The Journal of Experimental Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/vjxe20

Can a Tablet Game That Boosts Kindergarten Phonics Advance 1st Grade Reading?

Cassandra Potier Watkins & Stanislas Dehaene

To cite this article: Cassandra Potier Watkins & Stanislas Dehaene (2023): Can a Tablet Game That Boosts Kindergarten Phonics Advance 1st Grade Reading?, The Journal of Experimental Education, DOI: 10.1080/00220973.2023.2173129

To link to this article: https://doi.org/10.1080/00220973.2023.2173129

Published online: 31 Jan 2023.

THE JOURNAL OF EXPERIMENTAL EDUCATION
https://doi.org/10.1080/00220973.2023.2173129

Can a Tablet Game That Boosts Kindergarten Phonics Advance 1st Grade Reading?

Cassandra Potier Watkins[a,b] and Stanislas Dehaene[a,b]

[a]Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Saclay, Gif/Yvette, France; [b]Collège de France Université Paris-Sciences-Lettres (PSL), Paris, France

ABSTRACT
We report the effectiveness of using the tablet-game Kalulu Phonics immediately after intervention in kindergarten and on national evaluations the year after. In a previous intervention testing the software with 1st graders, fluency and comprehension were boosted, but only when used in concert with reading instruction at the start of the year. Here, we asked whether a similar intervention would be more efficient if it started a year earlier, in kindergarten. Forty classes (1092 = children) were randomized into playing Kalulu Phonics or the matched Kalulu Numbers control game for the first half of the year, with reversed assignments in the second half. Ten non-randomized business-as-usual classes also participated. In a cross-over effect, children who used the phonics version improved in letter naming, grapheme-phoneme correspondence (GPC) matching and fluency. Students with the number version improved in number knowledge. In a longitudinal follow-up, all intervention participants maintained an advantage in phoneme awareness and GPC matching at the start of 1st grade, but this advantage failed to translate into school literacy gains by mid-year. No longitudinal benefits were found for numbers. Our results support using tablet-based aids in spurring early reading skills but question the possibility that a short-term intervention may address the challenges of long-term educational goals.

KEYWORDS
Game software; kindergarten; longitudinal; phonics; reading

As one of the few examples of successful applications of cognitive science to classrooms, phonics instruction has emerged from the "reading wars" as the clear winner in producing the largest literacy gains for all children (Castles et al., 2018; Torgerson et al., 2006). This finding is supported by randomized control trials and longitudinal follow-up assessments (for national reports including meta-analyses, see National Reading Panel, 2000; Rose, 2006) and appears to be equally true for students at risk for reading failure (Galuschka et al., 2014). It converges with functional brain imaging studies highlighting a standard reading circuit whereby both grapheme and phoneme processing nodes are established early in reading acquisition, and are particularly boosted by phonics instruction (Brem et al., 2010; Dehaene et al., 2015; Maurer et al., 2010; Turkeltaub et al., 2003). Studies comparing children before and after they learn to read show that visual recognition

CONTACT Cassandra Potier Watkins cassandra.potier-watkins@college-de-france.fr Collège de France Université Paris-Sciences-Lettres (PSL), 11 Place Marcelin Berthelot, Paris, 75005, France.
At the project website (in French) materials and the game used in this research can be downloaded for use. The game researched in this project was developed using the open-source platform, Godot (https://godotengine.org). The game code is freely available on the authors GitLab: https://gitlab.com/casspw/kalulueducation, under a Creative Commons Attribution 4.0 International Public License. Anonymized data and scripts are available upon request (we are required to let the department of statistics of the ministry of education know if data is shared).

activation in the left occipito-temporal pathway becomes sensitive to letter strings and develops increasingly efficient connections to regions specialized in processing speech sounds in the planum temporale (Dehaene et al., 2010; Dehaene-Lambertz et al., 2018; Monzalvo & Dehaene-Lambertz, 2013).

Today, the results of this collaboration between science and education have come full circle, as phonics becomes more and more prevalent in literacy curriculum. Phonics mastery has become central to standardized testing in several different countries (e.g., the Common Core evaluations in the United States; the Phonics Check in the United Kingdom; the ÉvalAide program in France). The question now is not so much why this method works but enhancing best practices for success. Several tenets have emerged: grapheme-phoneme correspondences (GPC) should be explicitly taught (Byrne, 1992; National Reading Panel, 2000) and instruction should follow a systematic progression, in which GPCs are introduced in a rational order taking into account their frequency and consistency in a given orthography (de Graaff et al., 2009; Ehri et al., 2001). Given the ubiquitous and systematic requirements of phonics instruction, an increasing number of education apps have been developed to tackle early literacy (Abrami et al., 2020; Archer et al., 2014; Blok et al., 2002; Galuschka et al., 2014; Grant et al., 2012; Savage et al., 2013; Torgerson & Elbourne, 2002).

Numeric responses to literacy curriculum

In spite of interest in using education technology to boost school learning, there is little evidence that its increased use translates into greater educational gains (Dynarski et al., 2007; OECD, 2015; Torgerson & Elbourne, 2002). Critically, similar results have also been found for commercial based reading software (Dynarski et al., 2007; Grant et al., 2012). It is for this reason that software should be developed using principles from the cognitive sciences (Hirsh-Pasek et al., 2015) and its efficacy verified through randomized control trials (RCT). Under these criteria, literacy appears to be one avenue where technology can be of help as several examples of evidence-based applications for early literacy have emerged. For example, the active 'drill-and-practice' of GPCs in the GraphoGame (graphogame.com) boosts phonics knowledge, particularly with children at risk for dyslexia (Richardson & Lyytinen, 2014). The ABRA software (literacy.concordia.ca), based on recommendations from the National Reading Program (National Reading Panel, 2000), has been successful in yielding positive effects in vocabulary, phonemic awareness, fluency and comprehension (for a meta-analyses on size effects of the ABRACADABRA software, Abrami et al., 2020).

In France, where the current research takes place, in a randomized control study comparing the French phonics software ELAN to a math control game, we demonstrated that our phonics software was successful in supporting fluency and reading comprehension skills in 1st grade (Potier Watkins et al., 2020). Crucially in this study, gains were only achieved by children that participated in the reading intervention during the first half of the year, coinciding with the time point when children were just starting formal literacy training (1st grade is the first year reading is taught in France). Using the phonics application during the second half of the year, as a review or as an aid to struggling readers, failed to produce significant improvements. The phonics software taught children letter-sound relationships in a programmed order, without taking shortcuts to adapt to the student's level. Undoubtedly, students from the 2nd-half of the year did move through the levels quicker, and the game did adapt to their skills by requiring faster answers, with the goal of building automatization, but the experience was still mostly one of review, since the child would have had to pass through all levels of GPC learning. Additionally, children that used the reading intervention at the beginning of the year continued to make significantly more correct responses at the end of the year, maintaining an advantage of 0.34 SD in fluency and 0.31 SD in comprehension over the control software group.

The results from this study pointed to a possible period when phonics instruction is the most advantageous. Improved reading comprehension from the early intense phonics instruction supports the hypothesis that the automatization of decoding skills can free up cognitive resources for comprehension. It fits with the Simple View of Reading, which states that written comprehension is the product of decoding and spoken language comprehension (Hoover & Gough, 1990; Kendeou et al., 2009). But perhaps even more importantly for the goal of better understanding 'what works', our results with ELAN phonics highlight a very specific and early window for the potential of phonics education as setting the foundation for literacy. After this period, once the student has developed a tentative reading strategy, further GPC training does not appear to be affective in improving reading fluency and comprehension. In other words, phonics instruction needs to be operative and early in reading education. This idea also finds support from other recent reports that many GPCs taught early in learning provide the best literacy gains (i.e., learning up to two GPC a week, compared to the frequent practice of one a week, Goigoux, 2017; Sunde et al., 2020). The National Reading Panel meta-analysis (National Reading Panel, 2000) reported a larger effect size when phonics instruction began early ($d = 0.55$), compared to after 1st grade ($d = 0.27$).

The current study: research aims in Kalulu phonics

Encouraged by the previous results of our phonics software and support for the importance of phonics instruction early in education, we decided to retest the software a year earlier with kindergarten children. Based on teacher recommendations and pilot sessions with a kindergarten class, we made several principled adjustments and rebranded the ELAN game under the name "Kalulu". The main adjustment made to the game was to remove sentence and word reading comprehension tasks that pilot studies had demonstrated as too difficult for kindergarten classes who were not receiving formal reading instruction. Kalulu Phonics retained the games from ELAN and reinforced learning of French GPC through syllable and short word games to master each lesson. As a side note, we would like to mention that Kalulu was developed using the open-source language Godot (https://godotengine.org). Code for the game as it was developed for this project can be accessed on our GitLab repository (https://gitlab.com/casspw/kalulueducation). Our goal here was and is to promote high quality games that can be easily adapted to any language and used by teachers or researchers.

The first goal of our Kalulu Phonics software study was to continue to explore the generality of the game to help children learn GPCs, while reducing several confounding factors from the first study. For example, by providing the intervention to kindergarten classes we hoped to measure the possible benefits of the software while reducing possible confounds with teaching method. A common limitation to the ELAN study was one common to educational research: the intervention provided an extra experience to literacy education in the class, while the latter remained uncontrolled (Abrami et al., 2020). The ELAN suite was designed to complement many of the most common manuals used in reading education, but it could have been more complementary to certain methods than others. Our logic was that an intervention in kindergarten would provide a better controlled, standardized amount of phonics teaching, well adapted to the children's current instruction level, and would determine if such an approach can yield long term benefits.

In this goal, another reason for testing Kalulu Phonics with kindergarten students was to assess if an early window in phonics instruction would provide greater benefits to children, independent of the period in the school year that it was used. In French kindergarten, children are expected to apprehend that there are predictable relationships between sounds and letters, but "there is no pre-reading in kindergarten" (*Programme d'enseignement de l'école maternelle*, 2015). Our first concern was that the absence of reading instruction may be because children are considered developmentally too young. This concern was assuaged by previous research showing benefits of pre-1st grade phonics programs, for example, children can learn letter-sound relationships as early as preschool (Wolf, 2016); mentoring kindergarten teachers in phonics teaching techniques improves reading skills at the end of kindergarten and in 1st grade (Ehri & Flugman, 2018); early phonics and small group reading interventions provide the biggest effect sizes for kindergarteners at-risk for reading difficulties (for a meta-analysis, see (Cavanaugh et al., 2004). Furthermore, several computer-based programs for kindergarten literacy, that take into account evidence based pedagogy, have had positive outcomes (Comaskey et al., 2009; de Graaff et al., 2009; Macaruso & Walker, 2008; Richardson & Lyytinen, 2014; Savage et al., 2013). Taken together, testing our phonics software seemed relevant and not completely incongruent with the current kindergarten learning goals, and any positive impact of the software on reading would occur in the absence of formal reading instruction in the class.

