetic Ability. If SES affects Arithmetic Ability, we should see a positive trend across SES with increasing Arithmetic Ability scores as SES Quintile increases. In Fig. 7, we see this pattern of higher scores for children in higher SES schools across Grade and Repeater Status as revealed in a regression of Arithmetic Ability by SES [$F(1,314) = 38.81$, $P < .001$, $R^2 = .11$)]. This reveals that SES impacts arithmetic ability and that children in higher SES Quintiles do better than their peers in less advantaged schools, as found in other studies (Goldin et al., 2014; Klibanoff et al., 2006; Odic et al., 2016; Valle-Lisboa et al., 2016; Wilkinson & Pickett, 2010; Zosh et al., 2018).

Considering Repeater Status in Fig. 7, note that despite Repeaters being a year older than their Non-Repeater classmates they nevertheless performed below the level of the Non-Repeater children on the pretest of Arithmetic Ability as revealed by planned *t*-tests within each Grade (Second Graders: $t(72.53) = 3.47$, $P < .001$; Third Graders $t(49.68) = 3.35$, $P < .01$). This highlights the importance of considering Repeater Status as a factor influencing students' performance.

ANS and Arithmetic Ability

Next, we considered the relationship between Arithmetic Ability and ANS Ability in our sample. There is a continuing debate concerning whether ANS abilities are related to school mathematical abilities (Bugden & Ansari, 2016; Clayton, Gilmore, & Inglis, 2015; DeWind & Brannon, 2012; Fazio, Bailey, Thompson, & Siegler, 2014; Gilmore et al., 2013; Odic et al., 2016; Xenidou-Dervou, Molenaar, Ansari, van der Schoot, & van Lieshout, 2017). We used the preintervention scores from all children to determine whether Arithmetic Ability is correlated with ANS Ability. The top plot in Fig. 8 shows the simple correlation

