ORIGINAL PAPER

Development and validation of a semi-automated, scalable response to intervention framework in mathematics

Eudald Correig-Fraga[1,2] · Albert Vilalta-Riera[1,3] · Cecilia Calvo-Pesce[1,4]

Abstract

Response to Intervention (RTI) has been shown to be better at identifying and addressing learning difficulties than other frameworks. However, only a few educational systems have implemented it, and it can be difficult for schools to implement RTI on their own without policies in place. One solution is to create semi-automated digital tools that can assist with RTI without requiring a lot of extra time from the teacher. This study describes the development and implementation of such tools in schools that were not previously familiar with RTI. The effectiveness of the tools was tested by comparing 5 schools (149 students) that used RTI with 8 schools (269 students) that did not. The results showed that the extra practice was effective and that students in the intervention group made significantly more progress in arithmetic fluency than their peers in the control group. These students were also more likely to leave the "low achievement" and "at risk of Mathematics Learning Difficulties" zones than students in the control group. This suggests that it is possible to develop a scalable RTI framework and implement it in educational systems without RTI-based policies.

✉ Eudald Correig-Fraga
  eudald.correig@urv.cat

1   Innovamat Education S.L., Sant Cugat del Vallès 08174, Spain

2   Departament d'Enginyeria, Química Universitat Rovira i Virgili, Tarragona 43007, Spain

3   Departament de Didàctica de les Matemàtiques i les Ciències Experimentals de la Facultat de Ciències de l'Educació, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain

4   Escola Sant Gregori, Barcelona 08035, Spain

Theoretical background

Response to intervention

Response to Intervention (RTI) represents a major shift in educational methodologies. Instead of relying on reactive models, RTI places a significant emphasis on proactive and early interventions, specifically targeting students who present academic or behavioral challenges (Fletcher and Vaughn 2009; Hughes and Dexter 2011; Gersten et al. 2020). RTI's academic and societal intent seeks to ensure equitable opportunities for all students, irrespective of their inherent abilities or socio-economic backgrounds (Horner et al. 2017). Its distinct multi-tiered structure is crafted to optimize the distribution of educational resources, contrasting sharply with traditional models such as the IQ-discrepancy method used in the identification of learning disabilities (Fuchs and Fuchs 2006).

The RTI model is structured around a tiered system designed not only to identify students in need of additional support, but also to integrate specialized interventions with the core curriculum:

Tier 1: Universal instruction and screening

At this foundational level, all students participate in high-quality, research-based instruction provided within the general education environment. Periodic universal screenings occur, assessing students' progress and mastery of core curriculum content. These screenings serve as proactive measures to detect early signs of academic struggles, ensuring timely interventions (Greenwood et al. 2011).

Tier 2: Targeted interventions

For students identified through Tier 1 screenings as not making adequate progress in the core curriculum, Tier 2 offers targeted interventions. These interventions are tailored to students' unique needs, often in small group settings, and complement the foundational Tier 1 instruction. Progress monitoring tools are deployed more frequently here, providing educators with data on the efficacy of the targeted interventions and whether students are closing the academic gap (Cho et al. 2014).

Tier 3: Intensive interventions

This tier is reserved for students who, despite Tier 2 interventions, continue to display minimal academic progress. The interventions at Tier 3 are heightened in intensity, often delivered in one-on-one settings. The data from these interventions, combined with a comprehensive evaluation, can serve as a basis for considering students for special education services if required (Jimerson et al. 2016).

Throughout the RTI tiers, the connection between interventions and the core curriculum is vital. Each intervention layer is designed to scaffold and reinforce the foundational content provided in the core instruction. This structure ensures not

only the remediation of skill deficits but also the alignment of interventions with grade-level expectations and standards (Jimerson et al. 2016).

Extant literature underscores RTI's efficacy in both outcomes and cost-effectiveness compared to traditional methods (Van Der Heyden et al. 2007; Torgesen 2009). Of note is RTI's inclusiveness, especially benefiting low-socioeconomic-status students who might be underrepresented using older models like the IQ-discrepancy method (Fuchs and Fuchs 2006; Eccles and Roeser 2011).

RTI and multi-tiered system of supports

Another important educational framework is the Multi-Tiered System of Supports (MTSS). MTSS framework is widely adopted to optimize student outcomes by integrating both academic and non-academic interventions. MTSS is grounded in data-driven decision-making processes to identify and support students who may be at risk in various domains, from academic challenges to behavioral and socio-emotional concerns (Sugai and Horner 2003). Through its tiered structure, MTSS ensures that interventions are appropriately intensified based on individual student needs, promoting a holistic educational environment where students receive support tailored to their unique requirements (McIntosh et al. 2009).

Response to Intervention (RTI), while similarly tiered in its structure, is a narrower framework primarily concentrated on academic interventions. It serves as a mechanism to address learning difficulties in the early stages of education, typically during Kindergarten or grades 1 or 2, especially when children are acquiring foundational skills in areas like reading and mathematics (Gersten et al. 2017). In essence, while RTI can be viewed as a component or subset of MTSS focusing on academics, MTSS offers a more expansive and encompassing approach by addressing both academic and non-academic student needs.

Several studies highlight the effectiveness of both frameworks. RTI, with its proactive early academic intervention, has been correlated with improved student performance in reading and mathematics and a reduction in referrals to special education compared to traditional educational models (Fuchs and Fuchs 2006). MTSS, due to its broader scope, is associated not only with enhanced academic outcomes but also with a decrease in behavioral issues and an improvement in the socio-emotional well-being of students (Benner et al. 2013). In environments devoid of RTI or MTSS structures, the introduction of these systems often yields heightened positive student outcomes, streamlined resource allocation, and increased teacher satisfaction.

RTI's implementation

Despite these recognized advantages, few educational systems have implemented a RTI model (notable exceptions being in the USA (Berkeley et al. 2020) and Finland (Jahnukainen and Itkonen 2016)). The main reason is that, in most countries, the models of providing special education services are top-down, i.e., developed by governments and imposed on schools, but educational policymakers are historically slow at implementing innovations (Serdyukov 2017). A second challenge arises when

the model depends heavily on human resources for screening, monitoring, and intervention. This reliance makes it financially unfeasible for many educational systems. (Fletcher and Vaughn 2009).