The end goal of phonics training is improved reading fluency and comprehension, our second goal was to ascertain if children that used Kalulu Phonics in kindergarten would get a boost in literacy gains once formal reading education began the year after (1st grade, 6–7 y.o.). To do this, we collected the results of our population on National Evaluations given at the start and middle of the 1st grade year.

Kalulu phonics (Figure 1) teaches children the GPCs in an order that takes into account their frequency and consistency in written language. This progression was determined using an automatized neural-network developed to automatically detect the GPCs then calculate their frequency and consistency in any language (Potier Watkins et al., 2019). We used the Manulex children's print database to train our model (Lété et al., 2004). The lesson progression is downloadable on the project website (https://ludoeducation.fr). Several adjustments were made to the ordering of the lessons in concordance with other existing French phonics manuals familiar to teachers (e.g., teaching all the primary vowel sounds in the beginning, introducing fricative consonants, before occlusive ones). Each GPC lesson was introduced in three steps commonly used by teachers or supported by research as benefitting phoneme processing and grapheme memorization (Figure 1, a). In step 1, clicking on the grapheme starts a video of an older child pronouncing the corresponding phoneme, with a zoom on the child's mouth and explicit articulation to focus attention on the articulatory gestures underlying each phoneme (Boyer & Ehri, 2011; Castiglioni-Spalten & Ehri, 2003). In step 2, clicking on the grapheme produces a voice emphasizing the phoneme sound and its presence at the onset of the name of a picture (e.g., 'aaa, aaaple') to facilitate memorization of the sound by linking it to a known word. In step 3, the child traces the grapheme in both upper and lower case on the touch screen. Haptic exercises with letters, such as learning to write or to trace letters with the finger, improve reading skills by providing additional motor code to support memory for grapheme shape (Bara et al., 2004; Bara & Gentaz, 2011).

After completing the lesson, three games are proposed to automatize recognition of the GPCs (Figure 1, b–d). The minigames in Kalulu all have the goal of spurring the child to automatize the recently acquired GPC relationship with gradual difficulty. The simplest games develop phonics awareness through syllables. Children hear the GPC in a syllable, either as a single vowel, consonant-vowel or vowel-consonant, and must click on the corresponding written stimulus on screen. Once children have learned enough GPCs to build short words, they practice their knowledge in reading and spelling games. Another game was specifically designed to develop knowledge of French morphology by highlighting the silent letters that mark gender and plurals in words. Difficulty is created by varying the distractors and the rate of targets and distractors. All syllables and words presented in Kalulu are 100% decodable from the lessons learned.

The game algorithm aims to provide appropriate distractors to each target. Only previously learned GPCs can act as distractors. Furthermore, only vowels are presented as vowel distractors and consonants as consonant distractors. Distractors in a minigame are chosen from a visual similarity matrix adapted from another study of similarity judgments of letter pairs (Boles & Clifford, 1989). Initial items are chosen (when possible) to be >0.75 in similarity to the target (1

Figure 1. Kalulu phonics pedagogy.
Screenshots from the game. a) Introduction to a grapheme-phoneme correspondence (GPC) by seeing a video pronunciation of the phoneme, hearing the sound in a word with accompanying picture, tracing the grapheme in upper- and lower-case forms. After learning the GPC, different minigames are proposed to automatize learning. b) Syllable games require hearing the new sound in a syllable and finding it amongst distractors on a screen. c) Word building games require hearing a word then choosing the correct graphemes to spell the word. d) Comprehension and morphology games require reading the word silently, then making a decision of which letters in the word are a silent feature of French or deciding if a presented word exists or was made-up by the game mascot.

being the matrix diagonal, and 0 being maximally different items). If the player makes a mistake in the minigame, then the corresponding threshold on distractor similarity is lowered. The game also builds an internal model of the child's confusion between targets and distractors, by keeping track of performance for each GPC. Across games, the player is required to maintain a > 75% score for all GPC. If the child makes too many mistakes across several games, and a GPC score dips below the 75% threshold, then the corresponding grapheme is reintroduced as a target. Along with adapting the presented stimuli, each of the minigames has several levels of difficulty that lead to changes in the number of distractors presented and response time requirements. These levels increase or decrease in difficulty for every two identical minigames won or lost in a row.

Experimental study

The goal of this study was to retest our Phonics software and explore two research questions. 1) Can the Kalulu Phonics software teach GPC to kindergarten students before they begin formal reading education? and 2) Can early phonics instruction boost literacy gains during formal reading instruction in 1st grade, the school year after the intervention? To ascertain the impact of the phonological goal, we designed a well-matched mathematics control game that used the same game environment as Kalulu phonics to test for potential benefits that could be attributed to the phonics software beyond the novelty of introducing tablets to the class. To assess the second question of possible long-term benefits, we followed-up on participants' national standardized tests, given the following school year from the intervention. Several important standards from the previous project were maintained: both intervention and active control groups were randomized, the application was piloted for conformity to kindergarten usability as a stand-alone game needing minimal assistance from an adult, students played on the tablet individually but in small group (6 - 8 children) workshop stations (Potier Watkins et al., 2020). Many of the reported reading software interventions require supervision from a trained adult (Comaskey et al., 2009; de Graaff et al., 2009). This research practice is unrealistic in the goal of large-scale software deployment. We used a cross-over design to compare the phonics software to the active control.

Methods

Participants

The project was approved by the ethical committee CERES (Conseil d'évaluation éthique pour les recherches en santé, project N° 2017 – 27). Two different school academies in France participated in the project (Nice and Poitiers). Each recruited twenty public kindergarten classes to be randomized in the tablet-based intervention from a mix of urban and rural zones and five classes for a tablet-free "business-as-usual" control group. At the start of the school year, 1,154 students were estimated in the project. Teachers sent home with students a letter explaining the project, accompanied with an 'opposition form' for guardians who did not want their child to participate. No students were excluded from the study by the research lab. Fourteen families opted out of the project and one child refused to test. Five students were excluded from the project by the teacher because they attended specialized learning instruction. Thirty-six students were reported as having moved away during the intervention year.

Design

The school districts required that all children from intervention classes eventually have access to both Phonics and Numbers versions of Kalulu. Classes were randomized to Kalulu Phonics for the first half of the year, then Kalulu Numbers for the second half (group labeled "read/math", N = 411 children), or vice versa (group labeled "math/read", N = 449 children). Randomization to one of the two intervention groups required an equal number of classes in each group and was stratified by the school district (Poitiers or Nice) and the size of the class, which varied from 5 to 28 children. Smaller class size indicates a class split into two grade levels (i.e., a class may have students from middle-preschool and kindergarten), a common practice in rural areas. Ten 'business-as-usual' control classrooms (group labeled "control"; N = 232 children) also participated. The inclusion of both an active and passive control was to measure improved phonics knowledge due to the pedagogy separate from potential tablet affects. The passive control could not be randomized, as these teachers had been recruited by the school district by their interest in the project, but reluctance to introduce tablet games to the class. All children were tested at the beginning of the year (labeled pretest). To measure the evolution of learning, children were tested during mid-year before switching games (labeled midtest) and at the end of the year (labeled endtest). This 2-period crossover design was similar the prior project ELAN, excepting a randomization of classes instead of students (Potier Watkins et al., 2020). To assess our second question of long-term benefits to using the software, children's results on national evaluations given at the beginning (September) and middle (end of January to beginning of February) of the 1st grade were also examined, respectively 4- and 9- months after intervention.

Procedure

Five to eight tablets (depending on class size) were sent to classes at the start of the school year. A lab member contacted each teacher individually to present the project and dedicated website with supporting information. Teachers were encouraged throughout the year to call or email the designated lab member with their implementation questions. This person was not aware of students' results on the tests. Within each class, a workspace dedicated to tablet sessions was set up in a quiet space. Teachers organized small groups of students according to the number of tablets received. Children took turns using the tablets. Group ability was heterogenous in terms of their expected ability by teachers. The goal was for children to help each other and minimize the need for teacher assistance. Teachers informed us that rotating small groups between activity stations is a common practice in kindergarten. Research has shown that small group learning is especially beneficial for young children (for a meta-analysis see, Hattie, 2010). Furthermore, small groups rotating at a tablet workstation is far more economical for large-scale implementation than one tablet per child.

Each intervention session lasted ten weeks. Teachers were asked to ensure that students play for one hour a week for at least eight weeks of each session (three 20-minute sessions or two 30-minute sessions per week). Each of the test periods lasted three weeks. Several small teams of pedagogical councilors from each school were assembled to test students. Three weeks were necessary for these teams to test the students, as this job was done on top of their work in the schools and in addition to working around class schedules. Once the pretests were completed, teachers were informed of their intervention group (read/math or math/read). They also received an instruction book (on paper and pdf) detailing the cognitive science behind the pedagogy for their immediate group (Phonics or Numbers). Teachers were encouraged to familiarize themselves with the materials, but under no obligation to modify their own pedagogy in consequence. Communications between the lab and the classes were maintained throughout the year. Schools that had WIFI sent data to the lab, while others kept the lab abreast of progress through phone calls. Everyone reported positive remarks and high engagement from the class. Only one child in the intervention refused to play half-way through the year. All of the classes remained in the project throughout the intervention year.

The intervention game: Kalulu phonics

Kalulu Phonics was adapted from the previously developed Phonics instruction application ELAN (Potier Watkins et al., 2020). We piloted game usability with a kindergarten class not participating in the intervention over three months.

A matched active control: Kalulu numbers

A criticism of our previous study in 1st grade was that the control software for enhancing number knowledge was not directly matched to the phonics games and did not produce any benefits on our control measures for early math ability. Participating teachers had reported that the children's interest in the control game dissipated quickly. In other words, we had a lingering doubt that the performance advantage provided by using the Phonics game was possibly due to its high engagement level over the math game, as gains were only made in reading tasks, but not number tasks when using the math game application. For the current project, our goal was to use a better matched control. To do this, we adapted the pedagogical goal, lessons and minigames from Kalulu Phonics to the Kalulu Numbers program. In so doing, we were able to reconfigure the lessons and games appreciated by students using the same visuals, rules and procedures to make both games comparably challenging and engaging (Figure 2). Both Kalulu Phonics and Numbers were designed to build associations between symbol and meaning. In the case of numbers, this meant learning to associate symbolic numbers (either number names, e.g. 'two', or indo-arabic numerals, e.g. '2′) to the corresponding quantity ('**'). Playing either Kalulu Phonics or Numbers can be restricted through a teacher panel.

