J. Comput. Educ. (2020) 7(2):183–212
https://doi.org/10.1007/s40692-019-00152-6

Accelerating reading acquisition and boosting comprehension with a cognitive science-based tablet training

Cassandra Potier Watkins[1] · Julien Caporal[3] · Clément Merville[3] · Sid Kouider[4] · Stanislas Dehaene[1,2]

Received: 3 January 2019 / Revised: 22 October 2019 / Accepted: 19 December 2019 / Published online: 10 January 2020

Abstract Tablets and computers offer opportunities for learning, but their potential is only as great as the quality of the software they propose. Educational games must not only provide an engaging design, but also be based on principles from cognitive neuroscience and education research, and be evaluated in large-scale classroom tests. Here, we describe ELAN, an adaptive game that supports literacy acquisition through teaching and training phonics. It provides explicit systematic grapheme–phoneme correspondence instruction and reinforces full decoding through reading and spelling practice with 100% decodable text. The game also uses periodical lexical decision tasks to measure the transition from letter-by-letter decoding to fluent word recognition. The software was tested in a randomized control trial in 44 first-grade classrooms ($n = 975$ French children). Children who used ELAN software during the first term improved relative to two control groups, respectively, using math software or no-tablet "business-as-usual" classrooms. Improvements were significant in reading fluency (one-minute word and pseudo-word reading) and sentence reading comprehension, consistent with the idea that improved decoding can help the child focus on understanding. These results emphasize the importance of early, explicit and systematic phonics training, and provide a new software tool to facilitate it.

✉ Cassandra Potier Watkins
cassandra.potier-watkins@cea.fr

[1] INSERM, UMR992, CEA, NeuroSpin Center, University Paris Saclay, Gif-sur-Yvette, France

[2] Collège de France, 11 Place Marcelin Berthelot, 75231 Paris Cedex 05, France

[3] Manzalab, 108 Rue Saint Honoré, 75001 Paris, France

[4] Brain and Consciousness Group, Ecole Normale Supérieure, PSL University, EHESS, CNRS, Paris, France

Keywords Reading · Education · First grade · Tablets computers · Intervention · Cognitive science

Introduction

International evaluations in schools have reported that major investments in technology for the classroom do not necessarily lead to improvements in math and reading abilities (OECD 2015). This is in part due to the lack of quality research-backed software that not only takes advantage of the experiences that technology can provide, but also incorporates scientific findings from cognitive neuroscience and education about how the brain learns. Previous research has shown that games that adhere to evidence backed pedagogy, gamify drill-and-practice, and use adaptive algorithms to provide individualized learning can lead to improvements in math (Räsänen et al. 2009; Schacter and Jo 2016; Wilson et al. 2006a, b; Wilson et al. 2009) and reading (de Graaff et al. 2009; Kyle et al. 2013; Lyytinen et al. 2007; Saine et al. 2011). The goal of the current project was to test the ELAN software, a tablet-based game designed to support reading acquisition in the classroom. The game's pedagogy is inspired from evidence in cognitive neuroscience and education research as to how children learn to read. The game focuses on phonics teaching and training, an essential stepping stone to reading mastery as made evident by the meta-analyses of the National Reading Panel (Cunningham 2001). ELAN was not designed to supplant the role of the teacher but aims to provide 'drill-and-practice' and tailored game difficulty to accelerate consolidation of the grapheme–phoneme correspondences in initial decoding stage of reading.

The reading brain

The goal of learning to read is to train the visual system to decode letter strings and provide a novel input into the already developed cortical areas for spoken language comprehension, (Dehaene et al. 2010). Advances in functional MRI before and after learning to read have revealed that part of the left occipito-temporal pathway dedicated to visual recognition becomes sensitive to letter strings and develops increasingly efficient connections to regions, such as the planum temporale, specialized in the processing of speech sounds (Dehaene et al. 2010; Dehaene-Lambertz et al. 2018; Monzalvo and Dehaene-Lambertz 2013). In the early stages of reading, the learner must effortfully convert each letter or group of letters (called a 'grapheme') into the corresponding sound unit (called a 'phoneme') while 'listening in' to understand the word. With practice, reading becomes automatic, and words can be treated in their entirety, with all letters analyzed in parallel, permitting rapid access to the lexicon.

During the early decoding stages of reading, reading time increases linearly with word length, with a slope as high as ~ 200 ms/letter (Zoccolotti et al. 2005), reflecting a slow process of serial letter-by-letter reading. As reading automatizes, that slope decreases until all words between 3 and 8 letters are read equally fast

(New et al. 2006), except in young dyslexic readers (Zoccolotti et al. 2005). These results fit with a dual-route model of reading: the phonological path allows for a slow decoding of novel words, while the lexical path lends itself to fluent reading of familiar words (Coltheart 2005). Both neural paths exist in the proficient reader, and the goal of reading instruction should therefore be to quickly establish them. According to meta-analyses by the National Reading Panel (Cunningham 2001), the most efficient way to acquire literacy is phonics instruction, in both normally developing readers and children with increased risk for reading deficits.

The language caveat

The facility with which phonics can be acquired is modulated by the language's orthographic transparency. Transparency describes the degree to which grapheme-phoneme correspondences are consistent enough to support fluent reading of novel words, a key determinant of the speed of reading acquisition (Serrano et al. 2011). For example, English and, to a lesser extent, French have many letters that can make many different sounds. These two languages are considered opaque compared to Spanish or Italian—both of which have a highly consistent orthography, meaning that a given letter (or grapheme) has nearly always the same sound. For transparent languages, knowledge of the basic grapheme-phoneme correspondences suffices to read virtually any word. This difference in the transparency of written language influences the duration and difficulty of reading acquisition (Serrano et al. 2011; Ziegler et al. 2010) and the extent of cortex dedicated to the visual component of reading (Paulesu et al. 2000). Yet in spite of this, across all alphabetic languages, learning the grapheme–phoneme correspondences remains the most economical path to reading all words (Vousden et al. 2011).

Principles for phonics instruction

There are several key principles to a successful phonics instruction. Grapheme–phoneme correspondences should be taught explicitly (Castles et al. 2018; Cunningham 2001). The explicit teaching of grapheme–phoneme correspondence should follow a systematic progression (Ehri et al. 2001), meaning that they are taught in a carefully planned sequence, within a rational order that takes into account their frequency and their consistency. Systematic phonics software has been shown to provide better results than non-systematic software (de Graaff et al. 2009).

Practicing lesson-related words should also be a part of any systematic phonics instruction (Mason et al. 1974). Children should be presented with decodable text, i.e., using graphemes whose pronunciation has been previously taught. In spite of adequate phonics instruction, many children show difficulties in applying full decoding skills. These children have no trouble identifying the first letter of a word, but fail to apply their phonics knowledge to all successive letters in such a manner that they combine them into readable syllables. Reading activities that draw attention to the combinations of letters in a word help children apply the alphabetic principle to all letters (McCandliss et al. 2003). Learning to spell, which requires children to focus on each sound and its corresponding letter, also appears to improve decoding skills (Uhry and Shepherd 1993).

Tablet-based interventions for reading instruction

Tablet- or computer-based games may provide an excellent medium for helping children to automatize their grapheme–phoneme decoding skills. A prime cross-language example and an important source of inspiration for the present project is the GraphoGame, a Finnish computer/tablet-based game in which the primary goal is to incite the child to automatize the associations between graphemes and phonemes (Richardson and Lyytinen 2014). Functional brain imaging shows that when preschoolers play the GraphoGame for a few hours, the neural circuits for reading begin to emerge (Brem et al. 2010). Playing the GraphoGame helps children considered at risk for reading acquisition in both English (Kyle et al. 2013) and Finnish (Saine et al. 2011) improve their decoding and encoding skills above their respective control groups.

The ELAN software

Since the GraphoGame was at the time was not available in French, our goal for the ELAN software was to provide similar phonics training for children in France. We also included in ELAN a teaching section that explicitly introduces new grapheme–phoneme combinations in a systematic order, thus allowing children to advance in the game independently of learning in the classroom. ELAN also uses a lexical decision task to measure changes in reading fluency. We now describe these points in turn.

In ELAN, the child must explore 20 different islands where he or she will learn 1 to 5 different grapheme–phonemes relationships. After learning a new grapheme–phoneme lesson, automatization is reinforced through three mini-games that build on skills of syllable and word reading. Passage from one island to the next is allowed once the child has completed all mini-games with a minimum score of 80% correct responses and completed the island "boss". Children receive awards with each game won.

The systematic ordering of grapheme–phoneme correspondences was determined by examining the frequency and consistency of all grapheme–phoneme correspondences in the Manulex corpus, a corpus of 1.9 million words from French children's books (Lété et al. 2004). To further facilitate learning, we made several principled adjustments based on the progression usually taught in phonics manuals and teachers reports. In ELAN, the most frequent vowels are taught first. Regarding consonants, frequent fricatives are instructed before stop consonants due to the greater facility in pronouncing such phonemes for a prolonged duration in isolation and therefore in teaching how they blend into syllables (e.g., 'ffff aaa makes fa'). These adjustments allow ELAN to be used as a practice aid to support general learning in the class, as opposed to replacing methods used by teachers. The appendix presents the full order of grapheme–phonemes taught in ELAN.

