Enhancing reading skills through a novel video game mixing action mechanics and cognitive training

Angela Pasqualotto[1,4*], Irene Altarelli[2], Antonella De Angeli[3], Zeno Menestrina[3] Daphne Bavelier[4#], & Paola Venuti[1#]

Author affiliations

1 – Department of Psychology and Cognitive Science, University of Trento, Trento, Italy

2 – Université de Paris, LaPsyDÉ, CNRS, Paris, France

3 – Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

4 – Faculte de Psychologie et Science de l'Education, University of Geneva, Geneva, Switzerland

[*]Corresponding Author:

Angela Pasqualotto, Department of Psychology and Cognitive Science, University of Trento, Trento, Italy

[#]These authors contributed equally.

Authors' contributions.

Z.M., A.P., A.D. and P.V. contributed to the design of the video game.

A.P. and P.V. designed the experiment; I.A. and D.B. provided advice on the experimental design.

A.P. contributed and supervised the data collection.

A.P., I.A. and D.B. performed the data analyses and wrote the manuscript.

This study constitutes part of A.P.'s PhD thesis under the supervision of P.V.

Literacy is an essential requisite in modern societies. Poor readers, that is children encountering difficulties with decoding and/or understanding print material, face a disproportionately higher risk of struggling at school and in life[1–3]. Thus, developing proper training strategies for reading is a goal of upmost importance.

Several models of reading acquisition have emerged in the course of the last decades, highlighting the importance of oral language skills, and especially of phonological awareness, for the appropriate development of reading in childhood[4–7]. On top of phonological awareness skills – that is the ability to identify and manipulate sub-units within words, such as syllables and phonemes – other, more domain-general abilities have also been related to reading acquisition. Among these, attentional control – flexibly distributing attention on what is task-relevant and filtering out distractors – appears to be key at several levels. Attention needs to be correctly distributed on page to successfully decode text, as a printed page is typically quite crowded. In this context, attentional control is especially important to group letters into words and phrases as well as swiftly move from one line to another; this is especially so before perceptual expertise for printed material emerges[8–10]. Attentional control is also essential for suppressing close phonological competitors or orthographic competitors, and flexibly switching between different types of information to be extracted from the same string of letters (i.e. phonological, orthographic or semantic information)[11,12]. Interestingly, attentional skills are not only related to concurrent reading abilities during childhood, but early attentional abilities – measured before the acquisition of reading – have been identified as predictors of later reading skills, particularly for decoding[8].

Related aspects of executive functions (EFs) have also been associated with reading abilities. Among these, working memory appears to be essential for holding sequences of phonemes in memory as decoding proceeds, and for keeping track of successive words as sentences unfold[13–19]. In addition, cognitive flexibility may be relevant for building cross-modal connections and switching between the multiple dimensions of written content, such as phonology, morphology, semantics and syntax[20–24]. Finally, it has been argued that planning – the ability to identify, monitor and, if needed, revise strategies – may be relevant for reading comprehension[25,26].

Considering the importance of attentional skills for reading, training attentional control has been suggested as a promising way of lifting one of the several potential brakes on achieving literacy. To this end, one possible approach is making use of action video games, which are known to change for the better several attentional components in typical adults ([27–29], see[30] for an extensive meta-analysis of action video game impact on perceptual, attentional, and cognitive skills) and children[31]. Franceschini and collaborators[32] trained Italian dyslexic readers with child-appropriate commercially-available action-based video games and showed increases in reading speed without significantly altering reading accuracy. This effect, first established in dyslexic readers of a transparent orthography, Italian[32], has recently been replicated in dyslexic readers of a more opaque written language, English[33]. Though promising, these findings remain controversial[34,35] (for a review, see[36]). It is also the case that these past experiments mainly focused on reading-impaired children leaving it unclear whether training attentional skills would benefit reading abilities in typically developing children in a way that is relevant for academic achievement.

The current study evaluates whether, in a relatively large sample of typically-developing children, training attentional control will indeed result in benefits in reading abilities, as measured by not only standardized tests but also school grades. A first, novel aspect of this work is the development of an entertaining child-appropriate video game, *Skies of Manawak (SOM).* This game was designed to target the diversity of cognitive challenges that reading poses and to leverage the reported positive impact of action video game-based training on attentional control (see Fig. 1). The gameplay, situated in a child-friendly adventure world, revolves around two main game mechanics. One, *"The Flight",* that bears key characteristics of action video games, and in particular a need for pacing, a load on distributed attention and the necessity to shift between divided and focused attention states at proper times[37]. The other, *"The Village"*, bears key characteristics of an incentive world, where the players through slow-paced, exploratory moves discover how points earned during the rest of the gameplay can be redeemed to personalize his/her world. In the Village, players also discover the next quest they are assigned to, sending them back into the action game. In addition to the Flight, the game consists of 9 mini games, each targeting aspects of executive functions engaged during reading acquisition, such as visual and auditory working memory, inhibitory control and cognitive flexibility to cite a few. Best practice from cognitive training was naturally aligned with action video game mechanics in an effort to optimize impact on attentional control and its related EF[38,39]. Assigned mini-games, their in-game progression, as well as the action game difficulty was adaptively personalized. This game was developed over a period of 2 years through an iterative process that included game designers/artists, domain experts (cognitive and clinical scientists) and players (children between 8 and 14 years old who participated in story and aesthetics of the game through game ideation workshops, N=60, and helped refine the game through playtesting, N=250 - See SI, Section 5 - Training Regimens). We refer the reader to the Supplementary Table 1 for a synthesis of the attentional/executive skills targeted by Skies of Manawak, and their links to the various game elements in Skies of Manawak.