Kalulu Numbers teaches numbers 0-100 in forty lessons. Each lesson starts with a 3-part sequence to associate number with meaning using techniques supported by the literature (Figure 2a). First the number is presented in its counting sequence on a 10×10 game board. Playing on game boards that demonstrate the linear structure of numbers has been linked with improved numeracy (Elofsson et al., 2016; Laski & Siegler, 2014; Ramani & Siegler, 2008). Next, the child is introduced to different quantity representations of the number, a common practice shown to improve math readiness in school (Skwarchuk et al., 2014). Finally, children are asked to trace the number symbol twice to help memorize its form. The hypothesis with letters is that haptic exercises like tracing or writing may provide an extra motor code and reduce instances of mirror-writing (Gottfried et al., 2003). As mirror writing also exists in numeral writing, we reasoned here that this practice could also help cement the memory for numerals (Arnas et al., 2004; Fischer, 2011).

Once the learning session is complete, as with the reading path, several minigames are proposed to cement comprehension (Figure 2, b–f). The first type of minigame focuses on automatizing fluency of association between number name, number symbol and quantity. Once children learn enough numbers, a counting game is introduced. The counting game requires helping a game character find the correct number in a counting series starting from either '0′ or another random number and ending at specific number (e.g., "starting from the number 5, count all the numbers up to the number 8"). Once children learn numbers above 10, a game is introduced to associate numerals to quantities, taking into account place value (e.g., if asked to find the number matching the quantity of '28′, first the child would match the symbolic number '20′, then the symbolic number '8′). About mid-way through the game, arithmetic for small additions (<10) is introduced. Passage from one garden to the next is tested by a 'choose the biggest number task', where children have to choose the biggest between two numbers, twenty pairs are presented in total. Children have three minutes to complete the task.

Measures

To assess kindergarten phonics and number knowledge, we used the tablet version of the French national evaluation given to all children at the beginning of their 1st grade year. The evaluation assesses expected knowledge and predictors for reading and math achievement (Groupe de travail du Csen, 2019). By using it during the intervention, we could measure progress through the kindergarten year and follow-up on progress on comparative measures when our participants took

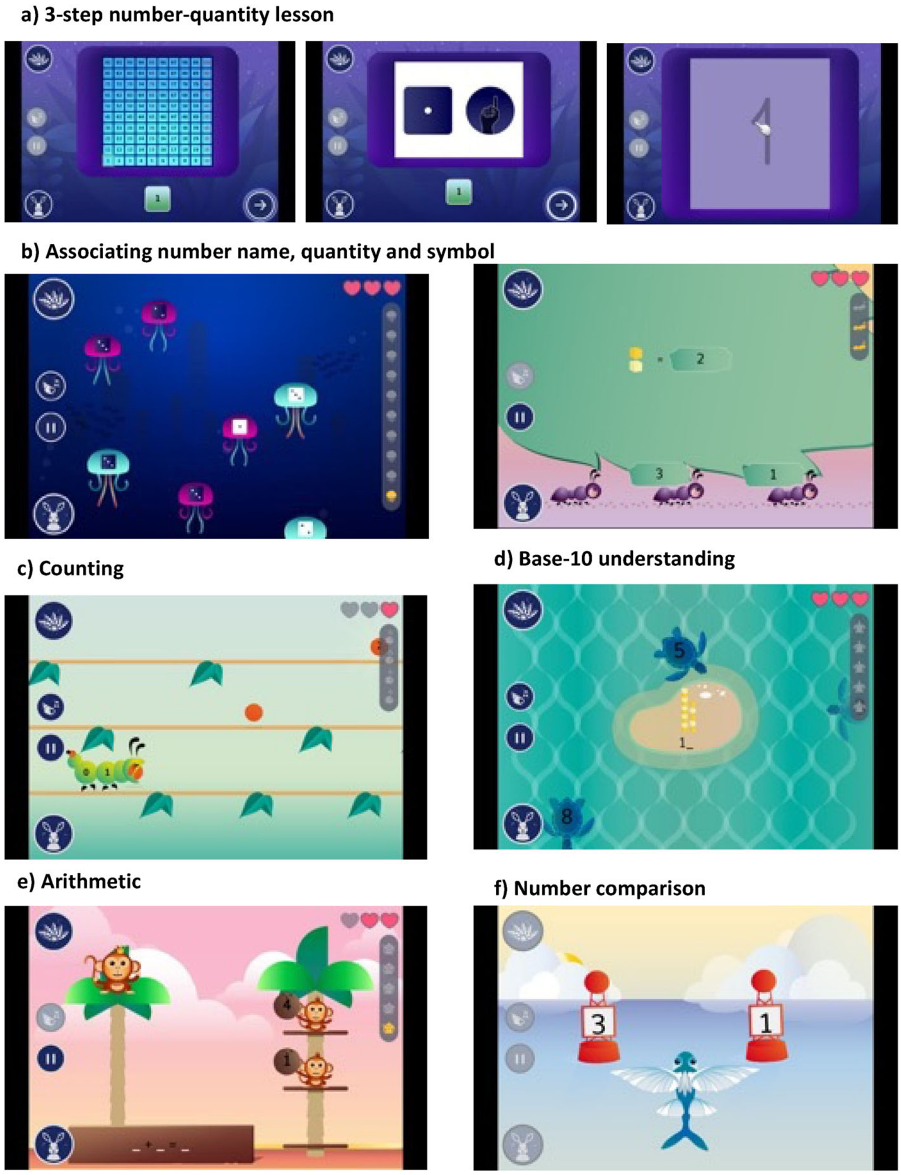

Figure 2. The matched control, Kalulu numbers pedagogy.
Screenshots from the game. a) Introduction to a number and its quantity by counting to the number on a game-board that highlights the linear order of the numbers, showing different quantity representations of the number and tracing the number symbol. Once the lesson is learned, children are introduced to different games to automatize number and meaning. b) Associating number name, quantity and symbol requires hearing or seeing a quantity then matching it with the correct symbol or quantity. c) Counting games require finding the correct number in a counting suite amongst distractors. d) Base-10 understanding games require associating quantity and symbol, while taking into account place value. e) Arithmetic games provide single digit addition problems. f) A number comparison task requires choosing the biggest number in a limited amount of time.

the national evaluations in 1st grade. Testing at all three periods during the kindergarten intervention was done on tablet by school employees specialized in pedagogical support. Testers attended an on-line training. Children did the test individually, wearing headphones and providing answers on their own on the touch screen. The testing administrator was present to repeat instructions and ensure fidelity. Data from the tests was subsequently sent to the lab.

Participant baseline composite predictor

As all children in France typically go to preschool (beginning 2-years before kindergarten), three tasks were done only at pretest to create a composite variable of pre-intervention knowledge. The first task tested visual attention (24 items). Children were presented with side-by-side letter suites of 3, 4 or 5 letters and asked to press on a button indicating if they were the same or different. Children were also assessed on vocabulary (15 items), adapted from the TVAP of expected French vocabulary for this age group (Deltour & Hupkens, 1980). Four pictures were shown to the child, after which they heard a word and were asked to choose the picture that best illustrated the word. Target words described objects (i.e., 'hammer) or actions (i.e., 'sew'). Distractor pictures had similar pronunciation or were categorically semantic. The last task assessed syllable awareness for beginning word sounds (10 items) and rhyme (5 items) (adapted from the EVALEC battery, Sprenger-Charolles et al., 2005). Children were presented with four named pictures. For the beginning syllable task, a word was heard by the child, who then chose the picture that started with the same syllable. In the rhyming task, the child chose which of the pictures had a word that did not rhyme with the other words. For each task, the number of correctly answered items was collected. A predictor variable was created from the average of the Z-scores for each individual on each task.

Assessment measures of early reading predictors and skills

Oral sentence comprehension. Children heard a short sentence then chose the correct corresponding image from a choice of four pictures including scenes with semantic and syntactic distractors. Ten items were presented and the number of correctly answered items was collected. This test was adapted from the standardized oral comprehension test (L'ECoSSe in French Lecocq, 1996; TROG-2 in English Bishop, 2003).

Phoneme awareness. Phoneme awareness was assessed for sensitivity to the first and last sound of a word. Children were presented with four named pictures, then heard another word and were asked to choose which picture started with the same sound (8 items) or ended with the same sound (7 items). Phoneme awareness is at the same time predictive of reading acquisition and improved by learning to read (Melby-Lervåg et al., 2012).

Letter name knowledge. In this task, the child was presented with three rows of five letters, told a letter name and asked to click on all the examples of the heard letter (7 questions, each with 3 examples of the target letter). The target letter could be written in majuscule or miniscule and cursive or print. Four consonants and three vowels were asked. Distractors were different random letters.

Grapheme-phoneme correspondence (GPC) matching. In this task, children were presented with five letters and heard a letter sound. The child had to choose the letter that matched the sound heard. Ten items were presented. Targets were all consonants.

Fluency. This test question was taken from the EVALEC (Sprenger-Charolles et al., 2005). This was the only test on paper done individually with a tester. Children were shown a list of words and asked to read as many as possible in one-minute. After, they were presented a list of matched pseudo-words and again asked to read as many possible in one minute. We measured the average number of words and pseudowords combined as a measure of fluency.

Control measures, number knowledge and math skills. On the numbers and math assessment only numbers up to ten were shown or required as answers, ten being the biggest number taught in French kindergarten. The goal of including these questions was to distinguish the intervention's specific phonics or number effect. An additional number comparison task was attempted, but the results were not included in our analyses due to a design issue that prohibited us from calculating the number of correct responses made in a minute time constraint, the typical measure of analysis for this task. The number of correctly answered items was collected and the percentage correct measured

Number knowledge. Children completed two different task to assess their number knowledge. In the first task, the child was asked to match an orally heard number name with its written form (20 items). All numbers from 1 to 10 were presented twice. In the second task, the child had to choose the correct number (from 1 to 9) that matched a quantity, presented as eggs in a basket (8 items).

Word problem solving. Math problem solving was assessed using simple everyday word problems (i.e., "Five kids need a pen. There are three pens in the box. How many more pens do they need?"). For each question an image demonstrating the two given quantities of the problem (e.g., In the given example, an image of five kids and three pens) was provided to help the child find the solution. Five items were asked.