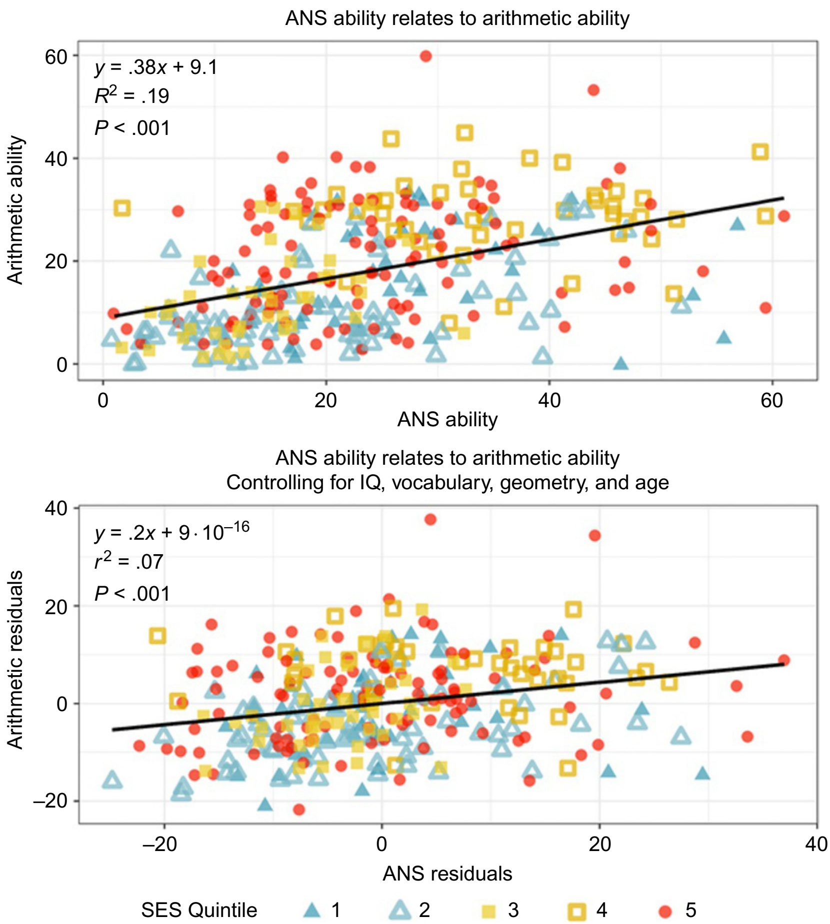

FIG. 8 Correlation and partial correlation (controlling for IQ, Vocabulary, Geometry Ability, and Age) between ANS ability and arithmetic ability. The top figure shows the simple correlation between the measures; the bottom figure shows the partial correlation. Each *symbol* corresponds to a child, and the *shapes* indicate the child's SES Quintile. Both regressions are significant and each SES Quintile contributes to these effects.

between these measures while the bottom shows the partial correlation of Arithmetic Ability and ANS Ability controlling for IQ, Vocabulary, Geometry Ability, and Age. The colors correspond to the SES Quintiles of students' schools. The first thing we notice is that the relationship is significantly positive in both cases ($P < .001$) suggesting a relationship between ANS and Arithmetic Ability, even when controlling for other factors (correlation: $F(1{,}304) = 70.07$, $P < .001$, $R^2 = .19$; partial correlation: $F(1{,}304) = 21.23$, $P < .001$, $r^2 = .065$). An additional age-normalized regression (not shown) between ANS Ability and Arithmetic Ability was also significant (this regression is one way of creating age-normalized standard scores for these tasks): $F(1{,}316) = 67$, $P < .001$, $r^2 = .17$.

Second, we see that the colored dots are evenly spread throughout the trend, indicating that no single SES Quintile is driving the effect. Indeed, individual regressions were also performed for each SES Quintile separately and the regression slopes were all positive suggesting that all SES Quintiles contribute to this effect. Therefore contrary to some claims in the literature (Bugden & Ansari, 2016; Clayton et al., 2015; Gilmore et al., 2013; Xenidou-Dervou et al., 2017) and consistent with others (DeWind & Brannon, 2012; Fazio et al., 2014; Odic et al., 2016) our data show a link between ANS Ability and Arithmetic Ability (across SES Quintiles and controlling for many relevant measures).

Engagement With the Intervention

Thus far we have seen noteworthy effects of SES and Repeater-Status on several measures in our sample. We also found that ANS Ability correlates with Arithmetic Ability while controlling for many factors. Before turning to consider how performance changed from pre- to posttest, we first report children's engagement with the intervention games. Recall that children in the Game Intervention group ($n = 197$) had access to three discrimination games on their tablets for 5 weeks and they were free to play them as much or as little as they liked while at home. Fig. 9 shows children's total number of games played over the course of the 5 weeks. The total number of games played suggests that children liked the games and did volunteer to play them. Because each discrimination game lasted approximately 3 min per play, this intervention can be considered a brief amount of exposure to the intervention games over the course of the 5 weeks.

Pre- to Postintervention Improvement

We turn now to considering how performance changed from pre- to posttest. Fig. 10 shows the percentage change from pretest to posttest performance (viz., performance 5 weeks later), collapsing across SES Quintiles, Grade, and Repeater-Status, for both Game Intervention and BAU Control children on each of our measures: Arithmetic Ability, ANS Ability, Vocabulary, Geometry, IQ. We saw significant gains for all groups of children on all tasks as shown by planned *t*-tests for each bar in Fig. 10 (all Ps $< .05$ except