Within a given ELAN lesson, a grapheme–phoneme correspondence is introduced in three steps that are commonly used by teachers and supported by research as benefitting phoneme processing and grapheme memorization (Fig. 1). In step 1, the child clicks the grapheme to hear and see a high-quality, noise-free video of an older child pronouncing the corresponding phoneme, with a zoom on the child's mouth and explicit articulation. Enhancing auditory signal-to-noise, for instance, through auditory aids, is known to facilitate reading acquisition (Tallal 2004). Pictures of distinct articulatory gestures underlying each phoneme facilitate the development of early reading (Boyer and Ehri 2011; Castiglioni-Spalten and Ehri 2003). In step 2, when the child clicks on the grapheme, a picture that starts with the phoneme is displayed and described (e.g., 'aaa, aaaple'). This type of exercise relies on the acrophonic principle, which is at the origins of alphabetic writing (each letter used to be the first letter of a common word, e.g., b = "beit" = a house). It is commonly used in the classroom and used here to bind the phoneme to a known word in order to facilitate memorization and to build phonemic awareness of letter sounds in words. The image is there only to ensure understanding of the spoken word. Children can repeat these two steps as much as needed. In step 3, the child traces the grapheme twice 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 are useful activities which have been shown to improve reading skills, probably because they

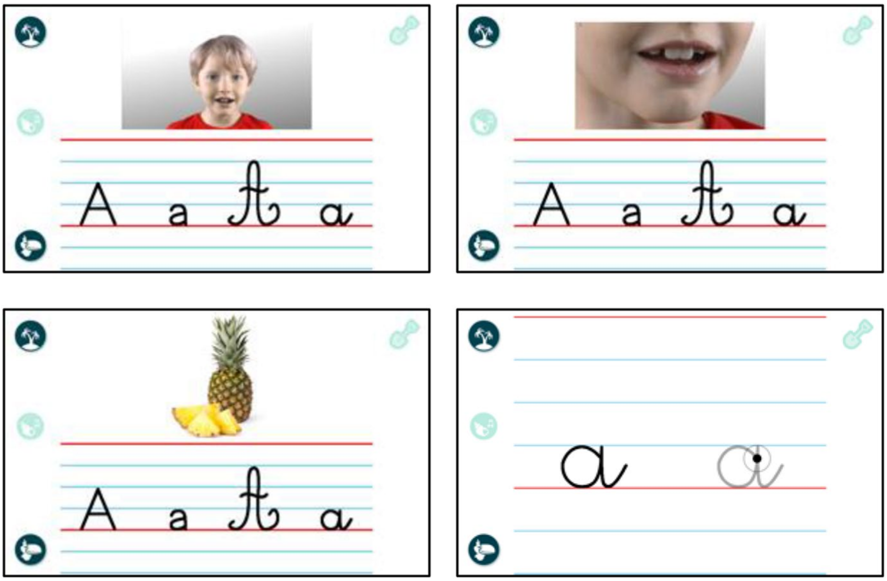

Fig. 1 Grapheme–phoneme learning in the ELAN software. The child is introduced to each grapheme–phoneme correspondence by seeing the grapheme and watching a video of another child pronouncing the sound (top images), seeing the grapheme and hearing the sound in a word with accompanying picture (bottom-left image), and writing the letter on the screen (bottom-right image)

provide an additional motor code to support memory for grapheme (Bara and Gentaz 2011; Bara et al. 2004, 2007, 2016; Longcamp et al. 2005). Because gestures are completely different for visually similar letters such as b and d, tracing may also help surmount the mirror-letter confusions that all children experience, due to the fact that the visual cortex generalizes across viewpoints (Dehaene et al. 2010).

Automatizing grapheme–phoneme association

Once the child has completed the lesson, the next goal is to overlearn the correspondence to the point that it becomes automatic. ELAN uses several mini-games that require children to make faster and faster responses to recently acquired grapheme–phoneme relationships, similar to the GraphoGame (see Fig. 2a). In each

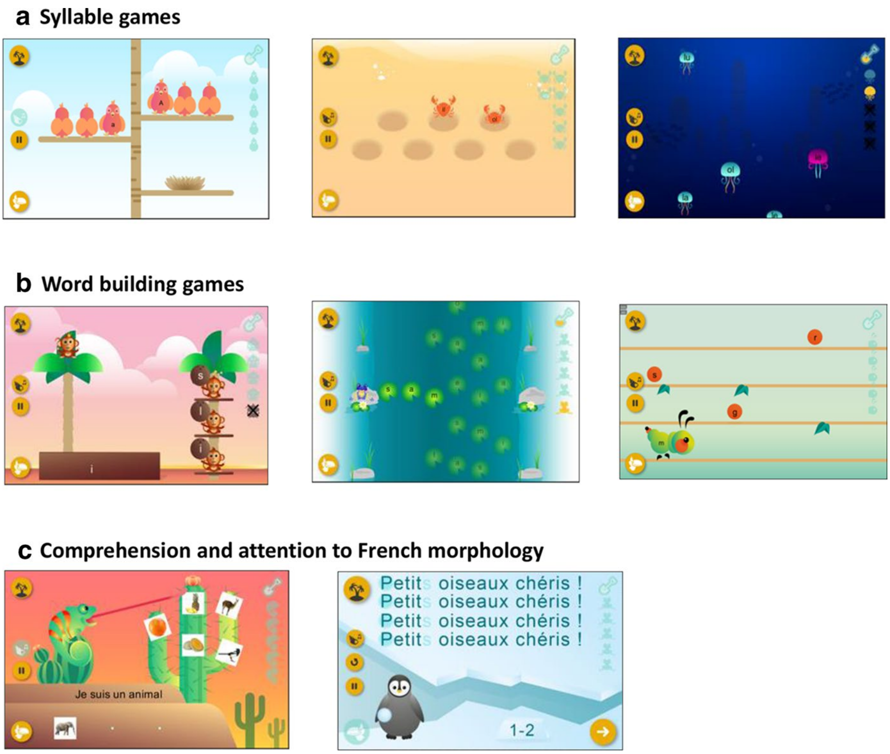

Fig. 2 Screenshots of various games. **a** Letter and syllable games. After learning a given grapheme–phoneme correspondence, children practiced automatizing it by playing various games inspired by the GraphoGame, in which they heard the grapheme within syllables and had to find them written in upper and lower-case letters on screen. The number and type of distractors and their rate of their appearance were automatically adapted to each child's performance. **b** Word building games required children to hear a word and then spell it. Difficulty was adjusted to each child by varying word length and the number and type of distractor letters. **c** Decoding and comprehension games required children to read sentences with 100% decodable text and demonstrate their understanding through an action (e.g., deciding whether the phrase was in the singular or plural)

game, children hear a syllable and must click on the corresponding written stimulus on screen. Syllables can be a single vowel sound, consonant–vowel (CV) or vowel–consonant (VC). Difficulty is created by varying the distractors and the rate of targets and distractors. Distractors can be syllables with the letters reversed, the vowel replaced or the consonant replaced (for example, for the target syllable 'la,' distractors would include 'al,' 'le,' and 'ta').

Once children have learned a grapheme–phoneme pair, they immediately apply this knowledge to reading and spelling small words (Fig. 2b). Finally, the third type of games in ELAN asks children to read short sentences and show their understanding through an action, for example, choosing a picture that corresponds to the written text, or showing the game character the silent letters in a sentence (Fig. 2c). Phonics methods are often criticized for not also helping children develop text comprehension. The goal of these games is to provide periodic comprehension practice using words that respect the phonics progression of the game. The algorithm for automatic word selection in ELAN ensures that all words that are presented for spelling and reading are 100% decodable.

Adapting to the child's level

ELAN's algorithm provides several types of adaptation to each child's level and errors. The local support, common to adaptive gaming, addresses children's punctual errors: if the child makes mistakes, an error bell and loss of a life-point ensue. If the child makes a second error, the game character appears to remind the child of the game objective and highlights the correct response. If the child responds correctly, the highlight is removed for the next question; otherwise, the game ends and another game proposes the same lesson through a different strategy.

Game difficulty is also adaptive. As the child makes more correct responses, the game requires quicker responses. For each game, five levels are set up. These five levels of difficulty control for, according to the mini-game environment, the number of distractors present, the number of locations the child must monitor in order to find the target, and the time the target is visible. Two successive mini-game wins or losses cause the level of difficulty to change. Requiring faster and faster responses in environments of growing difficulty would be difficult to reproduce with traditional pen and paper tests and is intended to promote automaticity.

Finally, the choice of targets and distractors also adapts to each child. The child is required to constantly maintain a performance level superior to 75% correct for all the grapheme–phoneme pairs learned. If across several games, the child makes several mistakes on the same grapheme–phoneme, its score may decrease below this criterion, causing it to become reintroduced as a target. The knowledge score of each grapheme–phoneme pair is based on the last 5 responses to the pair. At the onset of each mini-game, the game software chooses appropriate targets for the child. This pool of syllable or word items is chosen from the current lesson (80% of possible target items), previously learned items (10%) and items containing a grapheme–phoneme correspondence that tipped below the 75% correct response threshold (10%). If all items are above threshold, then random previously learned items are presented.

Using this technique allows for repeated spaced practice, which is known to benefit long-term retention (for reviews, see (Carpenter et al. 2012; Cepeda et al. 2006).

Distractor stimuli are also carefully chosen according to the target and the child's performance history. While only previously learned grapheme–phoneme correspondences can act as distractors (nothing is presented to the child that has not been taught in the game), these are chosen according to their orthographic or phonemic proximity to the target. The visual similarity matrix was adapted from another study of similarity judgments of letter pairs (Boles and Clifford 1989). The phoneme similarity matrix was adapted from French listeners' confusions in discriminating phonemes in CVC pseudo-words presented in natural background noise (Meyer et al. 2013). Only vowels are presented as vowel distractors and consonants as consonant distractors. Half of distractors were chosen based on the letter shape matrix and half based on the phoneme similarity matrix. Initial items are chosen (when possible) to be >0.75 in similarity to the target (1 being the matrix diagonal, and 0 being maximally different items). If the player chooses an erroneous distractor, then the corresponding threshold on distractor similarity is lowered.