Number-line. In the number-line task, children were shown a number-line from 0 to 10 with a tick mark crossing the line. Children had to choose from six options the number indicated by the tick mark.

Measures for the longitudinal follow-up

For the longitudinal follow-up in 1st grade, participants results on the national standardized tests were collected. These tests were administered by teachers simultaneously to the entire class. The lab was not in contact with these teachers. Data was sent to the lab by the Ministry of Educations's department of evaluation, of prospecting and performance (DEPP). Children wrote their responses with pen-on-paper, except for several questions that required reading aloud individually to the teacher. The test was given in the beginning of the 1st grade (September) and in the middle of the school year (late January to early February). The test given in September was identical to the intervention test (minus the fluency test). The mid-year test included more difficult material (i.e., bigger numbers, reading and spelling) to consider expected learning progress between the two time periods. The beginning-of-the-year test provides teachers with an objective measure of the student's achievement after the frequent loss of skills acquired over the two summer months (Alexander et al., 2001; McCoach et al., 2006). The middle-of-the-year evaluation provides information about expected progress. The goal of these evaluations is to measure school learning and alert teachers to each student's strengths and difficulties. The test booklets and teacher guides as edited by the ministry of education are open access (https://eduscol.education.fr/2295/evaluations-des-acquis-et-besoins-des-eleves-au-cp).

Statistical analysis for the intervention

We first analyzed responses for the randomized part of the experimental design using a mixed analysis of variance with a 2-level between-group factor (read/math versus math/read) and a 3-level within-group test period factor (pretest, midtest, endtest). We then re-ran the same analysis including the non-randomized no-tablet control (3×3 Anova). The composite predictor was used as a covariable of pre-intervention knowledge. Anonymized data and scripts are available upon request.

In a crossover design, an effect of training is expressed as a significant interaction between the group and test period. Because all children are expected to learn during the year, the passive control provides us with a measure of normal progression. To correct for multiple comparisons, we used a Bonferroni correction for each comparison by dividing the threshold on Type I error rate (α) by the number k of summary measures that we assessed, $\alpha/k = .05/5 = .01$ for the language and reading subtest, and $\alpha/k = .05/3 = .017$ for the number and math subtest. The results we report all passed this threshold.

Participant attrition and baseline characteristics

Only children that completed the three tests given during the kindergarten intervention were examined. This included 783 children (read/math= 300, math/read = 320, control = 163). To examine the baseline balance between the two intervention and control groups, we performed an overall F test or Chi-square test, as well as pairwise comparisons between the three groups in terms of baseline demographic variables, the three predictive variables at pretest (visual attention, vocabulary and syllable awareness) and the pretest results in each domain (see Table 1). Age and gender distribution were equal across groups (ps > .1). A significantly greater percentage of children from our intervention groups spoke another language besides French at home, two by two group comparisons revealed that this difference was restricted to the intervention compared to the passive control (read/math vs. math/read = $t(607)=0.79$, $p = 0.43$; read/math vs. control = $t(459)=3.89$, $p < 0.001$; math/read vs. control = $t(462)=3.25$, $p < 0.001$. It should be noted that national data shows that 10,2% of the population stems from immigration in France (Institut National de la Statistique et des Etudes Economiques, 2021). It is possible that the randomized intervention population was a better representation of France, thus highlighting the importance of RCT.

On the pretest measures, no differences were found between groups, except for on the fluency task. Critically, two by two comparisons revealed that a difference between the two intervention

Table 1. Baseline characteristics at pretest for children that completed the intervention.

	Group						
	Read/math		Math/read		Control		
N =	300		320		163		
N = female	152		144		82		
	M	SD	M	SD	M	SD	$\chi^2(2)=2.33$, $p = 0.31$
Age in months	62	3.98	62	3.54	63	3.59	$F(2, 780)=1.21$, $p = 0.30$
Bilingual	14%		12%		4%		$F(2, 780)=5.52$, $p < 0.01$
Pretest predictive measures							
visual attention	60%	.27	63%	.29	63%	.28	$F(2, 780)=0.62$, $p = 0.540$
syllable awareness	31%	.20	31%	.20	33%	.20	$F(2, 780)=0.87$, $p = 0.418$
vocabulary	42%	.28	41%	.29	44%	.28	$F(2, 780)=0.63$, $p = 0.531$
Reading predictor and phonics knowledge							
oral comprehension	49%	.30	47%	.31	51%	.30	$F(2, 780)=0.93$, $p = 0.395$
phoneme awareness	21%	.13	22%	.16	22%	.15	$F(2, 780)=0.35$, $p = 0.703$
letter-name knowledge	26%	.25	23%	.24	25%	.23	$F(2, 780)=0.87$, $p = 0.421$
GP matching	33%	.23	32%	.24	31%	.22	$F(2, 780)=0.54$, $p = 0.585$
One-minute word and pseudoword reading[a]	.29	.71	.21	.44	.48	1.3	$F(2, 780)=6.41$, $p = 0.002$
Math Predictors and number knowledge							
number knowledge	67%	.23	68%	.23	68%	.20	$F(2, 780)=0.33$, $p = 0.717$
word-problem	30%	.27	32%	.28	30%	.27	$F(2, 780)=0.36$, $p = 0.695$
number-line	23%	.23	20%	.21	21%	.20	$F(2, 780)=1.28$, $p = 0.278$

[a]Children were not expected to know how to read at the beginning of kindergarten and only attempted the question if they could read the first two trial words, otherwise they were scored 0. The significant group effect can be attributed to the greater percentage of children that were able to attempt this question (r/m = 25%, m/r = 20%, control = 29%). However, as the average number of words read for each group was < 1, we did not remove participants to create equal groups.

groups, $t(493)=1.82$, $p=0.07$, was not significant. The control group read more items than the math/read group $t(180)=2.59$, $p=.01$, but not the read/math group $t(214)=1.68$, $p=.09$. These differences were mainly due to the fact that few children attempted this test. Children were required to read the first two trial syllables in order to continue, and the proportions of children who did differed across groups (read/math = 25%, math/read = 20%, control = 29% of the participants succeeded in reading the trials). We decided to not adjust groups based on this variable as, on average, all groups read less than one item per minute.

Results

Intervention effects on pre-literacy skills

Children's progress over the three test periods on the reading measures reading is shown in Figure 3 and the math measures in Figure 4. As expected, there was a significant effect of session on all test tasks, $p < 0.001$, demonstrating learning progression through the school year. There was also a significant effect of the pretest predictor covariable for all measures, $p < 0.001$. Over and above those expected effects, we now report for each measure the crucial group by session interactions.

On the two orally presented early predictive measures for reading, i.e. oral comprehension and phoneme awareness, children in both intervention groups performed equally across all three

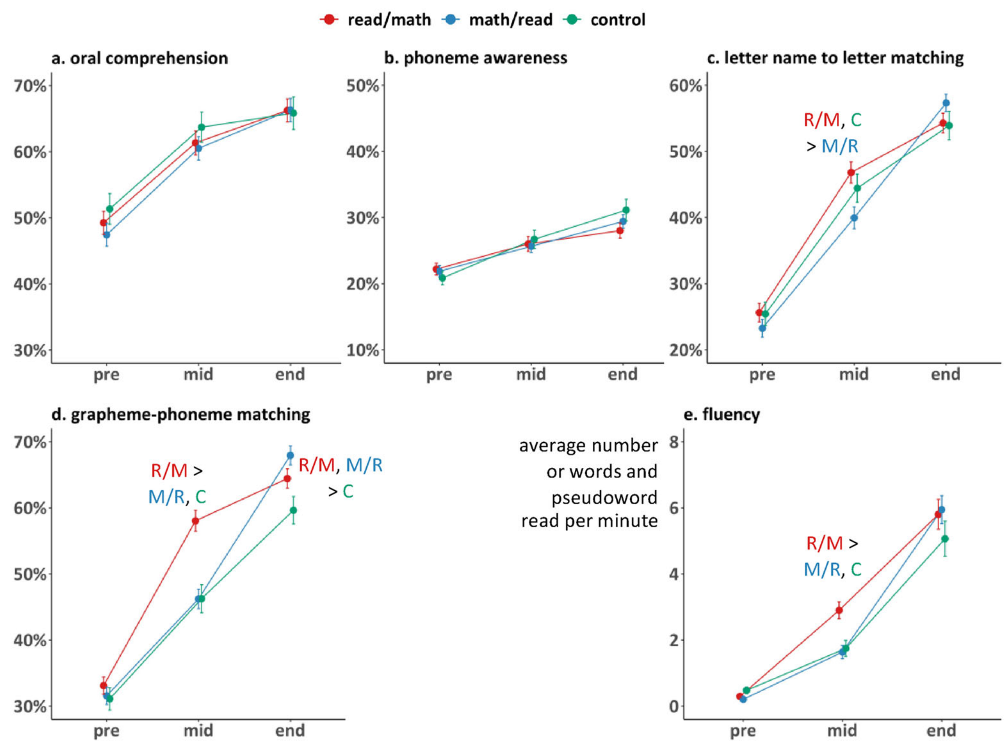

Figure 3. Group performance improvements on tasks of reading predictors and skills.
Group averages on language & reading tasks at the three intervention time-points. All tests improved across time. a) oral comprehension (listening to short sentence and choosing the corresponding picture), b) phoneme awareness (choosing a picture description of a word that either starts or ends with a target word), c) letter name knowledge (hearing a letter name and choosing three different orthographic presentations (upper or lower case; print or cursive) of the correct letter amongst distractors), d) grapheme-phoneme matching (hearing a phoneme and choosing the corresponding grapheme), e) word and pseudo-word reading (average number of items read in one minute). Pairwise differences that were significant at $p < 0.05$ are highlighted (RM read/math group; MR math/read group; c control group).