One of the educational systems not implementing either RTI or MTSS is the Catalan, where this study takes place. In Spain, each autonomous community holds competence over its educational system. Catalonia, as one of these autonomous communities, dictates its educational parameters independently. Historically, within the Catalan education system—and, indeed, across the entirety of Spain—there has been no official policy or framework implementing either RTI or MTSS. Despite international research supporting these frameworks, they have not yet been integrated into Spanish regional educational systems. This absence, while in line with broader global hesitations towards systemic changes in education, underscores the potential benefits that could be realized should such frameworks be considered for future adoption. Notably, none of the participating schools in this study had prior knowledge or exposure to the concepts of RTI or MTSS. This unfamiliarity offers a unique opportunity: to observe the potential impacts and benefits of these interventions in an environment devoid of any preconceived notions or biases related to these specific educational strategies.

Difficulties in implementing RTI

Given the assessment-centric nature of RTI, refining traditional screeners, especially in mathematics, becomes a pressing concern. Current screeners often narrow their focus on specific mathematical concepts, which can neglect broader areas of mathematical understanding (Butterworth 2003). This narrow focus not only restricts their diagnostic reach, but can also misrepresent a student's holistic mathematical abilities (Reigosa-Crespo et al. 2012). Additionally, many of these tools bear cultural and linguistic biases, which may disadvantage diverse student groups in increasingly multicultural classrooms (Geary et al. 2017).

Moreover, the substantial reliance on teacher observations, prevalent in many screeners, introduces the potential for subjective biases. While the insights of educators are invaluable, relying solely on them may not paint a comprehensive picture of a student's genuine capacities or needs (Butterworth 2003).

Digital, semi-automated tools emerge as a promising solution. Not only can they reduce the human resource costs associated with RTI's implementation, but they can also enhance the accuracy and consistency of screenings and interventions (Wilson et al. 2006; Cho et al. 2014; Butterworth and Laurillard 2016).

Arithmetic fluency

Our study focuses on a specific academic area: arithmetic fluency. Arithmetic fluency, defined as the ability to solve basic arithmetic operations quickly and accurately, serves as a cornerstone for higher-level mathematical reasoning and problem-solving (Jordan et al. 2009). Jordan et al. (2009) elucidate that a solid grasp of arithmetic in the early years significantly predicts success in algebra and other advanced math topics later on. Furthermore, research by Vasilyeva et al. (2015) indicates that deficiencies in

arithmetic fluency can cascade into challenges in more complex mathematical tasks, as students struggle with basic calculations, diverting cognitive resources from grasping higher-level concepts.

Recognizing its paramount importance, interventions targeting arithmetic fluency have shown promise in not only enhancing this foundational skill but also in better equipping students for subsequent mathematical challenges (Geary 2011). By aligning RTI with these evidence-based interventions, there lies a profound opportunity to substantially impact students' mathematical trajectories.

In this study, two critical cutoff points, based on indices of the screener, were employed to categorize students based on their arithmetic fluency. First, a cutoff at the 30th percentile was used to define students in the "low average" category. This cutoff is common in educational research for differentiating students who perform below the average range but are not necessarily at significant risk for learning disabilities (Desoete and Grégoire 2006). Importantly, the 30th percentile was chosen as it lies on the upper end of commonly used cutoff percentages, which often extend as low as 25% (Geary et al. 2007). The rationale for selecting the upper limit was to capture as broad a spectrum of students as possible.

Second, a more stringent cutoff at the 15th percentile was used to identify students "at risk of Mathematics Learning Difficulties" (MLD). This percentile roughly corresponds to being one standard deviation below the mean, a metric regularly used in both psychological and educational research to identify significant deviation from average performance (Jordan and Hanich 2003). While the 15th percentile is at the upper limit of commonly used cutoffs—some research employs even lower percentiles like the 10th—it serves as an inclusive measure that aims to reduce the risk of false negatives (Jordan and Hanich 2003; Desoete and Grégoire 2006; Geary et al. 2007).

It is worth noting that both the 30th and 15th percentile cutoffs were chosen to be intentionally inclusive. Given that our assessment tools might not be as precise as those employed by trained psychologists, the upper-threshold cutoffs helped to mitigate the risk of false negatives. The focus was to ensure that students who might require additional support were not inadvertently excluded from receiving targeted interventions.

Study overview

In this research, the principal objective was to examine the efficacy of semi-automated digital tools designed to bolster the implementation of Response to Intervention (RTI) in schools unfamiliar with the framework. This inquiry was driven by the understanding that while RTI is empirically proven to be more proficient in identifying and addressing learning difficulties than traditional models, its adoption remains limited. This is due to the challenges associated with its implementation in the absence of institutional policies.

The study encompassed 13 schools in total, with 5 schools that integrated the RTI framework through the utilization of the digital tools, and 8 schools that continued their regular teaching methods without the aid of RTI—serving as the control group. The results highlighted two key findings: First, students in the intervention group

demonstrated a significant improvement in arithmetic fluency compared to their peers in the control group. Second, these students were also more likely to progress out of zones categorized as "low achievement" or "at risk of Mathematics Learning Difficulties". Together, these results point to the statistical significance and practical importance of the intervention in improving arithmetic fluency and helping students move out of risk categories.

This study explores the potential of these tools within a k-12 curriculum context, offering an innovative approach that integrates seamlessly with regular curriculum tools. This integration seeks to address barriers such as resistance from educators and administrators while allowing for a scalable, grassroots approach to RTI in educational landscapes where policy does not back it.