Measuring the word length effect

As mentioned in "The reading brain" section, a marker of efficient reading is the disappearance of the word length effect, all written words being ultimately processed equally fast regardless of their length (between ~3 and ~8 letters). The challenge in Elan is to measure how reading is affected by word length without requiring the child to actually read aloud (given that high-quality child-oriented speech recognition was unfortunately not available at the time of programming). In order to measure reading time, ELAN uses a lexical decision game, presented as a "boss" allowing passage from island to island (see "The ELAN software" section), and requiring the children to send 'real words' to a green buoy and 'invented words' to a red buoy. In this game, the child is presented with 16 words and 16 pseudo-words (with a single letter change or inversion), 2–5 letters long, in random order. The challenge must be accomplished in under three minutes (indicated to the child by the sun that crosses the sky). This game thus emphasizes fast word recognition and allows us to measure the identification time for words and pseudo-words. Through this game, we aimed to evaluate the reading slope, i.e., the amount of increase in reading time with each additional letter. As described below, this strategy was successful: identification time initially increased linearly with each additional letter, and that slope decreased with automatization.

Experimental study

The Elan software was designed in several successive stages, each involving piloting with first-grade children as well as adult crash tests. The piloting sessions with children were conducted in classrooms with members of the lab observing the children's interactions of the games. Once the games and their parameters were set up, we tested it in a large test with 1st-grade French schools, contrasting it with both a low-level no-tablet control and with an active control (children playing a previously designed number game, the NumberCatcher). We now describe how this experimental study was organized. The project was approved by the ethical committee CERES (comité d'éthique pour la recherche en santé) on June 28, 2016.

Materials and methods

Participants

Testing of the ELAN software was done in collaboration with the public school district of Poitiers, a region in central-west France. School district personnel in charge of the project were asked to provide approximately 1000 children from classes representing the region. Teachers were approached by district employees, and 53 teachers agreed to participate, for a total of 53 classrooms from 45 schools and 36 towns (ranging in population from ~ 80,000 to ~ 3000 inhabitants). Representative of these demographics, 44 classrooms agreed to integrate our tablet intervention with their reading instruction curriculum, while 9 control classes (group labeled "control") followed a "business-as-usual" curricula for the entire school year. Note that the tablet/no-tablet variable was not randomly assigned, as it was not possible to force teachers to one of these two groups against their will.

Procedure

At the start of the school year, teachers sent home with students a letter explaining the project, accompanied by a 'letter of opposition' for parents who did not wish for their child to participate. In classrooms participating in the tablet intervention, the project followed a randomized crossover design (see Fig. 3). Students were randomly assigned to groups of 3 to 8 children playing the same game at the same time in the classroom's designated 'tablet zone.' The only request made by the lab was that teachers check that groups be heterogeneous in ability. These teacher-determined small groups were then randomized by the lab into either a group that played ELAN for the first half of the year, then the control math game for the second half of the year (group 1, labeled "read/math"), or, conversely, the math game first, then ELAN (group 2, labeled "math/read"). At the start of the year, 975 children were announced by teachers as participating in the project: 417 were assigned to read/math, 357 to math/read, and 201 to control. Whether a group was to play ELAN or the control math game was only announced to teachers on day one of the game intervention.

All children were pretested from September 20 to October 11, 2016, except for 5 children who were tested in the first week of the intervention. This was followed by two weeks of school vacation, and a week to install the application. The two tablet intervention groups then began intervention A as soon as the games were installed. The first period lasted from November 7, 2016, to until the end of the first week of February 2017, for a possible 11 weeks of game play (this excludes the 2 weeks of Christmas break). The first post-test (labeled post-test A) began on January 30,

Session Design

sept.	oct.-jan	feb.	mar.-may	june
pretest	intervention A	post-test A	intervention B	post-test B
	group(1) ELAN & group(2) math		group(1) math & group(2) ELAN	
	group(3) no tablets		group(3) no tablets	

Fig. 3 All children took reading and math tests at the beginning, middle,and end of the year (respectively, termed pretest, post-test A, and post-test B). In a classic crossover design, during the first half of the school year, children in the Intervention groups (group 1 and group 2) played with either ELAN reading software or the math NumberCatcher software. After post-test A, children switched to the other software for the second half of the year. Control classes (group 3) did not have tablets. Children played in small groups using the same game: ELAN (top picture) or NumberCatcher (bottom picture)

2017, and ended on February 18, 2017. After post-test A, children had 2 weeks of winter vacation, then switched games. At this point, schools were sent an updated version of ELAN correcting some bugs. Intervention B commenced as soon as the game was updated and lasted from March 13 to June 9, 2017. This second period was one week longer (excluding the 2 weeks of spring vacation), allowing teachers to make up for lost time due to four 3-day weekends during this period. The second post-test (labeled post-test B) began on June 12, 2017, and ended on June 30, 2017. Teachers were asked that, as much as possible, children be allowed to make up for lost days. Our goal was for each child to play for twenty minutes, three times a week during each intervention period.

The control game used for the intervention was NumberCatcher (www.thenumbercatcher.com). This game, a sequel to a previous game developed by our lab (Number Race: Wilson et al. 2009; Wilson et al. 2006a, b), was created for children between ages 5 and 10. The goal of NumberCatcher is to train basic concepts of number, arithmetic, and base-10 understanding and to cement the links between symbols and quantities by requiring quicker and quicker responses. No research had been conducted on the benefits of the game for strengthening number understanding at the time of the study.

Teachers were instructed not to change their own teaching method in light of the tablet intervention. Adding the tablet atelier was a challenge to incorporate for many teachers, but reports from the academy found this change to be overall positive, as it gave more time to take care of other children during this period. The games were created for autonomous play and teachers did not interact with the children while they played, unless asked for help by the student. All teachers reported that, while a subgroup of children were playing with the tablet, they continued their usual learning program with the rest of the class or provided other specialized work groups. When playing ELAN or NumberCatcher, children worked individually, wearing headphones, but could communicate and help each other since all children in a group played the same game.

Most of the participating classes did not have access to wifi, so a child's progression in the game was stored in the application and only occasionally uploaded to our servers. This required that a child always play on the same tablet, logging in by entering a private code. To share the child's progression data with the lab, teachers were asked to connect to the school wifi and upload the data to the lab server. Periodic reminders were sent to all actors in the project reminding them to upload data.

Pre- and post-tests

Except for a vocabulary test that was only administered at pretest, all tests were given at all three test periods. Forty employees from the Poitiers school district individually tested each child. All tests were done on paper. Twenty-three members of this team had attended a test training day organized by the lab and were in charge of training other testers. These school employees were often in contact with teachers but would only be knowledgeable of a student's group if they purposefully asked, which they were instructed not to do during testing. Given that most children would not yet have had formal reading training at pretest, we emphasized testing known predictive measures as well as measures of reading ability. Although this was not the primary goal of this study, we also included two number knowledge tests to provide a minimal evaluation of the results for our control group.

Passive vocabulary test (30 items) We used a standardized French vocabulary test (TVAP, Deltour and Hupkens 1990). A word was said by the experimenter. The child was asked to choose the best corresponding image from a choice of six pictures. Two points were awarded for the correct response, 1 point for a close response, 0 for all other responses, for a total possible score of 60.

Phoneme (24 items) and syllable suppression (10 items) These two tests were taken from EVALEC, a battery of French predictive reading acquisition measures (Pourcin et al. 2016). The child was asked to suppress the first phoneme/syllable of a pseudo-word pronounced by the experimenter. Ten syllable suppression questions then twelve consonant–vowel–consonant and twelve consonant–consonant–vowel pseudo-words were said to the child. For each type of question, there were two trials with feedback. The child's percentage correct was taken separately for syllable suppression and phoneme suppression.

Letter knowledge (52 items) Children were asked to give the name and sound of all 26 letters. The child's percentage correct out of 52 items was collected. Letters were presented on paper, one-by-one and out of order in lower case. Knowledge of letter sounds is a critical component of early literacy skills and the single best predictor of first-year reading achievement (Neuman and Roskos 1998).

One-minute word and pseudo-word reading (2 × 1 min reading) The child was asked to read a maximum number of words followed by a maximum number of pseudo-words, each in one minute (Sprenger-Charolles et al. 2017). Children were presented with 35 words and 30 pseudo-words at pretest and post-test A. Sixty words were presented at post-test B. Words and pseudo-words were all either mono- or bi-syllabic. All the presented words were frequent and taken from the Manulex children's word database (Lété et al. 2004). Pseudo-words were matched to the words in syllable structure and orthographic difficulty. If a child was unable to do these two tests, reading less than four words from the first ten words, then he or she was not asked to continue to the more difficult reading tests that followed. To stabilize the variance, children's mean combined score for these two tests was analyzed, after checking that they provided consistent results. This combined score from both tests provides us with a measure of reading fluency, in other words, how well the child can read familiar words as well as their ability to decode. Combining these two test is a main early predictor of later reading comprehension (Gentaz et al. 2015).

Non-sense text reading The child was asked to read a text of 265 words, all of which are frequent. The text, however, is non-sense and thus requires that the child avoid reading words by context. The text, known as the Alouette test, has been standardized with French primary school children and is used to evaluate reading age, diagnose dyslexia (Lefavrais 2005), and predict future reading difficulties (Bertrand et al. 2010; Cavalli et al. 2017).

Sentence comprehension (8 items) Eight items were adapted from the French standardized ECOSSE test and Reading Evaluation Test (Sprenger-Charolles et al. 2017; Lecocq 1996). The child had to read aloud a sentence such as "the man is eating an apple," then choose the correct corresponding image from a choice of four pictures including semantic and syntactic distractors. Items were changed from post-test A to post-test B. The sentences in this test were in order of increasing syntactic difficulty. The percentage correct score was calculated for each child.

Items from pre and post-test A were not changed as 57% of children did not attempt this test and the average score was below 1 response correct at pretest.