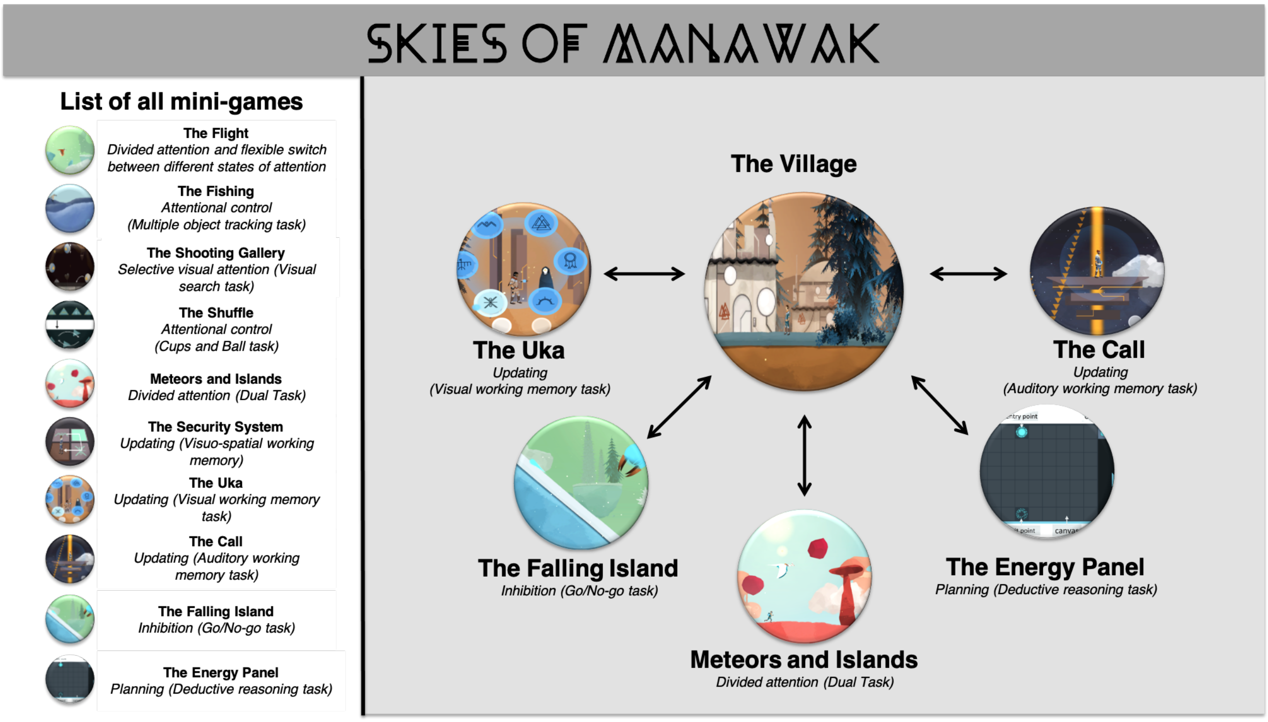

Figure 1| Skies of Manawak mini-games and targeted EF/attentional skills. We are reporting here only the main executive sub-processes that are tapped by each of the mini-games.

151 typically developing Italian children were recruited in this training study. Both the experimental group (N=79) and the active control group (N=72) were pre-tested for visuo-spatial attention, verbal and spatial working memory, planning and literacy skills (T1). All children then underwent 12 hours of training, distributed over 6 weeks, either on the experimental video game, named Skies of Manawak (SOM), or on a control video game, Scratch, a kid-tailored, engaging programming game. These two different interventions were presented as active to parents and teachers. Crucially,

training was implemented during class hours under the supervision of dedicated staff, trained respectively either in the experimental or the control game, ensuring balanced implementation across the two arms of our intervention. The same cognitive measures were collected at post-test, within a week from the end of training (T2), as well as six months later (T3). Importantly, among the three experimenters collecting data, two were blinded to the aims of the study, allowing us to control for possible experimenter effects. Academic achievement in Italian and in Mathematics was additionally tracked at 1.0 (T4) and 1.5 years follow-up (T5). We predicted greater improvements in attention and in reading skills from before to after training in the experimental group as compared to the control group. In addition, standardized reading assessments were expected to show more robust group differences than academic achievement scores, as the former are more targeted to the study design, while the latter represent further transfer. Still a key question concerned whether reading grades may show greater benefit in the experimental than the control group over the years, a test of the predicted transfer to everyday life activities.