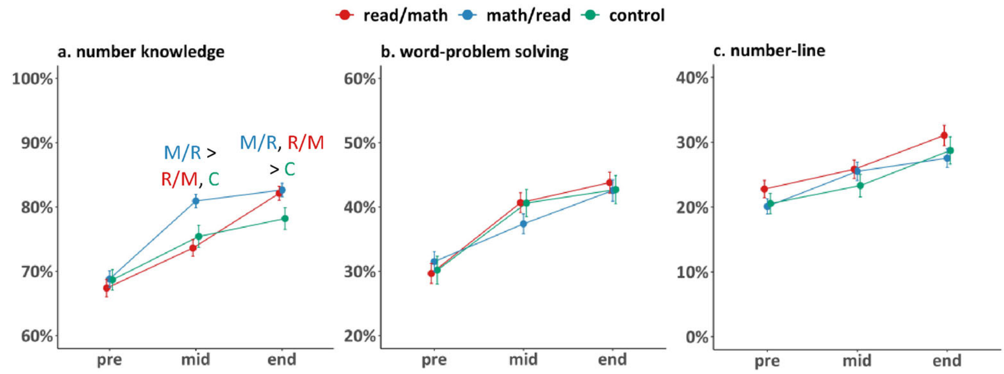

Figure 4. Group performance improvements on tasks of math predictors and number knowledge.
Group averages on number & math tasks at the three intervention time-points. All tests improved across time. a) number knowledge (a combined score of matching number name ('two') and number symbol ('2') and number symbol with quantity ('**')), b) word problem solving, c) number-line. Pairwise differences that were significant at $p < 0.05$ are highlighted (RM read/math group; MR math/read group; C control group).

periods, $F(2, 1234)=0.24$, $p = 0.788$ and $F(2, 1234)=0.79$, $p = 0.455$, respectively. Including the passive control group, we did not find a group by test interaction on either of the measures (oral comprehension: $F(4, 1558)=0.45$, $p = 0.769$; and phoneme awareness: $F(4, 1558)=1.40$, $p = 0.232$).

On letter name knowledge task, we reported an interaction, $F(2, 1234)=9.22$, $p < 0.001$. Post hoc analysis at each test period showed that the intervention groups started the year with similar knowledge, $F(1, 617)=2.04$, $p = 0.154$. After the first training period, children that used Kalulu Phonics recognized more letters than children that used Kalulu Numbers, $F(1, 617)=10.64$, $p < 0.001$. This initial advantage disappeared once both groups had had training with Kalulu Phonics, $F(1, 617)=2.27$, $p = 0.132$. Including the control group produced a group by session interaction, $F(4, 1558)=4.88$, $p < 0.001$. Post hoc analysis at each test period confirmed that all three groups began on par, $F(2, 779)=1.01$, $p = 0.363$. An effect of group at midtest ($F(2, 779)=5.36$, $p = 0.005$) reflected an advantage only for children that used Kalulu Phonics compared to Numbers (read/math vs. control, $F(1, 460)=1.77$, $p = 0.184$; math/read vs. control, $F(1, 480)=1.87$, $p = 0.172$). We did not find a group difference at the end of the year, $F(2, 779)=1.95$, $p = 0.143$.

For the GPC matching, we also found a significant group by session interaction, $F(2, 1234)$, 25.71, $p < 0.001$. Post hoc analysis at each test period showed no effect at pretest, $F(1, 617)=1.10$, $p < 0.294$. Children that used Kalulu Phonics made on average more correct GPC matches by the mid-year test, $F(1, 617)=35.94$, $p < 0.001$, an advantage that disappeared at the endtest once the other group was exposed to the Phonics game, $F(1, 617)=2.88$, $p = 0.090$. Including the passive control group again produced a group by test session interaction, $F(4, 1558)=14.03$, $p < 0.001$. All three groups started the year with similar phonics knowledge, $F(2, 779)=1.26$, $p = 0.283$. An effect of group emerged at midtest, $F(2, 779)=22.12$, $p < 0.001$. The Kalulu Phonics group performed not only better than children who played the Number game, but also the passive control, $F(1, 460)=25.40$, $p < 0.001$, while no difference was found between the active and passive control, $F(1, 480)=0.14$, $p = 0.704$. An effect of group extended to the endtest, $F(2, 779)=7.54$, $p < 0.001$. By the end of the year, both intervention groups made more correct GPC matches than the control (read/math vs. control, $F(1, 460)=5.88$, $p = 0.016$; math/read vs. control, $F(1, 480)=14.68$, $p < 0.001$). At the end of the school year, the group that had used Kalulu Phonics games in the first session, after 4 months of no longer playing, maintained an advantage of 0.18 standard deviations over the control group, while the group that had used the phonics suite in the second half of the year had a 0.32 standard deviation advantage over the control group.

On the measure of fluency, the average number of words and pseudowords read in a minute, we found a significant intervention group by test interaction, $F(2, 1234)=5.32$, $p = 0.005$. At the start of the year, the read/math group trended to better reading on average, $F(1, 617)=3.86$, $p = 0.050$. As mentioned in the section 'Base-line characteristics at pretest', this was mainly due to the slightly different numbers of children that attempted this test. By the middle of the year, the group that used Kalulu Phonics read more words than the active control, $F(1, 617)=17.32$, $p < 0.001$. This advantage disappeared by the end of the year, $F(1, 617)=0.01$, $p = 0.943$. We also found a significant three group by test session interaction, $F(4, 1558)=4.62$, $p < 0.001$. Once again, due to the percentage of students able to attempt the test, there was a significant group difference at pretest, $F(2, 779)=5.97$, $p = 0.003$, with the control group starting the year with a trend to better reading than the read/math group, $F(1, 460)=3.20$, $p = 0.074$, and significantly greater early reading ability compared to the math/read group, $F(1, 480)=10.71$, $p = 0.001$. By the mid-year, a significant group effect, $F(2, 779)=11.17$, $p < 0.001$, revealed that the passive control group read fewer items than the read/math group ($F(1, 460)=11.65$, $p < 0.001$ for the latter test), but not the math/read group, $F(1, 480)=0.01$, $p = 0.933$. On the endtest, we did not find an effect of group, $F(2, 779)=1.44$, $p = 0.238$.

Control task in number and math understanding

We now present the results for the number and math knowledge tests. For the combined number knowledge portion of the test, we found a significant group by test interaction, $F(2, 1234)=8.50$, $p < 0.001$. Post hoc analysis showed that both intervention groups started on equal footing at pretest, $F(1, 617)=0.40$, $p = 0.529$. By the mid-year, children that played Kalulu Numbers made more correct responses when matching symbolic numbers with their names and quantities, $F(1, 617)=19.76$, $p < 0.001$. This difference disappeared by the end of the year, $F(1, 617)=0.05$, $p = 0.818$. Including the control group revealed a similar group by test interaction, $F(4, 1558)=5.26$, $p < 0.001$. Post hoc tests revealed no difference between groups at pretest $F(2, 779)=0.26$, $p = 0.772$. At the mid-test, a significant group effect, $F(2, 779)=10.75$, $p < 0.001$ highlighted improved results after playing Kalulu Numbers compared to the passive control, $F(1, 480)=10.97$, $p < 0.001$. Playing Kalulu Phonics and the passive control showed similar knowledge $F(1, 460)=0.24$, $p = 0.627$. At the end test, a group effect, $F(2, 779)=4.15$, $p = 0.016$, showed that both intervention groups made more correct responses than the control (read/math vs. control, $F(1, 460)=6.38$, $p = 0.012$; math/read vs. control, $F(1, 480)=7.03$, $p = 0.008$). At the end of the school year, the group that had used Kalulu Numbers in the first session, after 4 months of no longer playing, maintained an advantage of 0.18 standard deviations over the control group, while the group that had it in the second half of the year had a 0.20 standard deviation advantage over the control group.

The word problem solving and number-line tasks did not produce a group by test interaction when taking into account only the two randomized groups (word problem solving, $F(2, 1234)=2.17$, $p = 0.115$; number-line, $F(2, 1234)=0.99$, $p = 0.372$) or when including the passive control (word problem solving, $F(4, 1558)=1.33$, $p = 0.258$; number-line, $F(4, 1558)=0.68$, $p = 0.608$).

Longitudinal follow-up

While using the Kalulu software boosted kindergartner's pre-reading in the trained domains, did it also boost 1st grade school readiness and lead to improved literacy success? Data was obtained from 94% of the intervention participants (read/math = 270, math/read = 312, control = 156). Consistent with the intervention analysis, we first looked at the randomized part of the study using one-way ANOVAs to test the main effect of group (read/math or math/read), the goal being to assess if children in the 1st grade retained more or less knowledge from the reading or math software depending on the period of intervention exposure. If no difference was found, we collapsed both groups and used the same ANOVA to compare them to the non-randomized control group. We used the same predictor covariable. We report results after a Bonferroni correction, obtained by dividing the Type I error rate by the number of measures assessed. For the beginning of the year evaluation, this is $\alpha/k = .05/4 = .013$ for the reading tests and $\alpha/k = .05/4 = .013$ for the math tests; for the mid-year evaluation, this is this is $\alpha/k = .05/6 = .008$ for the reading tests and $\alpha/k = .05/5 = .01$ for the math tests. The predictor continued to be significant on all tasks at $p < 0.001$. Degrees of freedom are adjusted to reflect the fact that some cells had missing data for certain questions within a test. This could have been due to a child being absent for a portion of the test. Figure 5 shows the performance at the beginning and middle of the 1st grade year for each of the read/math, math/read, and control groups, respectively, for all questions.

At the beginning of 1st grade, the two intervention groups showed similar abilities in all of the pre-reading tasks: oral comprehension, $F(1, 569)=0.03$, $p = 0.856$; letter-name knowledge, $F(1, 567)=0.09$, $p = 0.758$; phoneme-awareness, $F(1, 572)=0.37$, $p = 0.541$; and GPC matching, $F(1, 573)=0.15$, $p = 0.702$. We therefore collapsed the randomized groups and compared them to the passive control group using the same analysis. No difference was found between the intervention and passive control on the tasks of oral comprehension, $F(1, 725)=0.01$, $p = 0.943$ and letter-name