Symbolic number comparison (number of CR in 30 s) In the SYMP test of number comparison, children were asked to cross out the numerically larger of two Arabic one- digit numbers in thirty seconds. This symbolic number comparison test is correlated with math achievement from first to sixth grade in school (Brankaer et al. 2017). 56 pairs of numbers were presented on a page. Four trials with feedback were provided before beginning. Children did the same test at each of the three tests. The maximum number of correct responses was collected from each child.

Single-digit subtraction (12 items) Single-digit subtraction was used instead of addition, because subtraction provided a better measure of a child's ability to do arithmetic as answers are not typically memorized as with the addition table. Children saw the subtractions, presented one-by-one in p–q format, and heard them orally. There were 18 questions in total: six − 1 subtrahend questions, six small subtrahend questions (− 2, − 3, − 4), and six large subtrahend questions (− 6, − 7, − 8). The test was stopped if the child had less than three correct answers on the first 8 questions. Item order was changed from post-test A to post-test B. The percentage correct score was calculated for each child.

Participant attrition

The goal of this project was to evaluate the possible benefits in reading acquisition for children using the ELAN software. To measure this with confidence, we had set several requirements for inclusion in the research analysis. The first requirement was that all participants complete all three tests so as to ensure an accurate comparison of progression made, given that all children would undoubtedly progress thanks to normal class training. With each test, we lost a number of children due to absence on the testing day, children leaving schools, or testers forgetting to report id information on the tests. From an original count of 975 participants, 728 children completed all three tests (910, pretest; 867, post-test A and 728 post-test B). The large loss between post-test A and post-test B was due to two classes that left the project and three classes that used incorrect post-test B tests, they had re-used post-test A tests.

We also lost participants due to a software bug that damaged datafiles, causing progression data to be lost and sending many players back to square one in the game. If the child reported to the teacher this anomaly, we were able reinitialize the child's game. Unfortunately, not all children reported losing their progression and this problem was not always detected quickly enough. In order to keep a participant in our analysis, we required a normal game progression, meaning the following: (1) the child had spent sufficient time using the game (the threshold criteria was set at a minimum of 330 min of play) and (2) progression had been unhampered by problems with the software. The number of participants remaining from the intervention groups after these criteria was 350 (113 read/math; 103 math/read).

To examine the baseline balance between the two intervention and control groups, we compared the three groups, two-by-two, in terms of baseline demographic variables and a composite z-score of the four predictive reading acquisition tasks at pretest (vocabulary, syllable awareness, phoneme awareness, and letter knowledge). An ANOVA of this composite score revealed no significant difference between the two intervention groups [$F(1, 214)=0.55$, $p=0.46$]. A significant difference was however found between the read/math and control groups [$F(1, 245)=5.70$, $p=0.02$], but not between the math/read and control group [$F(1, 235)=2.47$, $p=0.12$]. Seven children with the lowest composite predictor scores in the control group were removed to ensure equal groups, see Table 1 for baseline information of the three groups.

Statistical analysis

Given our 2-period crossover design, for each task, we first analyzed responses from the children randomized into the read/math and math/read groups in a mixed analysis of variance with factors of intervention group (2 levels: read/math versus math/read) and test period (3 levels: pretest, post-test A, post-test B), corresponding to the randomized part of the experimental design. We then did a larger mixed analysis of variance with 3 levels for the group factors (the 2 intervention groups plus the control group) and the same test period factor. For the tests where reading was required by the child, to factor out pre-schooling differences amongst children, we added a covariable composite z-score computed from the individual's average z-scores on the four predictive reading acquisition tasks at pretest (vocabulary, syllable awareness, phoneme awareness, and letter knowledge). As shown in Table 2, this covariable was useful since there were significant correlations between the predictive reading acquisition tests and the reading tests.

In a crossover test, we expected to see a 2-step pattern of improvement: children in the read/math group should improve in reading tests at post-test A relative to the

Table 1 Baseline characteristics at pretest

Mean (SD)	Read/math	Math/read	Control	Group difference
N	113	103	127	
N = Male	61	53	63	$X^2(2)=.46$, $p=0.80$
Age in months at pretest	75.8 (3.5)	75.7 (3.8)	75.6 (3.5)	$F(2, 340)=.287$, $p=0.75$
Time using ELAN (min)	528 (114)	508 (139)	–	$F(1, 214)=1.36$, $p=0.25$
Vocabulary, $x=60$	41.06 (5.25)	39.60 (5.57)	40.50 (5.80)	$F(2, 340)=1.89$, $p=0.15$
Syllable suppression (%)	52% (41%)	55% (40%)	48% (38%)	$F(2, 340)=1.14$, $p=0.32$
Phoneme suppression (%)	46% (28%)	42% (29%)	39% (30%)	$F(2, 340)=1.80$, $p=0.17$
Letter knowledge (%)	73% (18%)	74% (19%)	71% (20%)	$F(2, 340)=.979$, $p=0.38$
Composite predictor	0.08 (0.68)	0.01 (0.73)	− 0.07 (0.77)	$F(2, 340)=1.20$, $p=0.30$

Table 2 Correlations (r^2) between reading predictors at pretest and the two post-tests

	Predictors at pretest			
	Vocabulary	Syllable suppression	phoneme suppression	Letter knowledge
Post-test A				
Syllable suppression	0.07	0.20	0.14	0.08
Phoneme suppression	0.06	0.25	0.37	0.23
Letter knowledge	0.04	0.10	0.16	0.22
Word and pseudo-word reading	0.04	0.21	0.32	0.23
Non-sense text reading	0.05	0.17	0.21	0.14
Sentence comprehension	0.04	0.07	0.11	0.05
Post-test B				
Syllable suppression	0.08	0.14	0.09	0.10
Phoneme suppression	0.07	0.19	0.24	0.17
Letter knowledge	0.08	0.10	0.12	0.20
Word and pseudo-word reading	0.02	0.16	0.19	0.14
Non-sense text reading	0.03	0.11	0.12	0.08
Sentence comprehension	0.06	0.04	0.07	0.04

(df = 341), r^2 scores superior to 0.03 are significant at $p < 0.001$

math/read intervention and control groups; reading scores should increase for the math/read group relative to the control group at post-test b. An effect of the training is therefore expressed as a significant interaction between the within-participant factor of test time point and the between-participant factor of group. Given a significant interaction effect, post hoc tests were done at each test period to untangle changes in progress made by each group.

Results

Figures 4 shows the evolution of performance across the three test periods for each of the read/math, math/read, and control groups, respectively, for the reading and math tests.

Syllable suppression

The analysis was done on the percentage of correct responses from the 10 items. The intervention group ANOVA revealed, as expected, a significant main effect of test session [$F(2, 428) = 94.8$, $p < 0.001$], consistent with an improvement over time. No group × test interaction was found [$F(2, 428) = 0.35$, $p = 0.705$]. Similar results were found when including the control group in our analysis: there was a significant main effect of test session [$F(2, 680) = 176$, $p < 0.001$], but no group × test interaction [$F(4, 680) = 1.59$, $p = 0.176$].

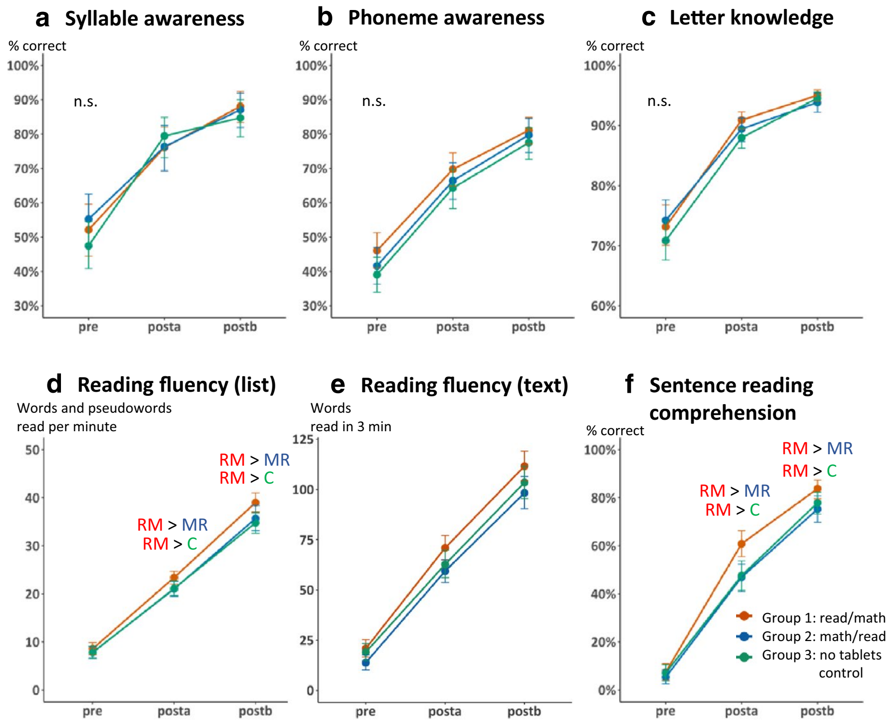

Fig. 4 **a**–**f** Performance improvements pertaining to reading acquisition. Panels indicate performance in tests of **a** syllable awareness; **b** phoneme awareness (combined score of CV and CVV suppression in pseudo-words); **c** letter knowledge (combined score of letter name and sound knowledge); **d** word and pseudo-word reading (average number of items read in one minute); **e** non-sense text reading (total number of words correctly read in a non-sense text within a 3-minute time limit); **f** sentence comprehension (reading a short sentence and choosing the correct image out of 4 semantically or syntactically related pictures). All tests improved across time. The p-value indicates the significance of the 3 × 3 group × time interaction. Pairwise differences that were significant at $p < 0.05$ are highlighted (*RM* read/math group; *MR* math/read group; *C* control group)

Phoneme suppression

The analysis was done on the percentage of correct responses combining both CV and CCV phoneme suppression tests, including a total of 24 questions. Again, the intervention group ANOVA showed a significant main effect of test session [$F(2, 428) = 230$, $p < 0.001$]. All children progressed equally in the task as was made evident by the absence of a group × test interaction [$F(2, 428) = 0.415$, $p = 0.66$]. Similar results were found when including the control group in our analysis: there was a significant main effect of test session [$F(2, 680) = 390$, $p < 0.001$], but no group × test interaction [$F(4, 680) = 0.333$, $p = 0.856$].