Materials and methods

Participants

The study involved 418 first-grade students from 13 schools in Catalonia, Spain (Fig. 1). Among these, 5 schools (comprising 149 children) were provided with RTI tools, while the other 8 schools (comprising 269 children) served as a control group. Among the intervention schools, 4 were public and had low socioeconomic status (SES), and the remaining one was a charter school with a high SES. The control
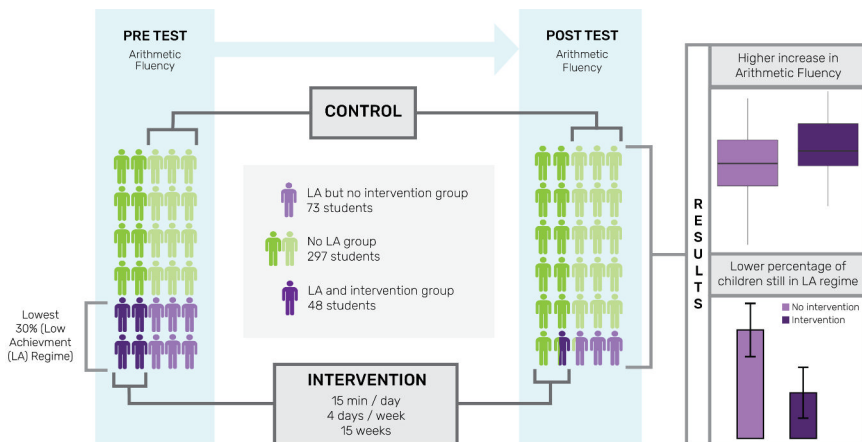
Fig. 1 An overview of the study's methodology and key findings. We administered an arithmetic fluency test to 418 grade 1 students in 13 schools in Catalonia, Spain. From these schools, 5 were in the intervention group and 8 in the control. We selected the 48 students in the overall lowest 30% percentile (Low Achievement (LA) regime). These students then participated in an intervention that consisted of 15 min of extra mathematics practice per day, four days per week, for 15 weeks. Afterward, we administered the same test again to all the students. We found that the children who underwent the intervention showed statistically higher arithmetic fluency. Additionally, after the intervention, a statistically smaller percentage of them remained in the Low Achievement (LA) regime compared to those in the control group. Note: if some normally achieving children fell into the LA regime in the post-test, they are still painted in green in order to make the abstract message clearer

schools included 6 public schools, with 4 having low SES and 2 with average SES. The remaining 2 control schools were charter schools, with one having a high SES and the other having an average SES.

All 13 schools were approached at the beginning of the academic year, and they were offered to participate in the study. They themselves chose whether to be included in the control or experimental group. Schools signed a data protection agreement with Innovamat Education SL (the curriculum provider), as well as with the families of the children participating in the study.

Tools

Tier I

Regarding the multi-tiered instruction required by RTI, the first Tier consists of the research-based mathematics curriculum proposal developed by Innovamat, which is currently implemented in more than 1,700 schools across 7 countries. The main teaching ideas behind Innovamat curriculum can be summarized as follows:

- Contents and processes: A global tendency in mathematical teaching practices is to opt for a dual approach which includes both content and processes (OECD 2018). This core curriculum achieves it by providing the teacher with a very detailed guide for every session, in which every activity is educationally justified. Each activity contributes to one or more of the mathematical processes as described by NCTM Principles and Standards (2008): problem-solving, reasoning and proof, connections, communication and representation. This dual focus is critical for performance outcomes, as it builds both conceptual understanding and procedural fluency (Hiebert and Grouws 2007).
- Fostering a problem-solving environment: Research shows that problem-solving fosters mathematical understanding and improves performance outcomes (Schonfeld 2016). Innovamat encourages students to engage in open-ended questions and peer discussions, fostering an environment conducive to research, analysis of mistakes, comparison of strategies, and reasoning. This aligns with Vygotsky's Social Constructivism, suggesting that peer interactions can significantly impact learning (Vygotsky 1978).
- From manipulation to abstraction: based on the ideas by V. M. D. Heuvel-Panhuizen (2008), a scaffolding mechanism where learning trajectories are developed from concrete materials to reduce abstraction is implemented. In next steps, abstraction is sought as materials are retired from the game until children, through representation, have understood the mathematical concepts behind them. This scaffolding technique has been shown to facilitate understanding and improve performance in mathematical tasks (Gersten et al. 2009).
- Practicing in meaningful contexts: two types of practicing are included, productive and reproductive. Productive practice is carried out by the whole class after an open-ended question has been introduced. Research indicates that context-based, meaningful practices improve retention and application of mathematical skills (Boaler 1993). Reproductive practice is carried out through a digital

application, which helps children practice mathematical concepts and procedures in a gamified setup that self adapts to every student's path, and provides continuous formative feedback. Numerous studies provide evidence that adaptive learning technologies can enhance performance outcomes (Kulik 2003; Shute 2008; Koedinger et al. 2010).

Tier II

In this study, a second Tier of support was designed specifically for low-achieving children. The Tier II support was facilitated using a digital app for 15–20 min, four times per week. While many activities in Tier II are analogous to those in Tier I, a distinct and targeted activity was integrated for children demonstrating low achievement in mathematics. This activity is rooted in the insights of Butterworth and Laurillard (2016).

Special activity for Tier II

At the heart of our Tier II intervention is a game-based activity inspired by Butterworth and Laurillard's (2016) work (Fig. 2). The activity deploys a series of beads that children can assemble or disassemble to produce sets with varied item counts. A student's task involves utilizing these digital beads to replicate a specific target number presented to them. Each unique bead count is color-coded, aiding students in associating a particular hue with a number. As students become accustomed to the game, the Arabic numeral equivalent of the bead count is incorporated into the design, fostering an association between numerosity and its symbolic representation. Once students exhibit mastery over this association, the color cues are withdrawn, ensuring that their numerical understanding is not color-dependent. The cardinality of these target numbers is progressively amplified, starting from 2 and culminating at 10.

Children spent around 5 min a day practicing this activity, 5 more minutes with Kindergarten level activities, and 5 more minutes with grade 1 activities. These were drawn from the usual tier I practice activities, and their details are elaborated in Supplementary Material I.

Fig. 2 Screenshot of the activity to build the concept of number as a set. In this activity, children can split or join strings of beads to match the objective set, above
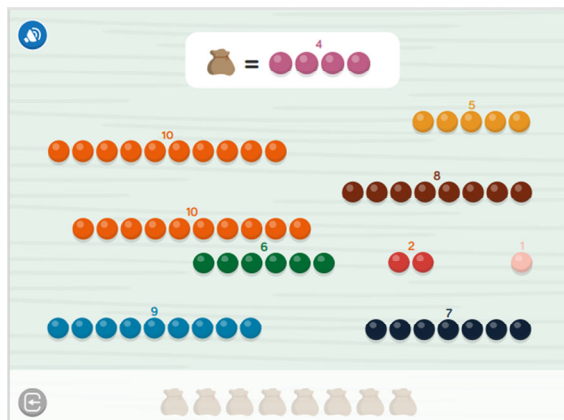
Screening

The screening materials have also been developed in the framework of this study and comprise a set of 8 tasks: motor speed, whereby the child has to press a stimulus appearing in the screen; visual processing speed, similar to WISC's symbol search (Wechsler 2014); number knowledge, where children have to select the bigger of two numbers presented simultaneously; an Approximate Number Sense task based on Halberda et al. (2008); a numerical line of range 0–100; a Working Memory task, where children need to remember some images while being distracted by other images; a reasoning task based on Raven's progressive matrices (Raven and Raven 2003) and a normed arithmetic fluency test, where children need to solve additions and subtractions in the 0–20 range as fast as possible. A detailed explanation of the tasks is provided in Table 1. For more information about their validity and reliability measures, please see Supplementary Material II.