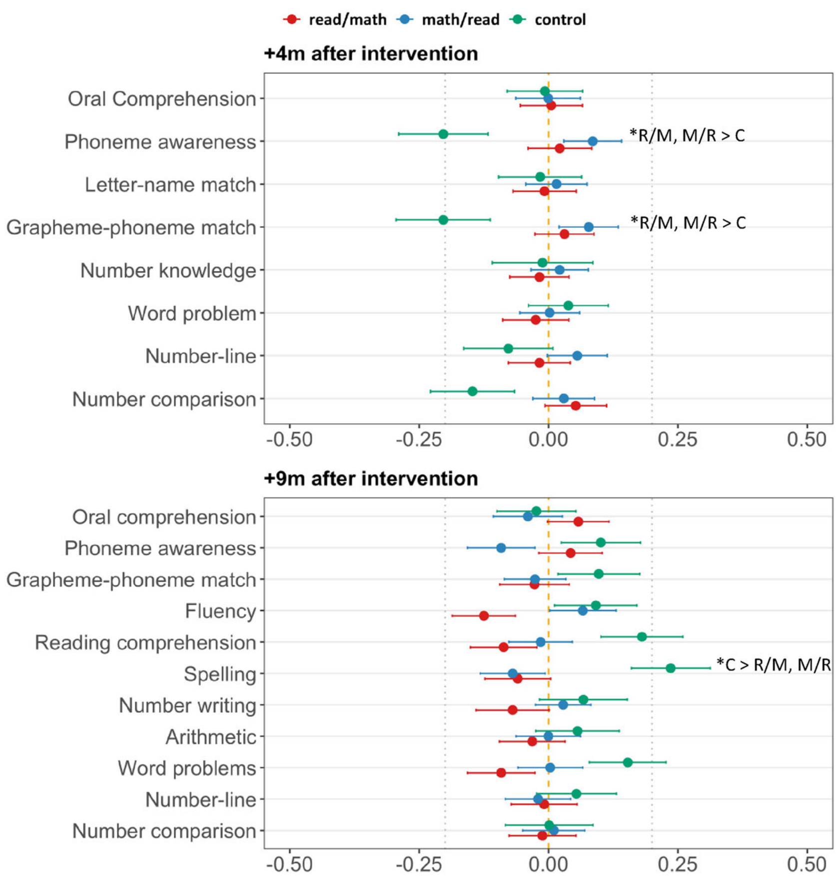

Figure 5. Longitudinal group outcomes on 1st grade evaluations.
Group averages on all tasks at the two 1st grade evaluation time-points. Z-scores, corrected for the composite predictor from the kindergarten analysis, for each question are presented to allow for comparison across all tasks. * Highlights differences that were significant at Bonferroni p-corrected <0.05/k (RM read/math group; MR math/read group; C control group).

knowledge, $F(1, 721)=0.00$, $p = 0.999$. In the two tasks requiring phoneme manipulation, a skill directly taught by Kalulu Phonics, the intervention groups did significantly better than the control, in spite of the summer break: phoneme awareness, $F(1, 727)=7.90$, $p = 0.005$; GPC matching, $F(1, 728)=7.19$, $p = 0.007$. After Bonferroni correction, both effects remained significant.

By the middle of 1st grade, the two intervention groups still performed equally on the measures of: oral comprehension, $F(1, 543)=0.69$, $p = 0.406$; GPC matching $F(1, 545)=0.01$, $p = 0.922$; reading comprehension, $F(1, 556)=0.48$, $p = 0.487$; spelling, $F(1, 531)=0.02$, $p = 0.901$. On phoneme awareness, this time, students that had used the reading intervention early in their kindergarten year showed a slight advantage over students that used the method during the second half, $F(1, 556)=5.48$, $p = 0.020$ while an advantage in the opposite direction was found for fluency, $F(1, 548)=5.82$, $p = 0.016$. As neither of these group effects were significant after Bonferroni correction, the two intervention groups were collapsed for comparison with the passive control. After half-a-year of formal reading education, all groups performed equally well in oral comprehension, $F(1, 692)=0.11$, $p = 0.737$; phoneme awareness, $F(1, 708)=0.77$, $p = 0.379$; GPC matching, $F(1, 693)=1.67$, $p = 0.196$; fluency, $F(1, 698)=1.04$, $p = 0.307$. In tasks of reading comprehension and spelling, the control group performed better than the intervention group, $F(1, 708)=5.85$, $p = 0.016$ and $F(1, 676)=11.43$, $p < 0.001$, respectively. After Bonferroni correction, only the spelling efffect remained.

Longitudinal assessment of the control tasks

Looking at the control questions in number and math understanding, no difference was found between the randomized intervention groups on any of the tasks: number knowledge, $F(1, 564)=0.15$, $p = 0.702$, word-problems, $F(1, 576)=0.02$, $p = 885.$, number-line, $F(1, 573)=1.29$, $p = 0.256$, number comparison, $F(1, 569)=0.06$, $p = 0.814$. We collapsed these two groups and compared them to the passive control group. The intervention groups and control performed comparably on tasks of number knowledge, $F(1, 718)=0.02$, $p = 0.902$; word-problems, $F(1, 732)=0.58$, $p = 0.445$ and number-line, $F(1, 727)=1.01$, $p = 0.315$. On the number comparison task, children from the intervention made significantly more correct responses than the control, $F(1, 725)=4.46$, $p = 0.035$, but this difference did not survive Bonferroni correction.

As for math, in the middle of 1st grade, no difference was found between the two intervention groups on number writing, $F(1, 545)=1.37$, $p = 0.243$; word-problem solving, $F(1, 556)=0.46$, $p = 0.497$; arithmetic, $F(1, 533)=0.25$, $p = 0.616$; number-line, $F(1, 555)=0.28$, $p = 0.600$; number-comparison, $F(1, 547)=0.53$, $p = 0.469$. Collapsing the two intervention groups and comparing them to the control did not produce a group effect in number-writing, $F(1, 693)=0.490$, $p = 0.483$; arithmetic, $F(1, 678)=0.69$, $p = 0.407$; number-line, $F(1, 706)=0.44$, $p = 0.507$; number-comparison, $F(1, 694)=0.10$, $p = 0.750$. Children in the control group made more correct responses than the intervention group in solving word problems, $F(1, 708)=5.26$, $p = 0.022$, but this effect was not significant after Bonferroni correction.

Discussion

This first goal of the study sought to develop evidence for using Kalulu Phonics as a tool for instruction of the alphabetic code, in this case, for French children in kindergarten. Our reasoning was that, since French kindergarteners do not receive formal reading instruction, this would allow us to measure, more directly, the benefits from using the software without interference from the intense literacy training that children receive in 1st grade.First graders had previously only showed benefits from using the software in the beginning of the year. Our crossover measure clearly showed a benefit to using Kalulu Phonics in correspondence to the intervention/active control period for a group. Compared to the passive control group, using our game provided a significant gain in GPC mappings knowledge. This benefit imparted from the software was present regardless of the timing of the training session and highlighted the utility of Kalulu. Progress was made on the questions tested, relevant to the treatment intervention and independently of whether the training took place in the first half or the second half of the year demonstrating that kindergarten students could use and benefit from the phonics software. Teachers reported high engagement from the children, and in fact, all participants except for one child, remained in the project for the entire school year. Both intervention groups, irrelevant of session, behaved similarly in the longitudinal follow-up. This benefit was not merely associated to the novelty of the tablet intervention as we also measured a benefit for the number game on tests of number-quantity knowledge. Comparing this study with kindergarten students to the first graders previously tested, we show that Kalulu Phonics can impart phonics knowledge, even in the absence of intense classroom reading instruction (Potier Watkins et al., 2020).

Our second goal was to assess long term benefits and transfer to literacy gains. Encouragingly, children that had taken part in the intervention arrived in 1st grade with an advantage in GPC knowledge and phoneme awareness, four months after the intervention. Furthermore, this test period showed that students were able to transfer their knowledge to pen-and-paper tasks. This outcome was important as it showed that learning by using the tablet was not just context dependent. Both phoneme awareness and letter-sound knowledge are important predictors of reading performance (Bradley & Bryant, 1983; Cunningham, 2001; Melby-Lervåg et al., 2012; Piquard-Kipffer & Sprenger-Charolles, 2013). At the beginning of 1st grade, there was also a trend to improved number comparison ability compared to the passive control, which has been previously shown to be a predictor of 1st grade math performance (Lyons et al., 2014). This supported continued momentum for both intervention and control tasks.

Despite these encouraging outcomes at the start of 1st grade, by the middle of the year, there was no longer any detectable advantage to having been exposed to Kalulu in kindergarten. In fact, children from the passive control did better in spelling, and showed a trend toward greater reading comprehension and word problem solving. What could have happened from the start to the middle of 1st grade is unknown to us, as we do not have access to school information. One possibility is that software training yields a long-term negative effect. In the past, methods promoting pure phonics instruction have been criticized for possibly developing an over-dependence on decoding skills while ignoring the end goal of fluent reading, i.e., word recognition. Software games may also distract children from more classical and possibly more efficient modes of classroom learning. While a thorough rejection of those hypotheses requires further research, we find them doubtful given our earlier results that comprehension was improved when 1st grade children used our phonics suite, with benefits lasting four months after the intervention (Potier Watkins et al., 2020).

Another explanation, supported by the results from our previous intervention, is that this project minimized the importance of introducing GPCs simultaneously with the focused experience of learning to read in the classroom. As stated, the goal of our phonics software was to provide children with many GPCs early in learning. In fact, children that participated in the intervention were exposed to more GPC training than children in the control group, as shown by the end test and early 1st grade results. While we assumed that this advantage would boost literacy learning, it may have had the adverse effect of decorrelating phonics education from classroom practices such as actual reading aloud. In other words, our students learned about phonics, but given that our software did not combine this with actual reading aloud, children could not apply this new knowledge to actual reading practice, perhaps leading to a disconnection of their abstract GPC knowledge from any practical application.

In this respect, this intervention underscores the importance of ensuring that apps clearly define their learning goals and are evidence-based. It has been proposed that apps for education heed the developmental principles of the science of learning: promoting active learning, engagement, meaningful learning, and social interaction (Hirsh-Pasek et al., 2015). Our phonics game did promote active and engaged learning as demonstrated by user fidelity to the project and the immediate learning gains. Social interaction within the small work groups was also reported by teachers. There is less evidence, however, that our intervention anchored GPC learning with a meaningful goal— learning how to read. Meaningful learning has been defined as the integration of newly learned knowledge to knowledge already possessed (Ausubel et al., 1978). It is described in juxtaposition to rote and drill-and-practice learning which do not aim to anchor new learning to knowledge. Research has shown that drill-and-practice does benefit transfer from slow and effortful to fast retrieval (Chen & Chan, 2019; Ojanen et al., 2015; Poncy et al., 2007; Walker et al., 2013). Traditionally, the teacher establishes the connection between meaning, and exercises to develop automaticity, and indeed, when Kalulu Phonics was used in this context of supporting classroom literacy education the greatest gains in our tests were in reading comprehension. However, in the current intervention, the kindergarten students' proficiency in GPC matching was not aligned with actual reading goals.

Furthermore, 1st grade teachers were not aware that students from the intervention had already started to build their phonics knowledge, and hence could not provide adapted literacy practice or refer to learned materials from the app to literacy in the class. Thus, students from the intervention probably began their 1st grade year reviewing learned GPCs, while children from the 'business-as-usual' control would have learned this novel information simultaneously with common classroom practices, such as reading-aloud. In this scenario, children from the control condition may have benefited from the advantageous situation of introducing phonics in combination with structured reading practice, while the intervention students could have been demotivated by receiving a mere review of what they already knew or not connecting previously learned material from the game to reading. Our design possibly recreated a similar situation to the one that we saw in our first experiment, whereby children that used the games as review did not benefit, only in this case the 1st grade classroom experience was the review (Potier Watkins et al., 2020).