Letter knowledge

The percentage correct for the child's combined letter name and letter sound score was collected in this test. The intervention group ANOVA showed a significant main effect of test session, [$F(2, 428) = 286$, $p < 0.001$], but no interaction between group and test period [$F(2, 428) = 1.16$, $p = 0.316$]. Similar results were found when adding the control group: all three groups made progress over time, as shown by a main effect of test session [$F(2, 680) = 441$, $p < 0.001$], in the absence of an interaction [$F(4, 680) = 1.44$, $p = 0.22$].

One-minute word and pseudo-word reading

We analyzed the mean score for word and pseudo-word reading combined. We were notified by testers that on several occasions the one-minute time limit was not observed. These instances were signaled on the tests and the scores of these children removed from analysis, leaving a smaller number of participants per group (read/math, $n = 99$; math/read, $n = 94$; control, $n = 123$). Both groups in the intervention group analysis improved over the three test periods, as revealed by a main effect of test period [$F(2, 382) = 1304$, $p < 0.001$]. An effect of the reading predictor covariate was also significant [$F(1, 190) = 81.37$, $p < 0.001$]. The group × test interaction approached significance [$F(2, 382) = 2.61$, $p = 0.075$]. When adding the control group, significant effects were found for the main effect of test period [$F(2, 626) = 2164$, $p < 0.001$], the composite reading predictor [$F(1, 312) = 178$, $p < 0.001$), and, most importantly, the group × test interaction $F(4, 626) = 2.96$, $p = 0.019$. A post hoc ANOVA at pretest showed no difference between the three groups at pretest $F(2, 313) = 0.451$, $p = 0.637$. An ANCOVA, using the predictive measures covariable, showed that at post-test A the general progress made by all three groups was no longer equal [$F(2, 312) = 4.44$, $p < 0.05$] with the read/math group reading significantly more items per minute than the math/read group [$F(1,190) = 7.19$, $p < 0.01$] and the control group [$F(1, 219) = 6.89$, $p < 0.01$]. In quantitative terms, after 3 ½ months of game play, the group that had used ELAN improved by 0.26 standard deviations over the control group, and 0.27 standard deviations over the math software group.

The advantage obtained by the read/math group continued into post-test B [$F(2, 312) = 4.61$, $p = 0.01$], with the read/math group continuing to read more items on average than the math/read [$F(1, 190) = 4.55$, $p = 0.03$] and the control group [$F(1, 219) = 9.60$, $p < 0.01$]. In quantitative terms, at the end of the school year, the group that had used ELAN in the first session, after 4 months of no longer playing ELAN, maintained an advantage of 0.34 standard deviations over the control group, and a 0.27 standard deviation advantage over the math software group.

Non-sense text reading

Children that were unable to read more than four words of the first ten words in the one-minute word reading task were not asked to continue with the more difficult text reading tasks. It was also reported that while some children read more than four words on the one-minute reading task, this was done with great difficulty and the child refused to continue to the more difficult reading tasks. As a consequence, the number of children taking the following reading tests were smaller (pretest: read/math = 70, math/read = 45, control = 64; post-test A: read/math = 112, math/read = 98, control = 121; post-test B: read/math = 113, math/read = 100, control = 126). Children unable to do these tests were scored at ‘0’ in our analysis.

Analyses of the intervention group ANCOVA revealed a significant main effect of improvement across the three tests [$F(2, 428) = 745$, $p < 0.001$] and an effect of the composite pretest predictor [$F(1, 213) = 42.9$, $p < 0.001$]. No group × test interaction was found [$F(2, 428) = 2.39$, $p = 0.09$]. When including the control group, again, there was a main of effect of test period [$F(2, 680) = 1182$, $p < 0.001$], and a significant effect of the composite pretest predictor [$F(1, 339) = 116$, $p < 0.001$], but no group × test interaction [$F(4, 680) = 1.69$ $p = 0.15$].

Sentence reading comprehension

In this test, children read 8 sentences and, for each, selected the picture depicting their meaning among 4 possible choices. The intervention group ANCOVA showed a significant main effect of test period [$F(2, 428) = 596$, $p < 0.001$] and a significant effect of the composite pretest predictor [$F(1, 213) = 44.38$, $p < 0.001$]. Most importantly, a significant group × test period interaction was also found [$F(2, 428) = 3.75$, $p = 0.02$]. Post hoc tests at each time point revealed no difference between the two intervention groups at pretest [$F(1, 214) = 0.74$, $p = 0.39$], but a significant difference at post-test A [$F(1, 214) = 11.49$, $p < 0.001$] and post-test B [$F(1, 214) = 6.41$, $p = 0.012$].

Similarly, the ANCOVA including the control group also revealed a significant main effect of test [$F(2, 680) = 880$, $p < 0.001$], an effect of the composite pretest predictor [$F(1, 339) = 95.95$, $p < 0.001$], and a group × test interaction [$F(4, 680) = 2.85$, $p = .023$]. A post hoc ANOVA at pretest showed no difference between the groups at the start of the year [$F(2, 339) = 0.44$, $p = 0.65$]. By post-test A, however, the general progress made by all three groups was no longer equal, [$F(2, 339) = 7.92$, $p < 0.001$] with the read/math group answering more questions correct than either the math/read group [$F(1, 213) = 12.34$, $p < 0.001$] or the control group [$F(1, 237) = 11.75$, $p < 0.001$]. No difference was found between the math/read and control groups [$F(1, 227) = 0.03$, $p = 0.87$]. In quantitative terms, after 3 ½ months of game play, the group that had used ELAN improved by .39 standard deviations over the control group, and .41 standard deviations over the math software group.

The group differences continued into the post-test B [$F(2, 339) = 3.49$, $p = 0.032$), with the read/math group continuing to answer more items correct than either the math/read [$F(1, 213) = 6.96$, $p < 0.01$) or the control group [$F(1, 237) = 3.92$,

$p < 0.05$]. No difference was found between the children that used ELAN in the second period and the control group [$F(1, 227) = 0.55$, $p = 0.46$]. At the end of the school year, the group that had used ELAN in the first session, after 4 months of no longer playing ELAN, maintained an advantage of 0.22 standard deviations over the control group and a 0.31 standard deviations advantage over the math software group.

Measuring the evolution of the word length effect

ELAN was developed to help automatize the child's knowledge of grapheme–phoneme correspondences and word reading. The game pushes the child to respond faster by increasing game difficulty and requiring faster and faster responses. As mentioned in section "Measuring the word length effect", we used a lexical decision task at the end of each game level to evaluate how reading speed varied with word length.

Our main goal was to measure if the child's word reading speed became progressively less affected by word length over the course of the game, in other words, whether ELAN could be used to measure when students move from slow decoding to parallel word reading. Therefore, we only looked at data from children that had passed at least the first 4 game levels, meaning that these children had learned all the vowels, fricative graphemes, and the most common occlusive consonants. Twenty-one children from the read/math group and 44 children from the math/read group were included in the following analysis. The software measured the time elapsed between the presentation of a target word or pseudo-word and the child's decision. We removed outlier trials with response times below 325 ms (less than 1% of trials) or above a fixed cutoff, defined for a given cell of the design as being 2 standard deviations from the mean of the within-participants medians in that cell (6% of trials). The medians of the remaining correct response times were entered into a mixed ANOVA with participant as the random factor, intervention as the between factor and word length (number of letters, from 2 to 5), test number (test numbers 1 through 4), and lexical status (word or pseudo-word) as within factors. All effects are reported as significant at $p < 0.05$. Degrees of freedom are adjusted to reflect the fact that some cells had missing data for some participants.

The mean of the median response times for each intervention group at each test period and for all lengths is shown in Fig. 5. There was a significant length effect: as expected, response times increased as the number of letters increased from 2 to 5 [$F(3, 175) = 190.21$, $p < 0.001$]. We also found a main effect of lexical status [$F(1, 51) = 243.81$, $p < 0.001$] indicating that responses were overall faster for words than for pseudo-words, and a length × lexical status interaction [$F(3, 180) = 35.39$, $p < 0.001$], due to the fact that the slope of the length effect was shallower for words than for pseudo-words. Finally, all of these effects were qualified by a length × test number [$F(9, 557) = 13.06$, $p < 0.001$] and a triple interaction of length × test number × lexical status [$F(9, 543) = 7.12$, $p < 0.001$]. As seen in Fig. 5, over the four successive tests, the length effect became shallower for words (length × test number interaction restricted to words only [F(9,