The test is administered collectively, and each task is preceded by up to 5 examples that the whole class completes together. The test administrator makes sure that all children have understood the task before instructing the entire class to perform it by themselves, individually. The total duration of the test is around 45 min.

Arithmetic fluency (reflecting better addition and subtraction strategies as well as automation of basic arithmetic facts) is pivotal in the Curriculum Based Measurement in grade 1 (Fuchs and Fuchs 2006; Hosp et al. 2016) and is used in this study as a single-skill probe of mathematical achievement. The rest of the tasks were given to the students for two reasons: first, we wanted to give teachers a broader cognitive profile of each student and second, we wanted to analyze the effect of the intervention on cognitive areas not directly related to numerosity.
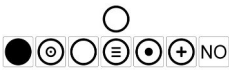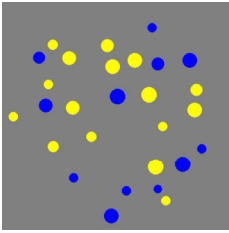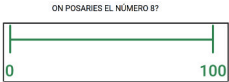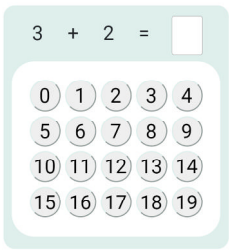
For the purpose of this study, the screening materials have been developed on a separate platform with different technology and a different user experience than the used by the students when learning with Innovamat (e.g., keyboard appearance). This was done in order to minimize the effect of children in the experimental group learning the tool.
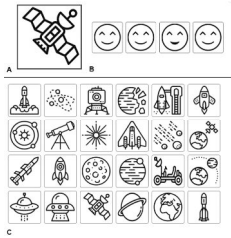
Procedure

At the start of the 2021–2022 school year, participating schools were informed of the details of the study. They were then given a 2-h individual session on the topic of learning difficulties in math and the RTI framework. They were then asked to identify a suitable time and location to carry out the intervention within the school and outside the normal mathematics classes. The intervention took place in the school because we couldn't ensure that the families would carry it out correctly, either due to lack of resources or willingness to participate from home.

In January 2022, all children were screened by members of the research team, and the statistical analysis of the results was conducted during the final week of the month. Based on these results, 30% of the children with the lowest performance on the arithmetic fluency task were selected for the intervention group. Schools were given two weeks to become familiar with the intervention materials and to report any issues or questions. In the third week of February, they were instructed to begin

Table 1 Explanation of the tasks involved in the pre- and post-screening tests

Task	Screenshot	Description
Execution speed		Measure the speed of the child when manipulating the device in which she will take the test. This information is used when analyzing timed tasks. The tasks consist of a four by five grids populated by a mole that changes position when clicked. The task consists of clicking on the mole as fast as possible. This task is timed and lasts for 30 seconds
Visual processing speed		Measures the ability to visually scan a series of objects and make decisions based on simple visual stimuli. The tasks consist in checking whether the upper symbol is repeated below, If it is, the child needs to click it, otherwise, she clicks "NO". This task is timed and lasts for 120 seconds
Number knowledge		Measures the degree of automation of the understanding of Arabic numerals. The child needs to click on the larger number as fast as possible. This task is timed and lasts for 30 seconds
Comparison of magnitudes		Measures the degree of acuity of the nonverbal Approximate Number System. Children are shown a stimulus with blue and yellow dots that range from 4 to 16 points per color and with ratios of 1:2, 2:5, 2:3, 3:4 and 4:5. The stimulus lasts for half a second, after which they have to choose in what color there was the higher number of points. This task is not timed and consists of 30 items balanced across ratios, overall number and blue or yellow as the correct answer. Furthermore, half of the trials are balanced across total color surface
Number line	Note: the stimulus reads "Where would you place number 8?"	Knowledge of the empty number line with range 0-100. Specifically, it measures the degree of linearization of the mental representation of natural numbers up to one hundred. Children are given a written numerical and asked to place it approximately in an empty number line. This task is not timed and consists of 20 items, and answers within 10% of the real value are considered correct
Arithmetic fluency		Measures arithmetic fluency in additive operations in the range 0-19. By measuring the speed of solving operations, we aim to measure the suitability of the strategies used by students when doing additions and subtractions and the degree of automation of basic math concepts. The child needs to do as many operations as possible in 120 seconds. Note that we designed a numerical entry object that looks like a normal calculator but holds the numbers from 0 to 19 so that students don't need to enter two separate digits when the result is over 9

Table 1 (continued)

Task	Screenshot	Description
Working memory		Measures the ability to hold information in memory and manipulate it without interference in the presence of distractors. The child is shown a picture (Fig. A below) and asked to memorize it. Then, she has a task where she has to choose the odd one out among 4 drawings (Fig. B below). This is done 8 times with 8 different stimulus to remember and tasks to perform. Afterward, she is shown a grid with 24 images (Fig. C below) and has to pick as many as possible of the 8 pictures she has tried to memorize. This task is not timed
Logical reasoning		Measures the ability to reason with concepts, to infer logical patterns and deduce their continuity. Based on Raven's matrices, it consists of a matrix with a missing piece and the child has to pick among 6 options. There are 24 items and the task is not timed

the daily 15 min practice sessions that comprised the tier II, which lasted for 15 weeks. After this, the research team re-screened the 13 schools using the same test as in January during the last two weeks of June.

Once a month, schools were sent a report with information on the level of participation and of successful completion of the activities that were given.

Analysis plan

We computed a final score for each task by adding the correct results and subtracting the wrong results, multiplied by a coefficient found in Supplementary Table 1. The first variable of analysis was the difference between the pre- and post-tests arithmetic fluency. Since it was a non-normally distributed continuous variable, we used Mann-Whitney's U to test the hypothesis of equality of medians between the control and intervention groups.