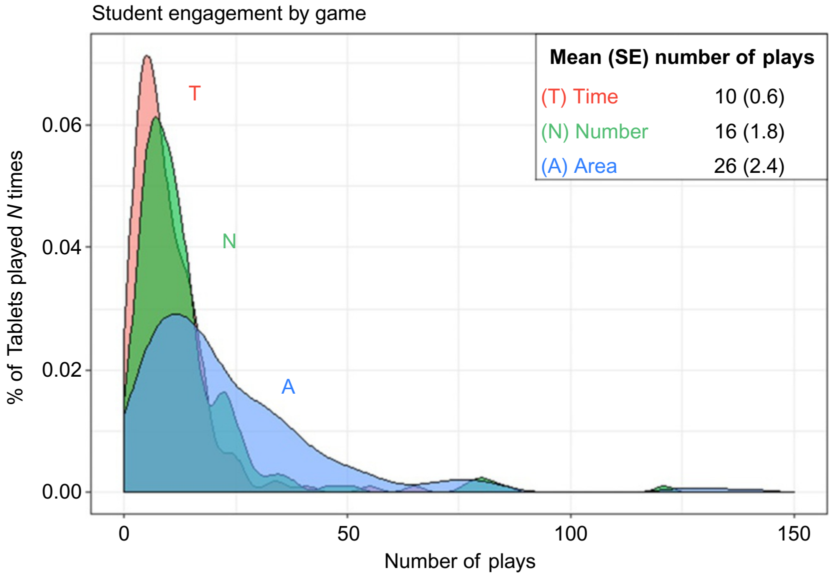

FIG. 9 Distribution of number of plays per child per game over the course of the 5-week intervention for all children in the Game Intervention group. The *graphs* are density distributions and show the proportion of children for each number of plays. Most children played each game around 10 times while the more active players pulled the mean number of plays higher for some games.

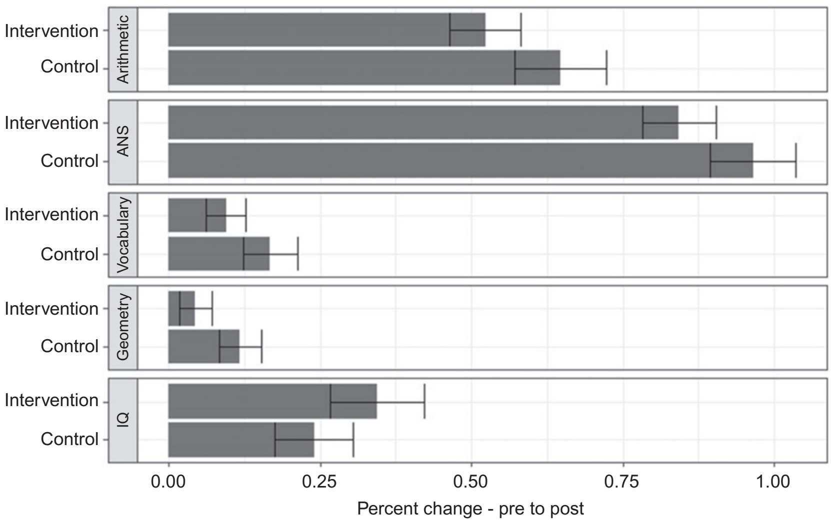

FIG. 10 Percent improvement on each assessment for both BAU Control and Game Intervention children. Both groups showed significant improvements from pre- to posttest on each of our measures (all $P < .05$ except Geometry, Intervention group $P = .09$), with the greatest improvement seen in the math assessments of Arithmetic Ability and ANS ability.

Geometry Ability in the Intervention group: $P = .09$). It is also noteworthy that we saw the greatest gains for our two math assessments: Arithmetic Ability and ANS Ability. Considering possible differences between the gains seen for BAU Control children and those seen for Game Intervention children, Fig. 10 does not reveal noteworthy improvements of the Game Intervention children above and beyond the BAU Control children. This is perhaps because of the large overall improvements we saw in all children. We next look at improvement as a function of SES Quintile and Repeater-Status as these factors may modulate who does and who does not improve.

We computed z-transformed change scores for both Arithmetic Ability and ANS Ability by taking each child's change in number of problems answered correctly from pre- to posttest and dividing by the SD of that child's classroom scores on the pretest:

$$\left(z_{diff} = \frac{Post_{child} - Pre_{child}}{SD(Pre_{child})} \right)$$

Recall that both our Arithmetic Ability assessment (Batería III Woodcock-Muñoz) and ANS Ability assessment were timed assessments which measured the number of problems correctly solved during 3 min. Dividing the difference from pre- to posttest by the SD of the pretest scores for each child's classroom is what makes this a z-normalized change score, and it has the consequence that each child's change score is roughly normalized across our factors of interest. That is, a z-normalized change score from a Grade 2 child in the Game Intervention group of the 5th SES Quintile will be comparable to the z-normalized change score from a Grade 3 child in the BAU Control group of the 1st SES Quintile. In both cases, the z-normalized change score indicates how much the child improved relative to the variability in the pretest scores of their immediate peers.