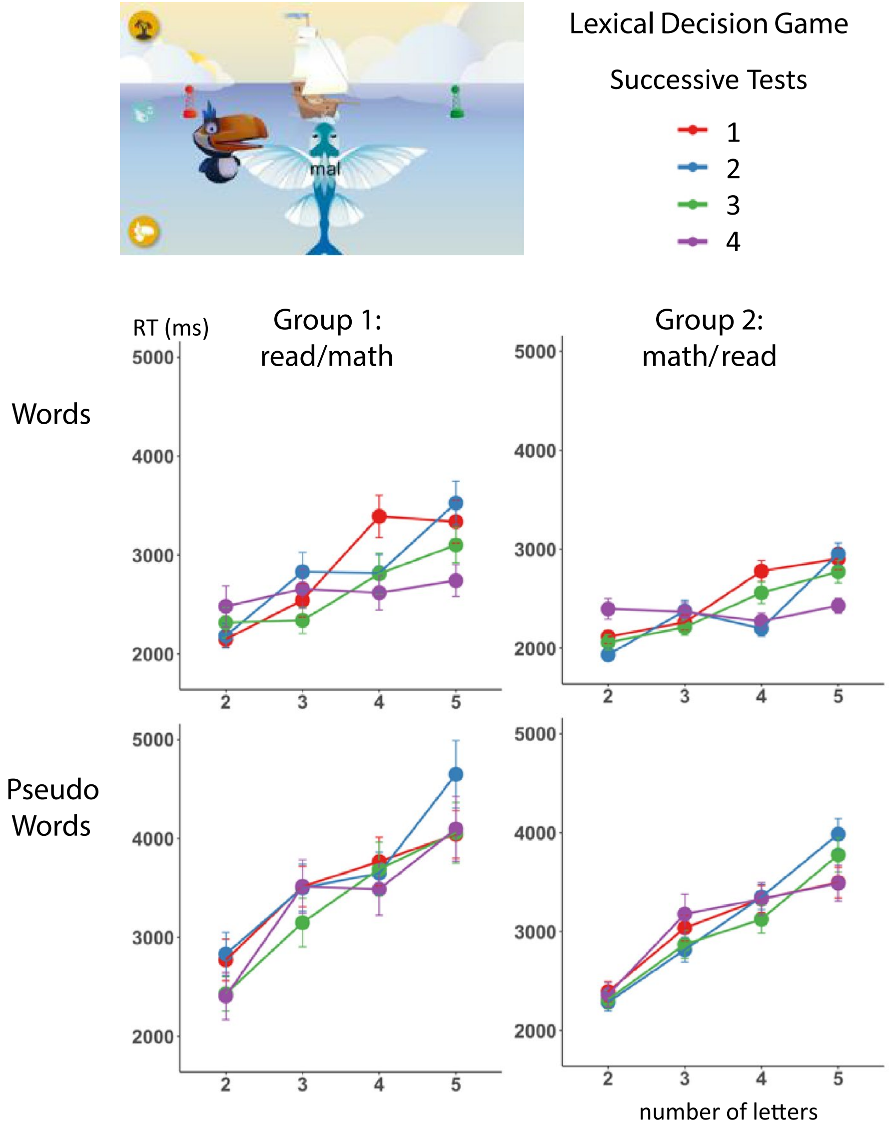

Fig. 5 Measuring the speed of reading and its improvement over time using a simple lexical decision game. Passing the lexical decision task was required at the end of each level to proceed to the next. Children had to decide if a word presented was either a real word or an invented word. Thirty-two words were presented: four words of 2, 3, 4, or 5 letters for both words and pseudo-words. All of the words were decodable from the lessons completed by the child. We measured reading response time by word length for both words and pseudo-words of the two different groups. Group 1 played ELAN during intervention A and Group 2 played ELAN during intervention B

548) = 21.15, $p < 0.001$], whereas this improvement was also present but less pronounced for pseudo-words [$F(9, 562) = 4.79$, $p < 0.001$]. In agreement with this interpretation, at test number 4, the length effect remained significant for pseudo-words [$F(3, 188) = 36.63$, $p < 0.001$) but no longer for words [$F(3, 178) = 0.95$, n.s.) (Fig. 5).

All of the above effects were true within both the math/read and the read/math groups. There were no interactions involving this intervention factor. Only a main effect of intervention was observed [$F(1, 51) = 5.52$, $p = 0.03$), reflecting the fact that the math/read group responded faster, in line with the fact that these children took the tests later in the year than the read/math group.

In summary, the lexical decision task concurred with paper tests in indicating that subjects in group read/math, who started with ELAN software, gained in reading acquisition. Superficially, the absence of any interaction with intervention in the lexical decision task seemed to suggest that the intervention did not have an effect on the evolution of the word length effect, and that the improvements were solely due to time elapsed at school. However, it should be remembered that (for this lexical decision test only) the children in the read/math group were tested within the first half of the year. Thus, their fast progression and the disappearance of the word length effect by test 4 are remarkable and do suggest that the software could have had an effect over and above the improvements due to schooling itself.

To evaluate this, we compared the performance of the read/math group at test 4 with that of the math/read group at test 1. The rationale was that if effects were due solely to time at school, the math/read group should exhibit a better performance on its test 1 (because it was taken later in the year and used grapheme–phoneme correspondences that children would already be very familiar with), whereas if the software intervention was the primary driver of performance, then the converse should hold.

We entered the corresponding median response times into a mixed ANOVA with subject as the random factor, intervention as a 2-level between factor and word length (number of letters, from 2 to 5) and lexical status (word or pseudo-word) as within factors. There were main effects of word length $F(3, 186) = 68.72$, $p < 0.001$ and lexical status $F(1, 60) = 96.70$, $p < 0.001$. Most importantly, the length × intervention was significant [$F(3, 186) = 3.50$, $p = 0.017$], confirming that, indeed, the read/math group, although tested at an earlier time, had a shallower length effect than the math/read group. Furthermore, the triple interaction intervention × length × lexical status was significant [$F(3, 179) = 6.14$, $p = 0.001$]. This was due to the fact that, with words, intervention reduced the length effect down to non-significance only for the read/math group (length × intervention interaction, restricted to words: [$F(3, 180) = 5.46$, $p = 0.001$]), and also reduced the impact of pseudo-word length [$F(3, 188) = 2.76$, $p = 0.04$]). In detail, for the read/math group, no effect of word length was found when reading words on test 4 [$F(3, 53) = .94$, n.s.], while the length remained significant for pseudo-word reading [$F(3, 59) = 18.19$, $p < 0.001$]. For the math/read group, a significant effect of word length was found at test 1 for words [$F(3, 127) = 36.67$, $p < 0.001$] and for pseudo-words $F(3, 129) = 44.36$, $p < 0.001$.

In summary, these results demonstrate that the children who used ELAN in the first session ceased to be affected by word length at an earlier time than children in the math/read group. These results lend support to ELAN as a catalyst to parallel reading.

Number comparison

The intervention group analysis on the number of comparison pairs correctly solved in one minute revealed a main effect of test period [$F(2, 428) = 116$, $p < 0.001$], but no significant group × test interaction [$F(2, 428) = 0.92$, $p = 0.40$]. The 3-group analysis yielded a significant interaction, however: [$F(4, 680) = 3.25$, $p = 0.01$]. Post hoc analysis revealed no difference between the 3 groups at pretest [$F(2, 340) = 0.32$, $p = 0.73$]. At post-test A, there continued to be no difference between the two intervention groups [$F(1, 214) = .14$, $p = 0.71$], but both the read/math and the math/read were superior to the control group [respectively, $F(1, 238) = 9.45$, $p < 0.01$; and $F(1, 228) = 12.54$, $p < 0.001$]. At post-test B, there continued to be no difference between the two intervention groups [$F(1, 214) = 0.54$, $p = 0.46$], but the differences relative to the control group remained [respectively, $F(1, 238) = 9.28$, $p < 0.001$; and $F(1, 228) = 5.79$, $p = 0.02$]. These results point to a possible tablet effect as all children that used the tablets significantly improved over the no-tablet control group (Fig. 6).

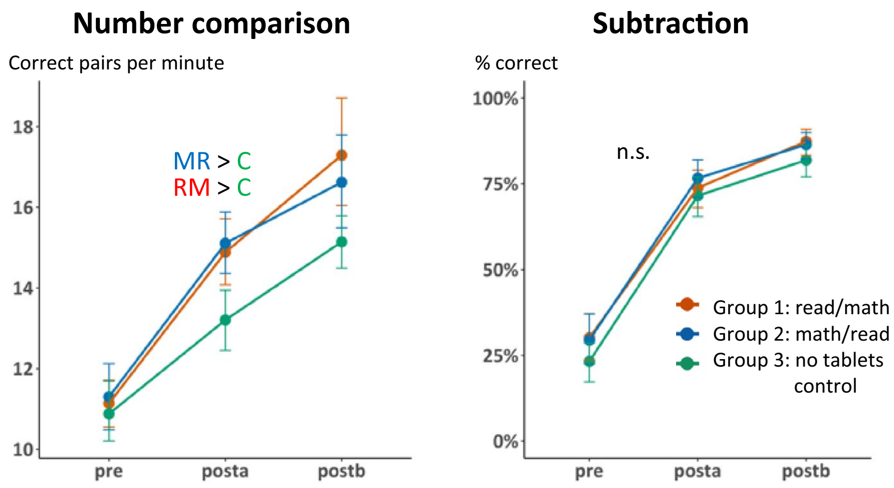

Fig. 6 Performance improvements in numerical tests. Two tests were used to measure possible effects of the control game intervention: **a** Symbolic number comparison (% correct in deciding which of two Arabic numerals is larger); **b** Subtraction (% correct). Same format as Fig. 4. All tests improved across time. The *p*-value indicates the significance of the 3×3 group × time interaction. Pairwise differences that were significant at $p < 0.05$ are highlighted (*RM* read/math group, *MR* math/read group, *C* control group)

Subtraction

The child's percentage of correct responses in subtraction was entered into an intervention group ANOVA, which revealed a significant main effect of the test period [$F(2, 428)=253$, $p<0.001$], but no group × test interaction [$F(2, 428)=0.314$, $p=0.73$]. In the 3-group analysis, the main effect of test was significant [$F(2, 680)=417$, $p<0.001$], but again, no group test interaction was found [$F(4, 680)=0.27$, $p=0.90$].