The second variable of study was the percentage of students still in the low achievement or in the "at risk of mathematics difficulties" zones. These categorical data were presented as absolute numbers and percentages, and tested using Chi squared test with Yates continuity correction to compare the control and intervention groups.

All analyses were performed with R statistical software version 4.1.

Results

Intervention adherence

Regarding the participation in the intervention, the average sessions per student per week was 2.22 (with standard deviation of 0.45 sessions per week). This is considerably less than the 4 sessions that they were instructed to do. Particularly problematic is the number of sessions at the very beginning of the intervention and after Easter break, when schools had a hard time picking up the pace. Variability among schools was very

small, and thus the results were not affected by the amount of intervention. We show a distribution of the sessions per student per week in Supplementary Fig. 1.

Arithmetic fluency

In Table 2, we show the results of the number of corrected operations in the fluency task before and after the intervention period, along with the difference between the pre and the post scores. We divide children in three groups: "Medium-High fluency" (i.e., scoring at or over the 30th percentile on the pre-test), "Low fluency: Control" (i.e., scoring below the 30th percentile but belonging to a control schools and thus not receiving intervention) and "Low fluency: Intervention", where belonged the children under the 30th percentile receiving intervention.

In Table 2 and Fig. 3, we can see that the pre-test arithmetic fluency scores are lower for students who received or needed the intervention (by construction), and somewhat higher for those who received the intervention, though not statistically significant. On the other hand, the post-test arithmetic fluency for students who received the intervention is statistically higher than for those who required it but did not receive it (11.1 vs. 7.2 operations, with a p-value 0.001). This indicates that the extra practice resulted in higher fluency. The effect size, calculated using Cohen's d, is 0.62, indicating a moderate effect of the intervention on arithmetic fluency. Looking at the gains, students in the intervention group also improved significantly more than the control group (7.3 vs. 4.2 operations, with a p-value = 0.007), with an effect size of d = 0.54, indicating again a moderate effect.

Other dimensions

We present descriptive statistics for all tasks in the Pre- and Post-tests, as well as their differences, in Table 3 and Fig. 4. When comparing the two "Low Achievement" groups, we find statistically significant results in the comparison of the differences for the Numerical Line ($p = 0.03$) and Number Knowledge (p-value: 0.01) tasks, but not in any other one (note that Working Memory is close to significance at p-value: 0.07).

Table 2 Arithmetic fluency task results

	Medium-high fluency	Low fluency: control	Low fluency: intervention	Low fluency control vs. intervention p-value
	N = 297	N = 73	N = 48	
Pre-test	12.5 [9.8; 16.8]	2.8 [−0.4; 5.0]	4.6 [0.0; 5.6]	0.19
Post-test	15.0 [11.0; 20.0]	7.2 [1.4; 10.4]	11.1 [7.3; 14.0]	<0.001
Difference	2.0 [−0.6; 5.0]	4.2 [−0.8; 9.4]	7.3 [3.8; 13.3]	0.007

Note Results of the arithmetic fluency task in the pre- and post-tests, as well as their difference in number of corrected operations in 2 min for the three groups in the study. The "Medium-High fluency" group corresponds to students in the 13 schools that scored at or over the 30th percentile, the "Low fluency: Control" and "Low fluency: Intervention" groups are those in the 8 control and 5 intervention schools respectively that scored under the 30th percentile. Results are given in medians and interquartile ranges. p-values correspond to Mann-Whitney U tests for the comparison of the two groups that needed intervention. You can find the Pre- and Post-scores and their differences for the remaining tasks in Table 3 and Fig. 3
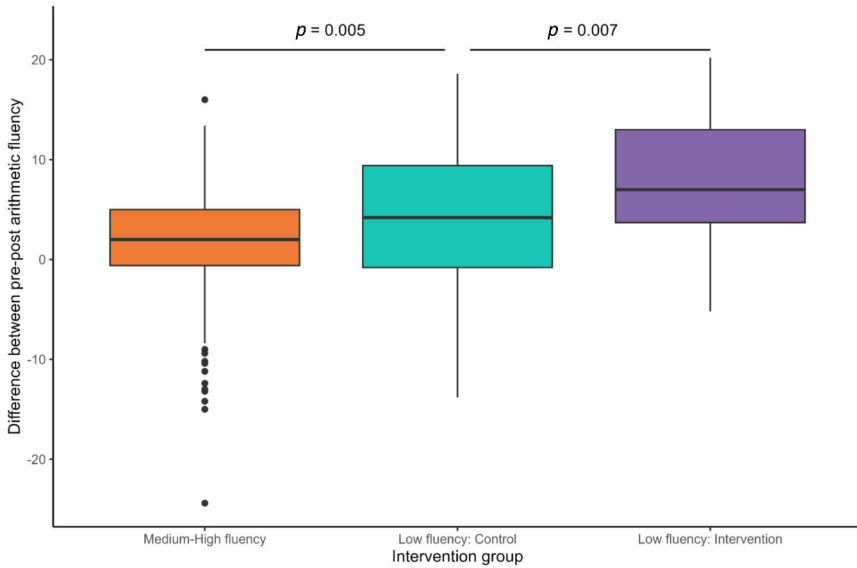
Fig. 3 Comparison of the difference between pre- and post-test arithmetic fluency, given in total corrected operations, for the three groups in the study. The boxes represent the interquartile range (IQR), stretching from Q1 to Q3, and contain the middle 50% of the data, for each study group. The two p-values shown on top represent the results of two Mann Whitney U tests between the study groups

Table 3 Medians and interquartile ranges for all the extra tasks given to students