Arithmetic ability. In Fig. 11, the left side shows the z-normalized change scores in Arithmetic Ability for Non-Repeater children and the right side shows the z-normalized change scores for Repeater children. The first pattern to notice is that all of the mean change scores are above 0 (the level of no change), indicating that all groups of children improved from pretest to posttest. This duplicates the pattern of bars in Fig. 10. The change scores across groups seems to hover around a value of 1. This means that children, in general, improved about 1 SD above the mean of their classroom group from pretest to posttest.

Next, on the left side we see that for no SES Quintile did the Game Intervention children improve significantly more than the BAU Control children, with the exception of trends in the 1st and 3rd SES Quintiles. Being conservative, we can say that this intervention did not show noteworthy improvement over the test-retest improvement in the matched BAU Control classrooms for children who did not repeat a grade.

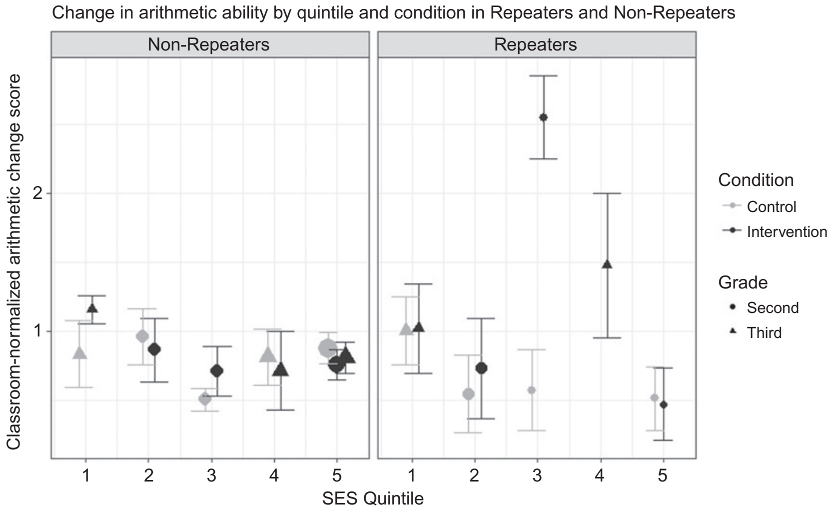

FIG. 11 Change in arithmetic ability by quintile and condition (BAU control vs. game intervention) for Repeaters (right) and Non-Repeaters (left).

Turning now to children who were or had already repeated a grade (right side), we see two noteworthy standouts for *z*-normalized change scores: the 3rd SES Quintile Game Intervention group and the 4th SES Quintile Game Intervention group. Planned *t*-tests compared the Game Intervention children to the BAU Control children in each SES Quintile. The test for the 3rd SES Quintile Repeaters (Game Intervention, $M=2.55$, $SD=0.600$; BAU Control $M=0.575$, $SD=0.588$) revealed that the Game Intervention children showed improvement significantly greater than the BAU Control children [$t(5.998)=-4.693$, $P<.01$]. The children in the 4th SES Quintile Game Intervention group are showing higher *z*-normalized change scores than most other groups; however, we did not have a child in the 4th SES Quintile BAU Control group who was a Repeater, so we cannot carry out the same *t*-test on this group. These groups suggest that the game intervention was more effective than simply test-retest improvement in at least some SES Quintiles for children who had or were repeating a grade and that there may be trends for lower SES children who did not repeat a grade (e.g., SES Quintiles 1 and perhaps 3), but these trends were insufficient in our sample to overcome the already large improvements seen in the BAU Control children and must be taken as merely suggestive.