Discussion

Summary of results

The progress of the read/math group at post-test A, which continued to post-test B on the one-minute reading test and sentence comprehension, coupled with the evidence from the lexical decision task for a reduced length effect for this group gives strong support for ELAN as a tool to aiding classroom reading instruction and accelerating first-grade reading acquisition. The observed effect sizes of 0.26–0.41 are considered as small to medium, but more importantly, the benefits seem to extend four months beyond the training period. Interestingly, these results pertain to those children that only used the software in the initial learning stages of reading. Given the short duration of the training period (read/math x = 8.8 h, SD 1.9; math/read x = 8.5 h, SD 2.3), it is likely that improved results would be obtained with prolonged and/or more intense exposure.

Our results are consistent with previous studies showing that early explicit systematic phonics provides the fastest route to reading acquisition (Cunningham 2001; Ehri et al. 2001).This study also gives strength to a recent observational report of 2500 first-grade students in France, which showed that reading ability improved faster in classes where a larger number of grapheme–phoneme correspondences were taught early in the year (Roland Goigoux 2016). Phonics methods are often criticized for not supporting comprehension due to the limiting constraints of word choice during the early lessons in methods that strictly adhere to 100% decodable text. In this context, however, it is particularly important to emphasize that sentence reading comprehension improved in the read/math group relative to both active and no-tablet control groups, even though our software was entirely focused on grapheme–phoneme teaching and 100% decoding practice. While we cannot disentangle the benefits of using ELAN from other instructional materials used by the participating teachers, this aspect of our results would seem to support the early use of phonics to help children improve access to comprehension. It is compatible with the *Simple View of Reading* formula that defines reading comprehension as the product of decoding and oral comprehension (Gough and Tunmer 1986; Lervåg et al. 2018). In other words, once a text is decoded, the reader uses the same mechanisms that he or she would apply to hearing language. By facilitating a faster automatization of decoding skills, the reader can focus on what they have read. Early intense practice of phonics may be key to alleviating the difficulty of the phonological path of slow decoding so that the child can quickly access the lexical path of fluent reading for understanding. The goal for developing ELAN was, in fact, that the software help automatize the rote task of grapheme–phoneme correspondence memorization and allow teachers to focus on developing oral comprehension and reading aloud.

While it is disappointing that the NumberCatcher game did not have a more specific influence on number comparison and had no detectable effect on subtraction abilities, the results suggest instead that tablet software itself may have a generic effect on speeded decision tasks, as both tablet groups improved in the number comparison test. Numerous studies on 'first-person shooter games,' video games that require the player to make combative actions from a first person perspective, have revealed improvements in a number of cognitive functions, including reaction time to 'go signals' (Colzato et al. 2013) and use of attentional resources in simple tasks of flanker compatibility and spatial localization (Green and Bavelier 2003, 2006). While neither ELAN nor Number Catcher would be classified as 'action games,' they both require players to make faster and faster actions in environments with many distractors. Thus, while our isolated result should be taken with caution, given that the benefit in number comparison was only obtained relative to a non-randomized no-tablet control group, future study of more general cognitive benefits of educational games that require timed responses in noisy environments is warranted.

Limitations

The two main limitations to this project can be attributed to the software design. Nine hundred and seventy-five children were announced as participating in the study. Of the 728 students that completed the three tests, minor software bugs complicated the large-scale deployment of the software to such an extent that only 47% of them experienced a game intervention adapted to their progress and level. While it is virtually impossible to create bug-free software without broad field testing, improved piloting was warranted in this case. Children were sometimes stymied from advancing in the game due to poor ergonomics that were not discovered until a close examination of the children's game progression.

A second limitation to this study concerns the absence of results with students that used ELAN during intervention B. ELAN adapts to the child's response rate, but requires completion of all lessons to move through the game. It is highly likely, given the improved reading scores for all children at post-test A, that for the math/read group, the grapheme–phoneme correspondences presented during the second session were mainly a review of already well-learned material in the class. Ideally, ELAN should use an adaptive mechanism that not only provides tailored game difficulty, but adjusts more quickly in order to specifically train the grapheme–phoneme correspondences that the child has not learned or needs to practice.

A third limitation to this study is that we measured the benefits of autonomous use of ELAN without taking into account other factors known to influence reading acquisition, such as socio-economic status on academic achievement (for a review of the effects of SES on academic achievement: (Sirin 2005) and the effects of materials and methods already in use by the participating teachers (Braibant and Gerard 1996). The benefits of ELAN might have been partially dependent on the activities already in place by the teacher. For example, it is unknown to us which teachers were already using systematic phonics methods and how many grapheme–phoneme correspondences were taught in the class. If ELAN were to be used in the classroom, it would be important to know if the software itself is sufficient in phonics teaching or if it should be used to provide repetition of the teacher's lesson.

Conclusion

In a recent review on the science of learning to read, it has been proposed that the "reading wars" debate between whole-word methods and phonics needs to end (Castles et al. 2018). Enough evidence has been accumulated to support explicit systematic phonics and our focus should be on developing guiding principles for "balanced instruction" that acknowledge the complexities of learning that underlie the individual child's progress from decoder to reading expert. This requires combining an understanding of the biological and cognitive aspects of learning to read, respecting the evidence-based practices already known and furthering research on the bricks of best practice that build a reading method from start to finish. In the case of this project, it appears that intense and focused learning of decoding skills may be more effective when taught early in learning. We also demonstrate that software that focuses on the repetition of decoding skills, and that adapts to the child's level, can indeed lead to beneficial effects for young readers. More importantly, this work builds onto the evidence that addressing 'which' strategies and 'when' should now be the focus of reading research.

Acknowledgements Funding for this work was provided by Bpifrance, the public French bank under the mandate of the Government under the Programme Investissement d'Avenir (investing for the future). The authors would like to thank many other people who contributed to this project: Lilliane Sprenger-Charolles, Gaëlle Olivier, Thibault Petit, Stéphanie Philippe, Bruce McCandliss, and Isabelle Denghien. We are exceptionally grateful to our partner the Academy of Poitiers, in particular, Dominique Quéré (Délégué Académique au Numérique), Cedric Couvrat (Délégué académique au numérique Adjoint pour le 1er degré), and the school district team that tested and accompanied the participating classes. A very special thanks to the teachers who participated in this project and their pupils and families.

Compliance with ethical standards

Conflict of interest We wish to confirm that there are no known conflicts of interest associated with this publication. We also wish to draw the attention of the Editor to the following facts: A 'Cooperation Accord' was signed between Manzalab, the labs and the BpiFrance to finance the development and testing of the ELAN software. Manzalab is a private company specializing in serious games, whose role was to develop the software ELAN as a serious game and, if so desired, commercialize the game after testing. The principle roles of the two participating labs were to oversee adherence by the game software to principles from cognitive neuroscience and education research, test game ergonomics and test the ELAN software in a randomized control intervention study. Neither research labs received payment from Manzalab for their involvement in the project, nor are they privy to any financial gain from the dissemination or licensing of the application ELAN. Research in this article was strictly conducted between INSERM and the school district of Poitiers, France. Testers from the school were blind to the intervention and progress made by students, while the researchers in the lab did not participate in testing. Only the 'global results' from the student's groups has been shared with Manzalab. We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing, we confirm that we have followed the regulations of our institutions concerning intellectual property.

Appendix

See Table 3.

Table 3 Order of grapheme–phonemes taught in ELAN. Phoneme code used from the MANULEX (Lété et al. 2004)

Lesson number	Grapheme orthography	Phoneme code	Lesson number	Grapheme orthography	Phoneme code (Manulex)
1	A, a	a	31	et, est	e,E
2	E, é	e	32	ent	silent
3	I, i	i	33	an, am	@
4	O, o	o	34	en, em	@
5	U, u	y	35	on, om	§
6	E, e	*	36	in, im, ym	5
7	Y, y	i	37	à, â	a
8	L, l	l	38	au, eau	O
9	M, m	m	39	è, ê, ë	E
10	S, s	s	40	oeu	9
11	R,r	R	41	gn	N
12	Ou	u	42	eConsonantConsonant	E
13	silent letters (e, s, t, x)	silent	43	ai, ei	E
14	Jj,	Z	44	ez er	e
15	F,f	f	45	et, fin d'un mot	E
16	V, v	v	46	Hh	silent
17	eu	2,9	47	ph	f
18	Z,z	z	48	ç, ce, ci, cy	s
19	ch	S	49	ge, gi, gy	Z
20	Nn, un	n,1	50	Xx	ks
21	T,t	t	51	silent end letters (all)	silent
22	D,d	d	52	ain, aim	5
23	C,c	k	53	ien	j5
24	P,p	p	54	eConsonant	E
25	Qq, qu	k	55	ti	s
26	B,b	b	56	Yy	j
27	Gg, gu	g	57	il, ill	ij
28	K,k	k	58	il,ill	j
29	s	z	59	ch	k
30	es	E	60	W,w	w

References

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. (2007). Haptics in learning to read with children from low socio-economic status families. *British Journal of Developmental Psychology*, *25*(4), 643–663. https://doi.org/10.1348/026151007X186643.

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.

Bara, F., Morin, M.-F., Alamargot, D., & Bosse, M.-L. (2016). Learning different allographs through handwriting: The impact on letter knowledge and reading acquisition. *Learning and Individual Differences*, *45*, 88–94. https://doi.org/10.1016/j.lindif.2015.11.020.

Bertrand, D., Fluss, J., Billard, C., & Ziegler, J. C. (2010). Efficacité, sensibilité, spécificité: Comparaison de différents tests de lecture, Abstract. *L'Année Psychologique, 110*(2), 299–320. https://doi.org/10.4074/S000350331000206X.

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.

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.

Braibant, J. M., & Gerard, F.-M. (1996). Savoir lire: Question(s) de méthodes? *Bulletin de psychologie scolaire et d'orientation*, 7-45.