Task	Low achievement: control			Low achievement: intervention		
	Pre	Post	Difference	Pre	Post	Difference
Execution speed	26 [20; 32.8]	32.4 [26.4; 37]	5 [0.4; 11]	26 [19; 31]	31.8 [26; 34.4]	5 [0.4; 10.2]
Visual processing speed	29.2 [22.6; 33.4]	33.8 [29.4; 38.6]	5.4 [1.6; 9.8]	27.8 [22; 32.6]	32 [27.8; 36.2]	4.2 [0.6; 9.4]
Number knowledge	2.7 [0; 10.7]	9.4 [4; 14.7]	2.7 [0; 8]	2.7 [0; 10.7]	8 [2.7; 13.3]	4 [−1.3; 6.7]
Magnitude comparison	10 [4; 16]	14 [8; 18]	2 [−2; 9.2]	8 [4; 14]	14 [6; 18]	2 [−4; 8]
Number line	2.7 [0; 10.7]	9.4 [4; 14.7]	2.7 [0; 8]	2.7 [0; 10.7]	8 [2.7; 13.3]	4 [−1.3; 6.7]
Working memory	5 [3; 6]	5 [3.5; 6.5]	0.5 [−1; 2]	4.5 [3; 6]	5 [4; 6]	0.5 [−1; 2]
Logical Reasoning	3.1 [1.8; 5]	4 [2.2; 6.1]	0.8 [−1.1; 2.4]	2.9 [1.7; 4.1]	3.3 [2.1; 5.4]	0.8 [−0.9; 2.4]

This result is yet another indicative that the intervention indeed succeeded in building the numerical abilities of children, and that other cognitive abilities were not mediating in the final improvement of arithmetic fluency.
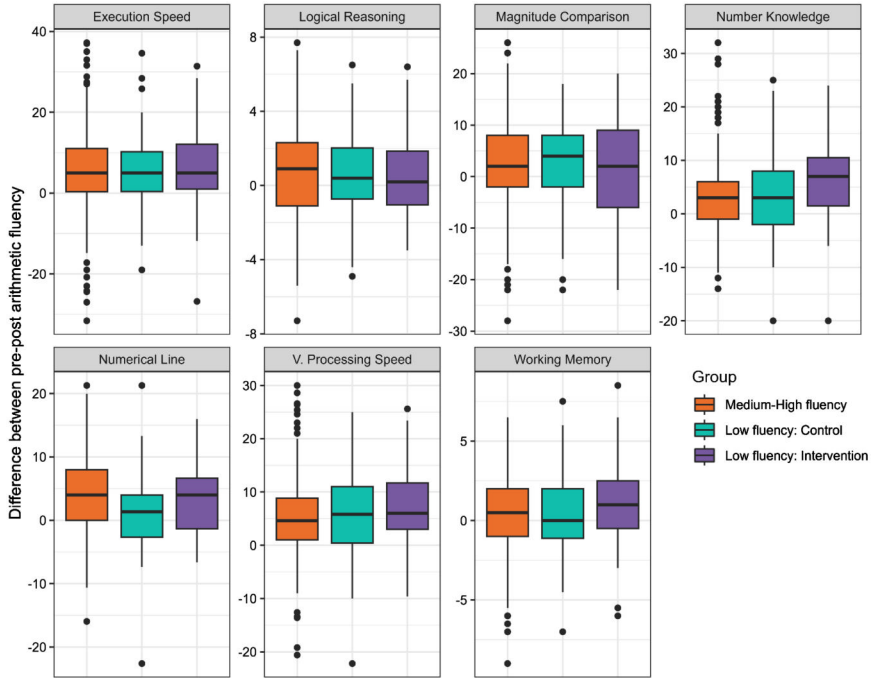
Fig. 4 Difference between pre- and post-scores for all extra tasks given to students

Low achievement and at risk of MLD zones

In table and Fig. 5, we show that the percentage of students who leave the "low achievement" zone (below the 30th percentile) and the "at risk of Mathematics Learning Difficulties" zone (below the 15th percentile) is significantly higher for students who received the intervention. Of the 48 students who received the intervention, only 9 remained in the "low achievement" zone afterward (19%), compared to 47% of students in the control group (p-value = 0.001, Fig. 5A). Similarly, only 4 students in the intervention group remained at risk of MLD afterward (8%), compared to 33% of students in the control group (p-value = 0.003, Fig. 5B). This indicates that the RTI intervention was effective in helping students leave the "low achievement" and "at risk of MLD" zones (Table 4).

Discussion

In this study, we evaluated whether it is feasible to develop a scalable semi-automated RTI framework in an educational system where RTI is not policy. For this purpose, we developed a screening test centered on arithmetic fluency that was administered to 418 first grade students in 13 schools. From these, we picked the
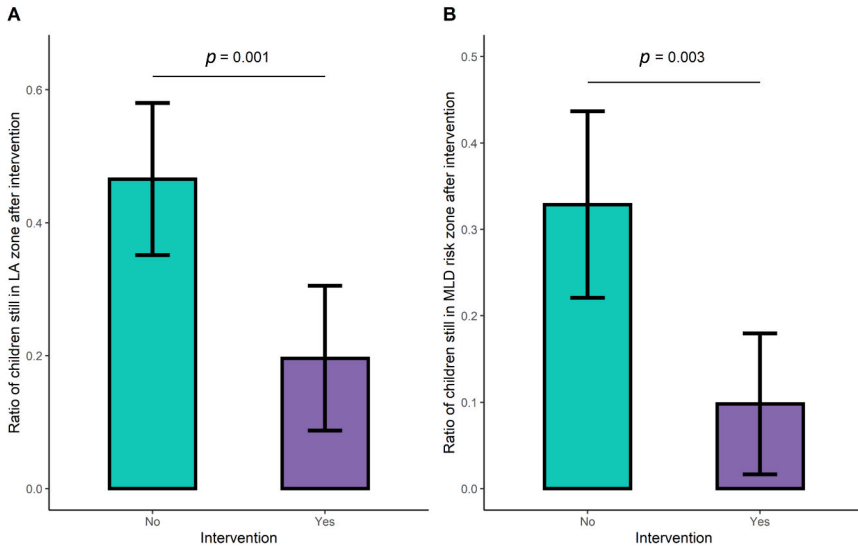
Fig. 5 Percentage of students who remain: **A** below the 30th percentile (low achievement regime); **B** below the 15th percentile (at risk of MLD regime), depending on whether they had intervention or not

Table 4 Absolute numbers and percentages of students who remain below the 30th percentile (low achievement regime) or 15th percentile (at risk of MLD regime) depending on whether they had intervention or not

| | | After intervention | | | |
| | | Still in low achievement regime | | Still in "at risk of MLD" regime | |
Low achievement group	N	No	Yes	No	Yes
Control	73	39 (53 %)	34 (47 %)	49 (67 %)	24 (33 %)
Intervention	48	39 (81%)	9 (19 %)	44 (92 %)	4 (8 %)

lowest performing (below the 30th percentile) of 5 of these schools, and gave them a digitized extra practice of 15 min per day, four days a week for 15 weeks.