A third explanation is possible: the changes seen during 1st grade may be due to the French national assessments. As described earlier, these tests are specifically designed to evaluate the pupils at the beginning of 1st grade and provide individualized feedback in a timely manner. Children from the control group began the year with significantly lower skills in phoneme awareness and GPC matching. It is not impossible that these students received additional targeted individualized attention in order to subvert this early deficit. The National Education system of France provides teachers with both detailed explanations of the tests and corresponding targeted remediation exercises.

Future directions

Teachers' reports on the Kalulu software were generally positive, and children learned, played, and enjoyed the games. Our participants made initial gains in reading readiness, and these improvements persisted after the summer-lag, whereas a slight decline across all academic areas is generally seen (Borman et al., 2005; Cooper et al., 1996; Heyns, 1987). We cannot say, however, that Kalulu training in kindergarten produced long-term gains in literacy. The present children improved in tasks of symbol to meaning understanding compared to the control group, but unlike with 1st graders, this knowledge did not transfer to fluency or comprehension improvement.

Any future direction of this project should seek to improve the disappointing follow-up results, i.e., the fact that, in spite of much previous research showing that phoneme awareness and letter-sound knowledge are predictors of reading acquisition (National Reading Panel, 2000; Piquard-Kipffer & Sprenger-Charolles, 2013) the initial boosts with Kalulu did not suffice to advance literacy. There are many aspects of the Kalulu game that could be improved. One of the most glaring issues is that the game teaches reading indirectly, by clicking on letters, words or pictures, rather than directly through reading aloud. To the best of our knowledge, this is true of all similar software that does not require an adult facilitator and is due to the unreliability of current technology in recognizing the nuances of children's vocal productions. A solution would be to integrate reliable voice-recognition or involve teachers or advanced students in reading aloud practice. Our game requires children to read in their heads and to associate an orally presented syllable or word with its written form. Some of the harder games also require encoding. For example, children hear a word, then have to break it down into its phonemes and assemble the corresponding graphemes one by one. Given that our software does not include voice recognition, children could not apply this new knowledge to actual reading practice, perhaps leading to a disconnection of their abstract GPC knowledge from any practical application. To our knowledge, there has never been research done on the use of voice recognition software to test children's reading. It would be interesting to test if this could be a possible future solution.

Another criticism may be that our intervention was too early, and that it would be better to introduce it in 1st grade. Not much is known about the optimal windows for reading acquisition, and we did obtain better results in a previous 1st grade intervention (Potier Watkins et al., 2020). We would also argue that there is enough research showing that reading instruction can begin as early as kindergarten (Comaskey et al., 2009; Cunningham, 2001; Wolf, 2016). Developing a kindergarten-based pedagogy that emphasizes early learning of GPCs and reading of extremely simple decodable text may free up time for 1st grade teachers to emphasize comprehension and learning of nontransparent words. Given that only a single child asked to stop using the Kalulu software, we see no reason why children would not benefit from such an integrated early reading program. Indeed, a clear limit of the present approach is that the teaching that was provided in 1st grade was entirely disconnected with the prior reading experience that our software provided in the previous year.

Future research should aim to develop a more integrated computer-aided curriculum that could be adopted by the teachers and take into account reading readiness in kindergarten. Learning goals and their finality must be clearly stated, as well as researched (Hirsh-Pasek et al., 2015). To this aim, the guidebook that we wrote to accompany the game has since been improved and made into a student reader with class lessons, reading aloud exercises and activities, all of which follow the Kalulu software pedagogical progression. Previous research using literacy apps has highlighted the importance of teacher fidelity implementation as central to intervention success (Abrami et al., 2020). Changes such as these would also be important to remediating another weakness in the Kalulu Phonics suite: Kalulu was only successful in imparting knowledge on the learning tasks that were made explicit by the game, transfer to fluency and comprehension should be the goal.

Conclusion

It is now undisputed that phonics instruction is an essential component in learning to read, but it is important that RCT field interventions examine the many facets of when and how to teach phonics in order to go beyond tenets and provide teachers with applicable tools. Furthermore, interventions that posit classroom gains should be evaluated using standardized tests and tested for long-term benefits. When gains are limited to short-term benefits and simply mirror the functions that were taught, they may not be followed by any subsequent classroom-relevant benefits. By evaluating our intervention using standardized school-based measures and looking at long term outcomes, we obtained a better, if somewhat disappointing, picture of the true benefits and shortcomings of our software. Obviously, much work remains to improve future implementations of the software. Working in tandem, teachers, researchers and school policy makers should all endeavor to build a system of checks-and-balances with the goal of building a curriculum that provides the best step-by-step support for teachers and the greatest impact on the child's learning.

We have no conflicts of interest to disclose.

Funding

This project was supported by INSERM, CEA and NeuroSpin and financed by the program: Investissements d'Avenir Action: « Innovations numériques pour l'excellence éducative » volet Espace de Formation, de Recherche et d'Animation Numérique (eFRAN). The project was approved by the ethical committee CERES (Conseil d'évaluation éthique pour les recherches en santé, project N° 2017 - 27).

ORCID

Cassandra Potier Watkins http://orcid.org/0000-0002-6588-0614
Stanislas Dehaene http://orcid.org/0000-0002-7418-8275

References

Abrami, P. C., Lysenko, L., & Borokhovski, E. (2020). The effects of ABRACADABRA on reading outcomes: An updated meta-analysis and landscape review of applied field research. *Journal of Computer Assisted Learning*, *36*(3), 260–279. https://doi.org/10.1111/jcal.12417

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis*, *23*(2), 171–191. https://doi.org/10.3102/01623737023002171

Archer, K., Savage, R., Sanghera-Sidhu, S., Wood, E., Gottardo, A., & Chen, V. (2014). Examining the effectiveness of technology use in classrooms: A tertiary meta-analysis. *Computers & Education*, *78*, 140–149. https://doi.org/10.1016/j.compedu.2014.06.001

Arnas, Y. A., Siğirtmaç, A. D., & Gül, E. D. (2004). A study of 60- to 89-MO.-old children's skill at writing numerals. *Perceptual and Motor Skills*, *98*(2), 656–660. https://doi.org/10.2466/pms.98.2.656-660

Ausubel, D. P., Novak, J. D., & Hanesian, H. (1978). *Educational psychology: A cognitive view* (2nd ed.). Rinehart and Winston.

Bara, F., & Gentaz, E. (2011). Haptics in teaching handwriting: The role of perceptual and visuo-motor skills. *Human Movement Science*, *30*(4), 745–759. https://doi.org/10.1016/j.humov.2010.05.015

Bara, F., Gentaz, E., Colé, P., & Sprenger-Charolles, L. (2004). The visuo-haptic and haptic exploration of letters increases the kindergarten-children's understanding of the alphabetic principle. *Cognitive Development*, *19*(3), 433–449. https://doi.org/10.1016/j.cogdev.2004.05.003

Bishop, D. V. (2003). *Test for reception of grammar: Version 2: TROG-2*. Harcourt Assessement.

Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research*, *72*(1), 101–130. https://doi.org/10.3102/00346543072001101

Boles, D. B., & Clifford, J. E. (1989). An upper- and lowercase alphabetic similarity matrix, with derived generation similarity values. *Behavior Research Methods, Instruments, & Computers*, *21*(6), 579–586. https://doi.org/10.3758/BF03210580

Borman, G. D., Benson, J., & Overman, L. T. (2005). Families, schools, and summer learning. *The Elementary School Journal*, *106*(2), 131–150. https://doi.org/10.1086/499195

Boyer, N., & Ehri, L. C. (2011). Contribution of phonemic segmentation instruction with letters and articulation pictures to word reading and spelling in beginners. *Scientific Studies of Reading*, *15*(5), 440–470. https://doi.org/10.1080/10888438.2010.520778

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature*, *301*(5899), 419–421. https://doi.org/10.1038/301419a0

Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., Lyytinen, H., Brandeis, D., & Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(17), 7939–7944. https://doi.org/10.1073/pnas.0904402107

Byrne, B. (1992). *Studies in the acquisition procedure for reading: Rationale, hypotheses, and data. Reading acquisition*. Routledge. https://doi.org/10.4324/9781351236904

Castiglioni-Spalten, M. L., & Ehri, L. C. (2003). Phonemic awareness instruction: Contribution of articulatory segmentation to novice beginners' reading and spelling. *Scientific Studies of Reading*, *7*(1), 25–52. https://doi.org/10.1207/S1532799XSSR0701_03

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological science in the Public Interest: A Journal of the American Psychological Society*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Cavanaugh, C. L., Kim, A.-H., Wanzek, J., & Vaughn, S. (2004). Kindergarten reading interventions for at-risk students: Twenty years of research. *Learning Disabilities: A Contemporary Journal* 2.1 (2004): 9–21.

Chen, R. W., & Chan, K. K. (2019). Using augmented reality flashcards to learn vocabulary in early childhood education. *Journal of Educational Computing Research*, *57*(7), 1812–1831. https://doi.org/10.1177/0735633119854028

Comaskey, E. M., Savage, R. S., & Abrami, P. (2009). A randomised efficacy study of web-based synthetic and analytic programmes among disadvantaged urban Kindergarten children. *Journal of Research in Reading*, *32*(1), 92–108. https://doi.org/10.1111/j.1467-9817.2008.01383.x

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, *66*(3), 227–268. https://doi.org/10.3102/00346543066003227

Cunningham, J. W. (2001). The national reading panel report. *Reading Research Quarterly*, *36*(3), 326–335. https://doi.org/10.1598/RRQ.36.3.5

de Graaff, S., Bosman, A. M. T., Hasselman, F., & Verhoeven, L. (2009). Benefits of systematic phonics instruction. *Scientific Studies of Reading*, *13*(4), 318–333. https://doi.org/10.1080/10888430903001308

Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews. Neuroscience*, *16*(4), 234–244. https://doi.org/10.1038/nrn3924

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science (New York, N.Y.)*, *330*(6009), 1359–1364. https://doi.org/10.1126/science.1194140

Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS biology*, *16*(3), e2004103. https://doi.org/10.1371/journal.pbio.2004103

Deltour, J., & Hupkens, D. (1980). Test de vocabulaire actif et passif pour enfants (5 à 8 ans). *Paris, Editions et Applications Psychologiques* (1980).