As one indication that these improvements may generalize—given greater training or more effective yoking of pre- and posttesting to minimize the large test–retest improvements we saw here in the BAU children—we looked at whether the groups who showed noteworthy gains were different from their peers (e.g., examining whether their gains were an epiphenomenal result of their having extremely low pretest scores, which made large gains easier to

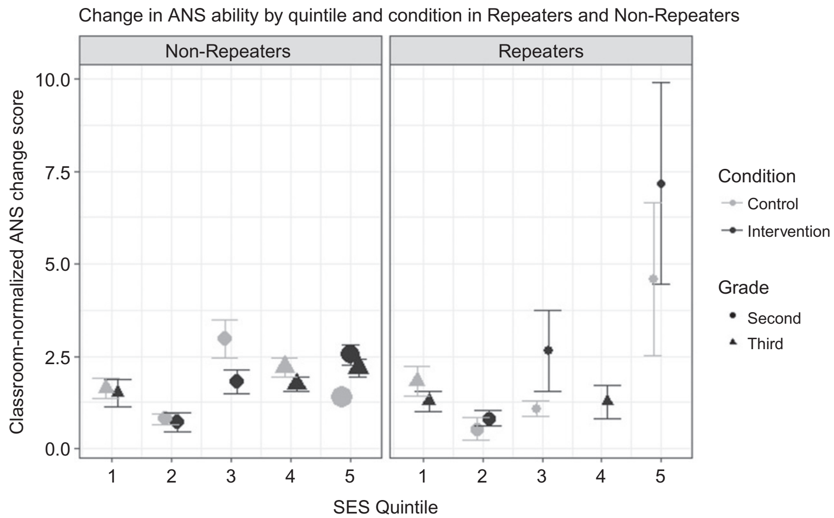

FIG. 12 Change in ANS ability by SES quintile and condition (BAU control vs. game intervention) for repeaters (right) and non-repeaters (left).

achieve). In both pretest ANS Ability and pretest Arithmetic Ability, children who improved more after the Game Intervention were no different than their peers—suggesting that these improvements did not merely result from aberrantly low pretest scores for these children.

ANS ability. Next, in Fig. 12, we consider the *z*-normalized change scores for performance on the ANS Ability assessment. Here again, we computed the change in number of correct answers from pre- to posttest and divided by the SD of scores within each child's classroom. In Fig. 12 we see these scores separated by intervention condition and Repeater Status. The first pattern to notice is that all of the mean change scores are above 0 (the level of no change), indicating that all groups of children improved from pretest to posttest. This again duplicates the pattern of bars in Fig. 10. The change scores across groups seem to hover around a value of 1.75. This means that children, in general, improved about 1.75 SDs above the mean of their classroom group from pretest to posttest.

On the left side we see that Game Intervention children in SES Quintile 5 improved significantly more than their peers in the BAU Control group as revealed by a planned t-test [$t(91.212) = -4.553$, $P < .001$]. However, children in SES Quintiles 3 and 4 showed the opposite trend with BAU Control children tending to show more improvement then the Game Intervention children [Q3: $t(20.019) = 1.890$, $P = .07$; Q4: $t(35.900) = 1.445$, $P = .16$]. Thus for children who were not repeating a grade, on the left side of Fig. 12, we cannot say that there was any systematic benefit of the Game Intervention above and beyond did the BAU Control children. All children improved on this assessment.

Considering now the children who had or were repeating a grade, on the right side of Fig. 12, we see a trend where Game Intervention children improved more than the BAU Control children in SES Quintile 3 ($p = .25$). However, none of these trends attained significance, perhaps due to the large variability from these small sample sizes.

Summary of Pre- to Postintervention Improvement

Across all measures, there is noteworthy improvement from pretest to posttest for both the Game Intervention and BAU Control children, most likely due to test-retest improvement as well as the positive effects of having a special visitor come to the classroom and engage children in special tests—note that the tester for the posttest (J.L.), who was blind to the experimental condition for each classroom, was a particularly special visitor as he is male (atypical in education settings in Uruguay), an American, and a Spanish speaker with an American-Iberian accent. This may have contributed to better focus and overall performance on posttest versus pretest (which was administered by female Uruguayan graduate students (D.L. & D.F.).