Interpretation of results

The core finding of our study is the efficacy of an RTI framework even in environments where RTI is not policy. Students who were provided with the RTI intervention displayed a marked improvement in arithmetic fluency compared to their peers in the control group. Notably, the interventions did not directly target arithmetic fluency, but instead aimed at building foundational number concepts. This outcome thus emphasizes that a better understanding and automatization of numbers can substantially bolster arithmetic capabilities, an insight consistent with other studies (Östergren and Träff 2013; Malone et al. 2020).

While we observed that 8% of students did not show a significant response to the intervention, considering the general prevalence of dyscalculia, this outcome aligns with

existing literature (Shalev et al. 2000). This showcases the RTI framework's sensitivity in identifying potential cases of dyscalculia, an invaluable asset for early intervention.

The challenges faced during the study year, notably the pandemic-induced disruptions and teacher strikes, highlight the resilience of the framework. Achieving over half the recommended sessions in such a challenging environment is testament to the framework's adaptability and effectiveness.

Implications of the findings

Our findings hold significant implications for educational practices globally. Even in regions without a formal RTI policy, schools can successfully integrate a semi-automated RTI framework to improve student outcomes. This holds promise for democratizing quality education, providing students with structured, individualized support irrespective of systemic policies.

The effectiveness of our intervention also underscores the significance of foundational number concepts in enhancing arithmetic fluency. This suggests that educators might need to re-evaluate and potentially redesign curricula that prioritize rote learning over foundational understanding.

Limitations of the study

Several limitations should be considered while interpreting our findings. The most significant limitation was the inability to complete all recommended RTI sessions due to unforeseen challenges during the study year. This may have impacted the full potential of the intervention's effectiveness.

Another limitation pertains to the specificity of the educational context in which the study was conducted. Given that educational systems, pedagogical practices, and student demographics can vary widely across regions and countries, our findings are rooted in the unique attributes of the participating schools and their respective environments. Thus, while our results provide valuable insights for this particular setting, caution should be exercised when attempting to generalize the outcomes to other educational contexts or diverse student populations. Further studies in varied settings would be essential to ascertain the broader applicability and adaptability of our RTI framework.

Conclusions

In summary, our study presents a robust case for the potential scalability of RTI frameworks even in systems where it is not formally integrated as policy. Our results suggest a strong link between foundational number understanding and arithmetic fluency, challenging traditional pedagogical approaches. With continuous refinement and integration of feedback, such RTI frameworks could transform the educational landscape, providing individualized, targeted support to those who need it most.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical approval Our intervention was primarily an educational enhancement, closely aligned with standard academic curriculum and typical pedagogical practices. Our approach posed no potential harm or discomfort to students and can be seen as a regular instructional strategy rather than a clinical intervention. Therefore, we didn't seek external ethical approval.

Informed consent We ensured complete transparency and responsibility by obtaining informed consent from all participants' guardians. Parents and guardians, and schools were provided with a clear overview of the study's intent and procedures, allowing them to make a well-informed decision about their child's participation. Moreover, rigorous measures were adopted to maintain data privacy and confidentiality and families were given the option to not participate in the study or to ask for the data regarding their child being eliminated from our databases.

Conflict of interest All three authors work at Innovamat Education SL.

References

Benner GJ, Kutash K, Nelson JR, Fisher MB, Fisher MB (2013) Closing the achievement gap of youth with emotional and behavioral disorders through multi-tiered systems of support. Educ Treat Children 36(3):15–29

Berkeley S, Scanlon D, Bailey TR, Sutton JC, Sacco DM (2020) A snapshot of RTI implementation a decade later: new picture, same story. J Learn Disabil 53(5):332–342. https://doi.org/10.1177/0022219420915867

Boaler J (1993) Encouraging the transfer of 'school' mathematics to the 'real world' through the integration of process and content, context and culture. Educ Stud Math 25(4):341–373. https://doi.org/10.1007/BF01273906

Butterworth B (2003). Dyscalculia screener: highlighting pupils with specific learning difficulties in maths age 6–14 years. nferNelson Publishing Company Ltd, Glasgow

Butterworth B, Laurillard D (2016) Investigating dyscalculia: a science of learning perspective. In: From the laboratory to the classroom. Routledge, pp 184–202. https://doi.org/10.1126/science.1201536

Cho E, Compton DL, Fuchs D, Fuchs LS, Bouton B (2014) Examining the predictive validity of a dynamic assessment of decoding to forecast response tier 2 to intervention. J Learn Disabil 47(5):409–423. https://doi.org/10.1177/0022219412466703

Desoete A, Grégoire J (2006) Numerical competence in young children and in children with mathematics learning disabilities. Learn Individ Differ 16(4):351–367. https://doi.org/10.1016/j.lindif.2006.12.006

Eccles JS, Roeser RW (2011) Schools as developmental contexts during adolescence. J Res Adolescence 21(1):225–241. https://doi.org/10.1111/j.1532-7795.2010.00725.x

Fletcher JM, Vaughn S (2009) Response to intervention: preventing and remediating academic difficulties. Child Develop Perspect 3(1):30–37. https://doi.org/10.1111/j.1750-8606.2008.00072.x

Fuchs D, Fuchs LS (2006) Introduction to response to intervention: what, why, and how valid is it ? Read Res Q 41(1):93–99. https://doi.org/10.1598/RRQ.41.1.4

Geary DC (2011) Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. Dev Psychol 47(6):1539–1552. https://doi.org/10.1037/a0025510

Geary DC, Hoard MK, Byrd-Craven J, Nugent L, Numtee C (2007) Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. Child Dev 78(4):1343–1359. https://doi.org/10.1111/j.1467-8624.2007.01069.x

Geary DC, Nicholas A, Li Y, Sun J (2017) Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: an eight-year longitudinal study. J Educ Psychol 109(5):680–693. https://doi.org/10.1037/edu0000159

Gersten R, Chard DJ, Jayanthi M, Baker SK, Morphy P, Flojo J (2009) Mathematics instruction for students with learning disabilities: a meta-analysis of instructional components. Rev Educ Res 79(3):1202–1242. https://doi.org/10.3102/0034654309334431

Gersten R, Haymond K, Newman-Gonchar R, Dimino J, Jayanthi M (2020) Meta-analysis of the impact of reading interventions for students in the primary grades. J Res Educ Eff 13:1–27. https://doi.org/10.1080/19345747.2019.1689591