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., Means, B., Murphy, R., Penuel, W., Javitz, H., Emery, D., Sussex, W. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. https://telearn.archives-ouvertes.fr/hal-00190019

Ehri, L. C., & Flugman, B. (2018). Mentoring teachers in systematic phonics instruction: Effectiveness of an intensive year-long program for kindergarten through 3rd grade teachers and their students. *Reading and Writing*, *31*(2), 425–456. https://doi.org/10.1007/s11145-017-9792-7

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: evidence from the national reading panel's meta-analysis. *Review of Educational Research*, *71*(3), 393–447. https://doi.org/10.3102/00346543071003393

Elofsson, J., Gustafson, S., Samuelsson, J., & Träff, U. (2016). Playing number board games supports 5-year-old children's early mathematical development. *The Journal of Mathematical Behavior*, *43*, 134–147. https://doi.org/10.1016/j.jmathb.2016.07.003

Fischer, J.-P. (2011). Mirror writing of digits and (capital) letters in the typically developing child. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, *47*(6), 759–762. https://doi.org/10.1016/j.cortex.2011.01.010

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PloS one*, *9*(2), e89900. https://doi.org/10.1371/journal.pone.0089900

Goigoux, R. (2017). *Rapport de recherche «Lire et écrire au CP–Étude de l'influence des pratiques d'enseignement de la lecture et de l'écriture sur la qualité des premiers apprentissages*. http://ife.ens-lyon.fr/ife/recherche/lire-ecrire/rapport/rapport-lire-et-ecrire

Gottfried, J. A., Sancar, F., & Chatterjee, A. (2003). Acquired mirror writing and reading: Evidence for reflected graphemic representations. *Neuropsychologia*, *41*(1), 96–107. https://doi.org/10.1016/S0028-3932(02)00130-6

Grant, A., Wood, E., Gottardo, A., Evans, M. A., Phillips, L., & Savage, R. (2012). Assessing the content and quality of commercially available reading software programs: Do they have the fundamental structures to promote the development of early reading skills in children? *NHSA Dialog*, *15*(4), 319–342. https://doi.org/10.1080/15240754.2012.725487

Groupe de travail du Csen. (2019). *ÉvalAide, un dispositif scientifique de prévention des difficultés en lecture et en mathématiques en CP et en CE1*. Canopé. https://www.reseau-canope.fr/fileadmin/user_upload/Projets/conference_role_experimentation_domaine_educatif/EvalAide_CSEN_Definitif_Mai2019.pdf

Hattie, J. (2010). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement (Reprinted)*. Routledge.

Heyns, B. (1987). Schooling and cognitive development: Is there a season for learning? *Child Development*, *58*(5), 1151–1160. https://doi.org/10.2307/1130611

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, *16*(1), 3–34. https://doi.org/10.1177/1529100615569721

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799

Institut National de la Statistique et des Etudes Economiques. (2021). *L'essentiel sur… Les immigrés et les étrangers*. https://www.insee.fr/fr/statistiques/3633212

Kendeou, P., Savage, R., & van den Broek, P. (2009). Revisiting the simple view of reading. *The British Journal of Educational Psychology*, *79*(Pt 2), 353–370. https://doi.org/10.1348/978185408X369020

Laski, E. V., & Siegler, R. S. (2014). Learning from number board games: You learn what you encode. *Developmental Psychology*, *50*(3), 853–864. https://doi.org/10.1037/a0034321

Lecocq, P. (1996). *L'ECoSSe: Une épreuve de compréhension syntaxico-sémantique*. Presses Universitaires du Septentrion.

Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *36*(1), 156–166. https://doi.org/10.3758/bf03195560

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science*, *17*(5), 714–726. https://doi.org/10.1111/desc.12152

Macaruso, P., & Walker, A. (2008). The efficacy of computer-assisted instruction for advancing literacy skills in kindergarten children. *Reading Psychology*, *29*(3), 266–287. https://doi.org/10.1080/02702710801982019

Maurer, U., Blau, V. C., Yoncheva, Y. N., & McCandliss, B. D. (2010). Development of visual expertise for reading: Rapid emergence of visual familiarity for an artificial script. *Developmental neuropsychology*, *35*(4), 404–422. https://doi.org/10.1080/87565641.2010.480916

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology*, *98*(1), 14–28. https://doi.org/10.1037/0022-0663.98.1.14

Melby-Lervåg, M., Lyster, S.-A H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological bulletin*, *138*(2), 322–352. https://doi.org/10.1037/a0026744

Monzalvo, K., & Dehaene-Lambertz, G. (2013). How reading acquisition changes children's spoken language network. *Brain and Language*, *127*(3), 356–365. https://doi.org/10.1016/j.bandl.2013.10.009

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769)*. National Institute of Child Health & Development. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

OECD. (2015). *Students, computers and learning*. OECD Publishing. https://doi.org/10.1787/9789264239555-en

Ojanen, E., Jere-Folotiya, J., Yalukanda, C., Sampa, F., Nshimbi, C., Katongo, M., Choopa, M., & Lyytinen, H. (2015). Mobile solution for better reading instruction in rural Africa. *2015 IST-Africa Conference*, 1–13 IEEE. https://doi.org/10.1109/ISTAFRICA.2015.7190559

Ojanen, E., Ronimus, M., Ahonen, T., Chansa-Kabali, T., February, P., Jere-Folotiya, J., Kauppinen, K.-P., Ketonen, R., Ngorosho, D., Pitkänen, M., Puhakka, S., Sampa, F., Walubita, G., Yalukanda, C., Pugh, K., Richardson, U., Serpell, R., & Lyytinen, H. (2015). GraphoGame – A catalyst for multi-level promotion of literacy in diverse contexts. *Frontiers in Psychology*, *6*, 671. https://doi.org/10.3389/fpsyg.2015.00671

Piquard-Kipffer, A., & Sprenger-Charolles, L. (2013). Early predictors of future reading skills: A follow-up of French-speaking children from the beginning of kindergarten to the end of the second grade (age 5 to 8). *L'Année Psychologique*, *113*(04), 491–521. https://www.researchgate.net/profile/Liliane_Sprenger-Charolles/publication/278736652_Prdiction_ds_la_maternelle_des_habilets_de_lecture_de_fin_de_CE1._Suivi_de_85_enfants_de_langue_maternelle_franaise_de_4__8_ans/links/558bc80c08aee99ca9ca5b8b.pdf

Poncy, B. C., Skinner, C. H., & Jaspers, K. E. (2007). Evaluating and comparing interventions designed to enhance math fact accuracy and fluency: Cover, copy, and compare versus taped problems. *Journal of Behavioral Education*, *16*(1), 27–37. https://doi.org/10.1007/s10864-006-9025-7

Potier Watkins, C., Caporal, J., Merville, C., Kouider, S., & Dehaene, S. (2020). Accelerating reading acquisition and boosting comprehension with a cognitive science-based tablet training. *Journal of Computers in Education*, *7*(2), 183–212. https://doi.org/10.1007/s40692-019-00152-6

Potier Watkins, C., Dehaene, O., & Dehaene, S. (2019). Automatic construction of a phonics curriculum for reading education using the transformer neural network. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial intelligence in education* (pp. 226–231). Springer International Publishing. https://doi.org/10.1007/978-3-030-23207-8_42

Programme d'enseignement de l'école maternelle. (2015, March 25). Ministère de l'Education Nationale de la Jeunesse et des Sports. https://www.education.gouv.fr/bo/15/Special2/MENE1504759A.htm

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child development*, *79*(2), 375–394. https://doi.org/10.1111/j.1467-8624.2007.01131.x

Richardson, U., & Lyytinen, H. (2014). The GraphoGame Method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*, *10*(1), 39–60. https://doi.org/10.17011/ht/urn.201405281859

Rose, J. (2006). *Independent review of the teaching of early reading: Final report*. DfES. http://dera.ioe.ac.uk/5551/3/5d970d28fc535dc54eb4dee9995bef36.pdf

Savage, R., Abrami, P. C., Piquette, N., Wood, E., Deleveaux, G., Sanghera-Sidhu, S., & Burgos, G. (2013). A (Pan-Canadian) cluster randomized control effectiveness trial of the ABRACADABRA web-based literacy program. *Journal of Educational Psychology*, *105*(2), 310–328. https://doi.org/10.1037/a0031025

Skwarchuk, S.-L., Sowinski, C., & LeFevre, J.-A. (2014). Formal and informal home learning activities in relation to children's early numeracy and literacy skills: The development of a home numeracy model. *Journal of Experimental Child Psychology*, *121*, 63–84. https://doi.org/10.1016/j.jecp.2013.11.006

Sprenger-Charolles, L., Colé, P., Béchennec, D., & Kipffer-Piquard, A. (2005). French normative data on reading and related skills from EVALEC, a new computerized battery of tests (end Grade 1, Grade 2, Grade 3, and Grade 4)1. *European Review of Applied Psychology*, *55*(3), 157–186. https://doi.org/10.1016/j.erap.2004.11.002

Sunde, K., Furnes, B., & Lundetrae, K. (2020). Does introducing the letters faster boost the development of children's letter knowledge, word reading and spelling in the first year of school? *Scientific Studies of Reading*, *24*(2), 141–158. https://doi.org/10.1080/10888438.2019.1615491

Torgerson, C., Brooks, G., & Hall, J. (2006). *A systematic review of the research literature on the use of phonics in the teaching of reading and spelling*. DfES Publications.

Torgerson, C., & Elbourne, D. (2002). A systematic review and meta-analysis of the effectiveness of information and communication technology (ICT) on the teaching of spelling. *Journal of Research in Reading*, *25*(2), 129–143. https://doi.org/10.1111/1467-9817.00164

Turkeltaub, P., Gareau, L., Flowers, D., Zeffiro, T., & Eden, G. (2003). Development of neural mechanisms of reading. *Nature Neuroscience*, *6*(7), 767–773. https://doi.org/10.1038/nn1065

Walker, D., Mickes, L., Bajic, D., Nailon, C. R., & Rickard, T. C. (2013). A test of two methods of arithmetic fluency training and implications for educational practice. *Journal of Applied Research in Memory and Cognition*, *2*(1), 25–32. https://doi.org/10.1016/j.jarmac.2013.02.001

Wolf, G. M. (2016). Letter-sound reading: Teaching preschool children print-to-sound processing. *Early childhood Education Journal*, *44*(1), 11–19. https://doi.org/10.1007/s10643-014-0685-y