While this overall improvement made it difficult to see improvement in our Game Intervention children above and beyond the improvement in our BAU Control children, and in spite of the small sample sizes in some of these groups, we did see some significant results and trends suggesting that there may be a positive influence of the game-based intervention training. But for certain, the strongest results in the current sample concern the effects of SES and Repeater-Status on Arithmetic Ability, along with the differential links between IQ and these factors, and the relationship between Arithmetic Ability and ANS Ability in our sample. Continued work to tailor interventions to the community is necessary.

CONCLUSIONS AND FUTURE DIRECTIONS

Researchers and educators have a shared interest in understanding the factors that influence students' educational outcomes, and our work is consistent with previous findings that the socioeconomic status of a child's community is related to their performance in the classroom. Even before our intervention started, students in lower SES schools scored lower on measures of math achievement compared to students at more advantaged schools. We also saw that lower SES students are more likely to repeat a grade compared to higher SES students, and that having a lower IQ has a disproportionately negative impact on the likelihood of repeating a grade for lower SES children. One reason for this may be that lower IQ students at higher SES schools are better supported, either within the classroom or at home, compared to similarly scoring students at lower SES schools. An important future direction may be investigating the specific ways that more affluent schools succeed in supporting students at risk for repeating a grade so that these methods can be implemented at lower SES schools. Certainly availability of resources has a role to play here, but interventions such as the one presented in this study suggest that there are possibilities for reducing the gap between high- and low-achieving students using existing infrastructure.

One method for enriching the educational experience of vulnerable students may be through technology. The Uruguayan educational system provides a unique opportunity to investigate this question, given the country's investment in educational technology in schools that span the SES spectrum. While not conclusive at this stage, our results are consistent with the idea that the tablet can serve as an important vehicle for intervention and that it may be useful for improving math performance for students who are the most disadvantaged. What is more certain is that students seemed to show across-the-board improvement in a number of areas when special adults visited their classrooms and evaluated their performance. To the extent that lower SES students feel different societal expectations in terms of their academic potential, providing special attention to these students may prove an important tool in helping them reach the same levels of achievement as their more affluent peers.

The design of the current study had some limitations which should be addressed in future interventions. A more fine-grained way to measure the impact of the game would have been to look at dosage effects; however, this stage of the project included no standardization of the amount of time that the students interacted with the intervention media. In part, this was by design; the project is a collaboration with the teachers who agreed to use the game in their classrooms. By giving them the freedom to implement the game in the ways they saw fit, we stood to learn from their feedback about how best to integrate the intervention game with the normal classroom curricula. Future deployments of this intervention may include lesson planning guides based on suggestions from teachers in the current study.

While the effects of SES and Repeater Status are central to our interpretation of this study's results, we were unable to balance our sample of students by age across SES Quintiles because classrooms in the study were included based on their willingness to participate. For the most part, only one grade level was represented within each SES Quintile, so it made comparisons across grade levels difficult given the large effect of Grade on our outcome measures (viz., Arithmetic Ability score). Similarly, we would have liked to have been able to examine in more detail the effect of repeating a grade, but the number of Repeater students was impossible to control in the sample. Future studies might focus on schools with high levels of repetition to examine the effects of the intervention in these groups. Having a more complete sample would allow us to identify which students are most helped by our intervention; however, this would depend on the availability of classrooms willing to participate in the study.

This study's main goal was practical: to create a tool for teachers to use in the classroom that could improve students' math performance. However, these results also speak to important scientific questions about the relationship between underlying cognitive abilities and classroom math performance. One of the strongest results in this study was the evidence for a relationship between Arithmetic Ability and ANS Ability. We found that ANS Ability related to Arithmetic Ability across all SES Quintiles controlling for IQ, Vocabulary, Geometry Ability, and Age. The inclusion of these control measures was important because there is a continuing debate concerning whether ANS abilities are related to school mathematical abilities when controlled for other factors (Bugden & Ansari, 2016; DeWind & Brannon, 2012; Fazio et al., 2014; Gilmore et al., 2013; Odic et al., 2016; Xenidou-Dervou et al., 2017). Contrary to some claims in the literature (Bugden & Ansari, 2016; Clayton et al., 2015; Gilmore et al., 2013; Xenidou-Dervou et al., 2017) and consistent with others (Halberda et al., 2012; Odic et al., 2016) our data show a consistent link between ANS Ability and Arithmetic Ability across SES Quintiles.