Gersten R, Jayanthi M, Dimino J (2017) Too much, too soon? Unanswered questions from national response to intervention evaluation. Except Child 83(3):244–254. https://doi.org/10.1177/0014402917692847

Greenwood C, Bradfield T, Kaminski R, Linas M, Carta J, Nylander D (2011) The response to intervention (RTI) approach in early childhood. Focus Except Child 43. https://doi.org/10.17161/fec.v43i9.6912

Halberda J, Mazzocco MM, Feigenson L (2008) Individual differences in non-verbal number acuity correlate with maths achievement. Nature 455(7213):665–668. https://doi.org/10.1038/nature07246

Heuvel-Panhuizen VMD (2008) Children learn mathematics (Dutch design in mathematics education). Sense Publishers, Rotterdam

Hiebert J, Grouws D (2007) The effect of classroom mathematics teaching on students' learning. In: Second handbook of research on mathematics teaching and learning. Information Age Publishing, Charlotte, pp 371–404

Horner RH, Sugai G, Fixsen DL (2017) Implementing effective educational practices at scales of social importance. Clin Child Fam Psychol Rev 20(1):25–35. https://doi.org/10.1007/s10567-017-0224-7

Hosp MK, Hosp JL, Howell KW (2016) The ABCs of CBM: a practical guide to curriculum-based measurement. Guilford Publications, New York

Hughes CA, Dexter DD (2011) Response to intervention: a research-based summary. Theory Pract 50(1):4–11. https://doi.org/10.1080/00405841.2011.534909

Jahnukainen M, Itkonen T (2016) Tiered intervention: history and trends in Finland and the United States. Eur J Spec Needs Educ 31(1):140–150. https://doi.org/10.1080/08856257.2015.1108042

Jimerson SR, Burns MK, Van Der Heyden AM (eds) (2016) Handbook of response to intervention: the science and practice of multi-tiered systems of support. Springer, New York

Jordan NC, Hanich LB (2003) Characteristics of children with moderate mathematics deficiencies: a longitudinal perspective. Learn Disabil Res Pract 18(4):213–221. https://doi.org/10.1111/1540-5826.00076

Jordan et al (2009) Early math matters: kindergarten number competence and later mathematics outcomes. Dev Psychol 45(3):850–867. https://doi.org/10.1037/a0014939

Koedinger KR, McLaughlin EA, Heffernan NT (2010) A quasi-experimental evaluation of an on-line formative assessment and tutoring system. J Educ Comput Res 43(4):489–510. https://doi.org/10.2190/EC.43.4.d

Kulik JA (2003) Effects of using instructional technology in elementary and secondary schools: what controlled evaluation studies say. Sri International, Arlington, pp 140–145

Malone SA, Burgoyne K, Hulme C (2020) Number knowledge and the approximate number system are two critical foundations for early arithmetic development. J Educ Psychol 112(6):1167–1182. https://doi.org/10.1037/edu0000426

McIntosh K, Horner RH, Sugai G (2009) Sustainability of systems-level evidence-based practices in schools: current knowledge and future directions. In: Sailor W, Dunlap G, Sugai G, Horner R (eds) Handbook of positive behavior support. Springer US, pp 327–352. https://doi.org/10.1007/978-0-387-09632-2_14

National Council of Teachers of Mathematics (2008) Principles and standards for school mathematics

OECD (2018) PISA for development assessment and analytical framework: reading, mathematics and science. OECD Publishing, Paris

Östergren R, Träff U (2013) Early number knowledge and cognitive ability affect early arithmetic ability. J Exp Child Psychol 115(3):405–421. https://doi.org/10.1016/j.jecp.2013.03.007

Raven J, Raven J (2003) Raven progressive matrices. In: McCallum RS (ed) Handbook of nonverbal assessment. Kluwer Academic/Plenum Publishers, pp 223–237. https://doi.org/10.1007/978-1-4615-0153-4_11

Reigosa-Crespo V, Valdés-Sosa M, Butterworth B, Estévez N, Rodríguez M, Santos E, Torres P, Suárez R, Lage A (2012) Basic numerical capacities and prevalence of developmental dyscalculia: the Havana survey. Dev Psychol 48(1):123–135. https://doi.org/10.1037/a0025356

Schonfeld A (2016) Learning to think mathematically: problem solving metacognition and sense making in mathematics (reprint). J Educ 196(2):1–38. https://doi.org/10.1177/002205741619600202

Serdyukov P (2017) Innovation in education: what works, what doesn't, and what to do about it? J Res Innovative Teach Learn. https://doi.org/10.1108/JRIT-10-2016-0007

Shalev RS, Auerbach J, Manor OHAD, Gross-Tsur V (2000) Developmental dyscalculia: prevalence and prognosis. Eur Child Adolesc Psychiatry 9(2):S58–S64. https://doi.org/10.1007/s007870070009

Shute VJ (2008) Focus on formative feedback. Rev Educ Res 78(1):153–189. https://doi.org/10.3102/0034654307313795

Sugai G, Horner R (2003) The evolution of discipline practices: school-wide positive behavior supports. In: Luiselli J, Diament C (eds) Behavior psychology in the schools. Routledge, New York, pp 51–66

Torgesen JK (2009) The response to intervention instructional model: some outcomes from a large-scale implementation in reading first schools. Child Dev Perspect 3(1):38–40. https://doi.org/10.1111/j.1750-8606.2009.00073.x

Van Der Heyden AM, Witt JC, Gilbertson D (2007) A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. J Sch Psychol 45(2):225–256. https://doi.org/10.1016/j.jsp.2006.11.004

Vasilyeva M, Laski EV, Shen C (2015) Computational fluency and strategy choice predict individual and cross-national differences in complex arithmetic. Dev Psychol 51(10):1489–1500. https://doi.org/10.1037/dev0000045

Vygotsky LS (1978) Mind in society: development of higher psychological processes. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wechsler D (2014) WISC-V: technical and interpretive manual. Pearson, Bloomington, MN

Wilson AJ, Revkin SK, Cohen D, Cohen L, Dehaene S (2006) An open trial assessment of "the number race", an adaptive computer game for remediation of dyscalculia. Behav Brain Funct 2(1):1–16. https://doi.org/10.1186/1744-901-2-20