Brankaer, C., Ghesquière, P., & Smedt, B. D. (2017). Symbolic magnitude processing in elementary school children: A group administered paper-and-pencil measure (SYMP Test). *Behavior Research Methods*, *49*(4), 1361–1373. https://doi.org/10.3758/s13428-016-0792-3.

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

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*(3), 369–378. https://doi.org/10.1007/s10648-012-9205-z.

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, 19*(1), 5–51. https://doi.org/10.1177/1529100618772271.

Cavalli, E., Duncan, L. G., Elbro, C., Ahmadi, A. E., & Colé, P. (2017). Phonemic—Morphemic dissociation in university students with dyslexia: An index of reading compensation? *Annals of Dyslexia, 67*(1), 63–84. https://doi.org/10.1007/s11881-016-0138-y.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354.

Coltheart, M. (2005). Modeling reading: The dual-route approach. In *The science of reading: A handbook* (pp. 6–23).

Colzato, L. S., van den Wildenberg, W. P. M., Zmigrod, S., & Hommel, B. (2013). Action video gaming and cognitive control: Playing first person shooter games is associated with improvement in working memory but not action inhibition. *Psychological Research Psychologische Forschung, 77*(2), 234–239. https://doi.org/10.1007/s00426-012-0415-2.

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., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., et al. (2010). How learning to read changes the cortical networks for vision and language. *Science, 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.-J., & Hupkens, D. (1990). *Test de vocabulaire actif et passif pour enfants de 3 à 5 ans—TVAP*. Issy-les-Moulineaux: Ed. scientifiques et psychologiques.

Dye, M. W. G., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science, 18*(6), 321–326. https://doi.org/10.1111/j.1467-8721.2009.01660.x.

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.

Gentaz, E., Sprenger-Charolles, L., & Theurel, A. (2015). Differences in the predictors of reading comprehension in first graders from low socio-economic status families with either good or poor decoding skills. *PLoS ONE, 10*(3), e0119581. https://doi.org/10.1371/journal.pone.0119581.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10. https://doi.org/10.1177/074193258600700104.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*(6939), 534–537. https://doi.org/10.1038/nature01647.

Green, C. S., & Bavelier, D. (2006). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1465–1478. https://doi.org/10.1037/0096-1523.32.6.1465.

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.

Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG rime and GG phoneme. *Reading Research Quarterly, 48*(1), 61–76. https://doi.org/10.1002/rrq.038.

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

Lefavrais, P. (2005). *Alouette-R*. Paris: Test d'analyse de la lecture et de la dyslexie.

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development, 89*(5), 1821–1838. https://doi.org/10.1111/cdev.12861.

Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 156–166. https://doi.org/10.3758/BF03195560.

Longcamp, M., Zerbato-Poudou, M.-T., & Velay, J.-L. (2005). The influence of writing practice on letter recognition in preschool children: A comparison between handwriting and typing. *Acta Psychologica, 119*(1), 67–79. https://doi.org/10.1016/j.actpsy.2004.10.019.

Lyytinen, H., Ronimus, M., Alanko, A., Poikkeus, A.-M., & Taanila, M. (2007). Early identification of dyslexia and the use of computer game-based practice to support reading acquisition. *Nordic Psychology, 59*(2), 109–126. https://doi.org/10.1027/1901-2276.59.2.109.

Mason, G., McDaniel, H., & Callaway, B. (1974). Relating reading and spelling: A comparison of methods. *The Elementary School Journal*, *74*(6), 381–386. https://doi.org/10.1086/460844.

McCandliss, B., Beck, I. L., Sandak, R., & Perfetti, C. (2003). Focusing attention on decoding for children with poor reading skills: Design and preliminary tests of the word building intervention. *Scientific Studies of Reading*, *7*(1), 75–104. https://doi.org/10.1207/S1532799XSSR0701_05.

Meyer, J., Dentel, L., & Meunier, F. (2013). Speech recognition in natural background noise. *PLoS ONE, 8*(11), e79279. https://doi.org/10.1371/journal.pone.0079279.

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.

Neuman, S. B., & Roskos, K. A. (1998). *Children achieving: Best practices in early literacy*. Newark, DE: Order Department, International Reading Association.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13*(1), 45–52. https://doi.org/10.3758/BF03193811.

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

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., et al. (2000). A cultural effect on brain function. *Nature Neuroscience, 3*(1), 91–96. https://doi.org/10.1038/71163.

Pourcin, L., Sprenger-Charolles, L., El Ahmadi, A., & Colé, P. (2016). Reading and related skills in Grades 6, 7, 8 and 9: French normative data from EVALEC. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology, 66*(1), 23–37. https://doi.org/10.1016/j.erap.2015.11.002.

Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development, 24*(4), 450–472. https://doi.org/10.1016/j.cogdev.2009.09.003.

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*, *10*(1), 39–60. https://doi.org/10.17011/ht/urn.201405281859.

Roland Goigoux, « Apprendre à lire et à écrire au cours préparatoire : enseignements d'une recherche collective », Revue française de pédagogie [En ligne], 196 | 2016, mis en ligne le 30 septembre 2016, consulté le 06 janvier 2020. http://journals.openedition.org/rfp/5069.

Saine, N. L., Lerkkanen, M.-K., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2011). Computer-assisted remedial reading intervention for school beginners at risk for reading disability: Computer-assisted reading intervention. *Child Development, 82*(3), 1013–1028. https://doi.org/10.1111/j.1467-8624.2011.01580.x.

Schacter, J., & Jo, B. (2016). Improving low-income preschoolers mathematics achievement with Math Shelf, a preschool tablet computer curriculum. *Computers in Human Behavior, 55*, 223–229. https://doi.org/10.1016/j.chb.2015.09.013.

Serrano, F., Genard, N., Sucena, A., Defior, S., Alegria, J., Mousty, P., et al. (2011). Variations in reading and spelling acquisition in Portuguese, French and Spanish: A cross-linguistic comparison. *Journal of Portuguese Linguistics, 10*(1), 183–204. https://doi.org/10.5334/jpl.106.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453. https://doi.org/10.3102/00346543075003417.

Sprenger-Charolles, L., Theurel, A. & Gentaz, E. (2017). Evaluer les capacités de lecture et les capacités reliées chez des enfants de 6-7 ans : Cadre théorique et bilan des études de Gentaz et al. (2013 et 2015). Document edited by FAPSE-UNIGE.

Tallal, P. (2004). Improving language and literacy is a matter of time. *Nature Reviews Neuroscience, 5*(9), 721–728. https://doi.org/10.1038/nrn1499.

Uhry, J. K., & Shepherd, M. J. (1993). Segmentation/spelling instruction as part of a first-grade reading program: Effects on several measures of reading. *Reading Research Quarterly, 28*(3), 219–233. https://doi.org/10.2307/747995.

Vousden, J. I., Ellefson, M. R., Solity, J., & Chater, N. (2011). Simplifying reading: Applying the simplicity principle to reading. *Cognitive Science, 35*(1), 34–78. https://doi.org/10.1111/j.1551-6709.2010.01134.x.

Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., & Cohen, D. (2006a). Principles underlying the design of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, *2*(1), 19. https://doi.org/10.1186/1744-9081-2-19.

Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. (2006b). An open trial assessment of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, *2*(1), 20. https://doi.org/10.1186/1744-9081-2-20.

Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain & Education, 3*(4), 224–234. https://doi.org/10.1111/j.1751-228X.2009.01075.x.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., et al. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science, 21*(4), 551–559. https://doi.org/10.1177/0956797610363406.

Zoccolotti, P., De Luca, M., Di Pace, E., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language, 93*(3), 369–373. https://doi.org/10.1016/j.bandl.2004.10.010.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cassandra Potier Watkins is a PhD student, under the supervision of Stanislas Dehaene, at the Neurospin Cognitive Neuroimaging lab in France. Her doctoral research investigates the common threads that allow our brains to master cultural skills like reading and math, and in turn, how learning changes the brain. She takes a multidisciplinary approach that encompasses the fields of cognitive neuroscience and on-the-field education research. The goal of her work is to advance the science of how we learn and develop applicable solutions for an evidence-based education. She is an associate member of the scientific council for the French Education Ministry.

Julien Caporal For over 29 years, Julien Caporal has been passionately devoted to new organisational and learning practices. His PhD focused on hypermedia and GIS. He has been fully immersed in the web from the early years, creating one of the first French web agencies. He founded and participated in many Start-up to address collaborative solutions, Enterprise Content Systems, Blended-learning solutions. Today at Manzalab he is Associated Partner and Pedagogical Director.

Clément Merville A computer science engineer and HEC Paris graduate, Clément Merville has held various management and business development roles at major video-game publishers. In 2010, Clément anticipated new uses of video games, especially in the field of training, and identified gaming as a powerful learning vector: He created a successful company specialized in Serious Game : Manzalab.

Sid Kouider is a professor of cognitive neuroscience at the brain and consciousness unit at Ecole Normale Supérieure in Paris. His research investigates the behavioral and neural determinants of cognitive functions such as learning and consciousness in humans. His work relies on psychophysics and brain imaging methods to address in particular the development of these functions in children and adult populations. Sid Kouider is a member of the scientific council for the French Education Ministry.

Professor Stanislas Dehaene holds the Chair of Experimental Cognitive Psychology at the Collége de France in Paris. He directs the NeuroSpin center in Saclay, south of Paris – France's advanced neuroimaging research center. His research investigates the neural bases of human cognitive functions such as reading, calculation and language, with a particular interest for the differences between conscious and non-conscious processing, and for the impact of education on the brain. Prof. Dehaene is a member of six academies and a recipient of the Brain Prize (with G. Rizzolatti and T. Robbins, in 2014). He has written several books for the general public, including The Number Sense, Reading in the Brain, Consciousness and the Brain, and How we learn, which were translated into more than fifteen languages.