Considering the importance of intervention research, to the extent that magnitude training can improve formal math ability, this suggests a causal link between intuitive number sense and children's formal understanding of math. Our data are consistent with other findings that have shown such a link (DeWind & Brannon, 2012; Odic et al., 2016; Odic, Libertus, et al., 2013), but further work is needed to explore more fine-grained questions about training type and specific formal skills. The intervention in this study trained children on three different magnitude tasks (area, time, and approximate number). A future study might look at the correlation between amount of student interaction with each of these games and amount of improvement in various achievement domains (e.g., addition/subtraction, geometry, time estimation). For example, if the partial correlation between ANS training and formal math ability were significant when accounting for approximate area training, this would suggest a privileged connection between the ANS system and symbolic math, as found by Lourenco et al. (2012) for adults. This would align with previous work that has shown a privileged relationship between ANS and symbolic math compared to symbolic math and time discrimination (Odic et al., 2016). It may also be the case that different magnitude training tasks are more strongly linked to performance in various outcome domains; for instance, we might observe privileged relationships between area training and formal geometry as well as between ANS training and formal arithmetic. These are important scientific questions about the foundations of human numerical understanding that could be addressed in future studies.

Understanding underlying mechanisms is key to progress in many areas of science. When we succeed in applying this understanding to improving the lives of people, our work becomes even more thrilling. The current project reveals that there are children who will benefit from interventions to improve math ability, and that these improvements may vary as a function of IQ and SES. It also shows that there is a link between ANS Ability and Arithmetic Ability across SES Quintiles. Nevertheless, our work also highlights the challenges of developing an intervention that will bridge the gap between game-based improvements and gains in math understanding. Through our partnership with teachers and students, we aim to continue exploring how equal access to technology and purpose-built software can improve educational outcomes, and ultimately help to create a more equitable society.

REFERENCES

ANEP [Administración Nacional de Educación Primaria]. (2012). *Relevamiento de contexto sociocultural de escuelas de educación primaria [Survey of primary school's socio-cultural context]*. Montevideo: ANEP. Obtenido de http://www.anep.edu.uy/anep/index.php/codicen-publicaciones/category/117-publicacionesdivision-de-investigacion-y-estadistica-educativa?download=1132:relevamiento-de-contextosociocultural-2010.

Bugden, S., & Ansari, D. (2016). Probing the nature of deficits in the 'Approximate Number System' in children with presistent developmental dyscalculia. *Developmental Science*, *19*(5), 817–833.

Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: the effect of different visual controls on ANS measurement. *Acta Psychologica*, *19*(5), 177–184.

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Frontiers in Human Neuroscience*, *6*, 1–10.

Dillon, M. R., Kannan, H., Dean, J. T., Spelke, E. S., & Duflo, E. (2017). Cognitive science in the field: a preschool intervention durably enhances intuitive but not formal mathematics. *Science*, *357*(6346), 47–55.

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123*, 53–72.

Feigenson, L., Dehaene, S., & Spelke, L. (2004). Core systems of number. *Trends in Cognitive Science*, *8*(7), 307–314.

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., et al. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE*, *8*(6).

Goldin, A. P., Hermida, M. J., Shalom, D. E., Costa, M. E., Lopez-Roesenfeld, M., Segretin, M. S., et al. (2014). Far transfer to language and math of a short software-based gaming intervention. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(17), 6443–6448.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": the approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457–1465.

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(28), 11116–11120.

Hart, B., & Risley, T. R. (1975). Incidental teaching of language in the preschool. *Journal of Applied Behavior Analysis*, *8*(4), 411–420.

Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378.

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances exact symbolic arithmetic in children. *Cognition*, *131*(1), 92–107.

INEEd. (2017). *Informe sobre el estado de la educación en Uruguay: 2015–2016*. Montevideo: INEEd.

Izard, V., Sann, C., Spelke, E., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10382–10385.

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool children's mathematical knowledge: the effect of teacher "math talk." *Developmental Psychology*, *42*(1), 59–69.

Larson, K., Russ, S. A., Nelson, B. B., Olson, L. M., & Halfon, N. (2015). Cognitive ability at kindergarten entry and socioeconomic status. *Pediatrics*, *135*(2), 440–448.

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science*, *16*(11), 841–845.

Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(46), 18737–18742.

Mailhos, Á., Maiche, A., Langfus, J., De Leon, D., Fitipalde, D., Sánchez, I., & Halberda, J. (2018). Towards a paper-and-pencil ANS assessment tool. (in prep.).

Odic, D., Lisboa, J. V., Eisinger, R., Gonzalez, M. O., Maiche, A., & Halberda, J. (2016). Approximate number and approximate time discrimination each correlate with school math abilities in young children. *Acta Psychologica*, *163*, 17–26.

Odic, D., Libertus, M., Feigenson, L., & Halberda, J. (2013). Developmental change in the acuity of approximate number and area representations. *Developmental Psychology*, *49*(6), 1103.

Odic, D., Pietroski, P., Hunter, T., & Lidz, J. (2013). Young children's understanding of "more" and discrimination of number and surface area. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 451–461.

Pilli, O., & Aksu, M. (2013). The effects of computer-assisted instruction on the achievement, attitudes and retention of fourth grade mathematics students in North Cyprus. *Computers & Education*, *62*, 62–71.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. San Antonio, TX: Pearson.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453.

Sokolowski, H. M., Fias, W., Ononye, C. B., & Ansari, D. (2017). Are numbers grounded in a general magnitude processing system? A functional neuroimaging meta-analysis. *Neuropsychologia*, *105*, 50–69.

Trucano, M. (2011). *What's next for plan ceibal in Uruguay?* EduTech: A World Bank Blog on ICT use in Education. 10 de June de. Obtenido de: http://blogs.worldbank.org/edutech/planceibal2.

Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene x socioeconomic status interaction on intelligence. *Psychological Science*, *27*(2), 138–149.

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, *14*(6), 623–628.

Valle-Lisboa, J., Cabana, Á. E., Mailhos, Á., Luzardo, M., Halberda, J., & Maiche, A. (2016). Cognitive abilities that mediate SES's effect on elementary mathematics learning: the Uruguayan tablet-based intervention. *Prospects: Comparative Journal of Curriculum, Learning, and Assessment*, *46*(2), 301–315.

Verdine, B. N., Irwin, C. M., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Contributions of executive function and spatial skills to preschool mathematics achievement. *Journal of Experimental Child Psychology*, *126*, 37–51.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Science*, *7*(11), 483–488.

Wilkinson, R., & Pickett, K. (2010). *The spirit level: Why equality is better for everyone*. London: Penguin.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001/2007). *Woodcock-Johnson III tests of cognitive abilities*. Rolling Meadows, IL: Riverside.

World Computer Exchange. (2016). *WCE FY16 annual report*. Hull, MA: World Computer Exchange.

Xenidou-Dervou, I., Molenaar, D., Ansari, D., van der Schoot, E., & van Lieshout, C. (2017). Nonsymbolic and symbolic magnitude comparison skills as longitudinal predictors of mathematical achievement. *Learning and Instruction*, *50*, 1–13.

Zhang, M., Trussell, R. P., Gallegos, B., & Asam, R. R. (2015). Using math apps for improving student learning: an exploratory study in an inclusive fourth grade classroom. *TechTrends*, *59*(2), 32–40.

Zosh, J., Verdine, B., Halberda, J., Hirsh-Pasek, K., & Golinkoff, R. (2018). Preschoolers' approximate number system varies by socio-economic status. (in prep.).