Results

We first consider the impact of the experimental video game, Skies of Manawak (SOM), as compared to that of the control training game, Scratch, from pre-test (T1) to post-test (T2), and possible long lasting effects at 6 months (T3) on both reading skills (word, non-word and text reading speed and accuracy, as well as types of reading errors made in the text reading task) and on a measure of attentional control. We refer the reader to the Supplementary information (Section 2 – Results: Additional training effects and Supplementary Table 5) for the more distal outcomes of short-term memory, working memory, writing and comprehension of a short text passage, for which we only have measurements at two time points. We also consider here planning, an exploratory outcome and the only other measure collected at all 3 time points. Second, we tested the hypothesis that the experimental video game training enhanced reading skills through attentional enhancements as proposed by Franceschini and colleagues[40]. Finally, the impact of our training in the academic domain was evaluated by analysing Italian grades and Mathematics grades collected at five time points (T1, T2, T3, but also 12 months after the end of training – T4, and 18 months after the end of training – T5).

1. Analysis Plan

Data was analysed using SPSS Statistics for Mac, Version 16.0 (IBM Corp). Group differences were examined by means of Analyses of Covariance using repeated measures general linear model (GLM) analyses with *Time* as a within-subject factor and *Group* (SOM versus Control) as a between-subject factor. The Time X Group interaction was our primary effect of interest and was followed with post-hoc analyses when significant. In order to exclude possible confounding effects, *chronological age* and *IQ scores* (Raven's CPM[41]) at T1 were entered as covariates and *Sex* as a between-subject factor for all MANCOVA/ANCOVA analyses. For each of our four different hypotheses (1.a. reading skills, 1.b. types of reading errors; 2. attentional control and planning; 3. academic performance), we applied Bonferroni correction for multiple comparisons. In addition, effect sizes were presented as partial eta square (SSeffect / (SSeffect + SSerror). We refer the reader to the Supplementary information (Section 2 – Results: Covariates effects) for covariates effects.

All participants' performance measures were transformed into z-scores on the basis of age norms, while for IQ the standard score (M=100; DS=15) was considered. Reading errors were analysed as raw scores. Importantly, the two groups were comparable at baseline in all demographic and neuropsychological variables, but text comprehension and proportion of word substitutions errors (Supplementary Table 2). The descriptive statistics for the main variables at the first assessment time point (T1) are presented in Supplementary Table 2 (Literacy skills performance; attention and planning; other EF skills) and in Table 3 (Demographic and IQ characteristics).

Second, to evaluate the proposal that video game training enhanced reading skills through attentional enhancements, we ran separate analyses for reading speed (Δ T1-T2 general reading speed) and for reading accuracy (Δ T1-T2 general reading accuracy). Three steps, fixed entry, hierarchical multiple regression analyses were computed taking into consideration covariates effects (Model 1: age, IQ, sex), pre-test reading performance (Model 2: T1) and changes in attention (Model 3: Δ T1-T2), with changes in planning added as an exploratory variable to the latter model. Analogous regression models were also performed on the two main categories of reading errors observed in our sample.

Finally, in order to evaluate the generalization of potential improvements to the academic domain, we computed two separate ANCOVA on Italian grades and Mathematics grades respectively, obtained at all five time points.

2. Skies of Manawak enhances reading skills

Reading skills were assessed by measuring speed and accuracy on the reading of Word lists, of Non-word lists (DDE-2 battery[42]) and of meaningful Text (MT reading test[43,44]). Separate 3x3x2 MANCOVA with Time (T1, T2, T3) and Task (Word, Non-word, Text) as within-subject factors, Group and Sex as between-subject factors and Age and IQ as covariates were carried out for speed and accuracy.

For reading speed, a Time X Group interaction [$F(2,144)= 27.211$, $p <.001$, $\eta^2_p = .274$] indicated greater improvement in reading speed in the SOM group than in the control group, that was largely maintained at T3 (see Fig. 2A and Supplementary Table 4). The only other significant effect was a triple interaction Time X Group X Task [$F(2,144) = 5.280$, $p=.002$, $\eta^2 p= .129$] highlighting that the highest improvements in reading speed were achieved by the SOM group on the text reading fluency task, compared to word and non-word reading tasks. When considering the covariates, a significant interaction between Time and Age [$F(1,145) = 5.117$, $p= .014$, $\eta^2 p = .066$] revealed that, regardless of the group, younger children improved more than older children.

For reading accuracy, a Time X Group interaction [$F(2,144) = 15.630$, $p =.002$, $\eta^2 p = 0.178$] confirmed greater improvement in reading skills in the SOM group than in the control group (Fig. 2B); no other effects were significant.

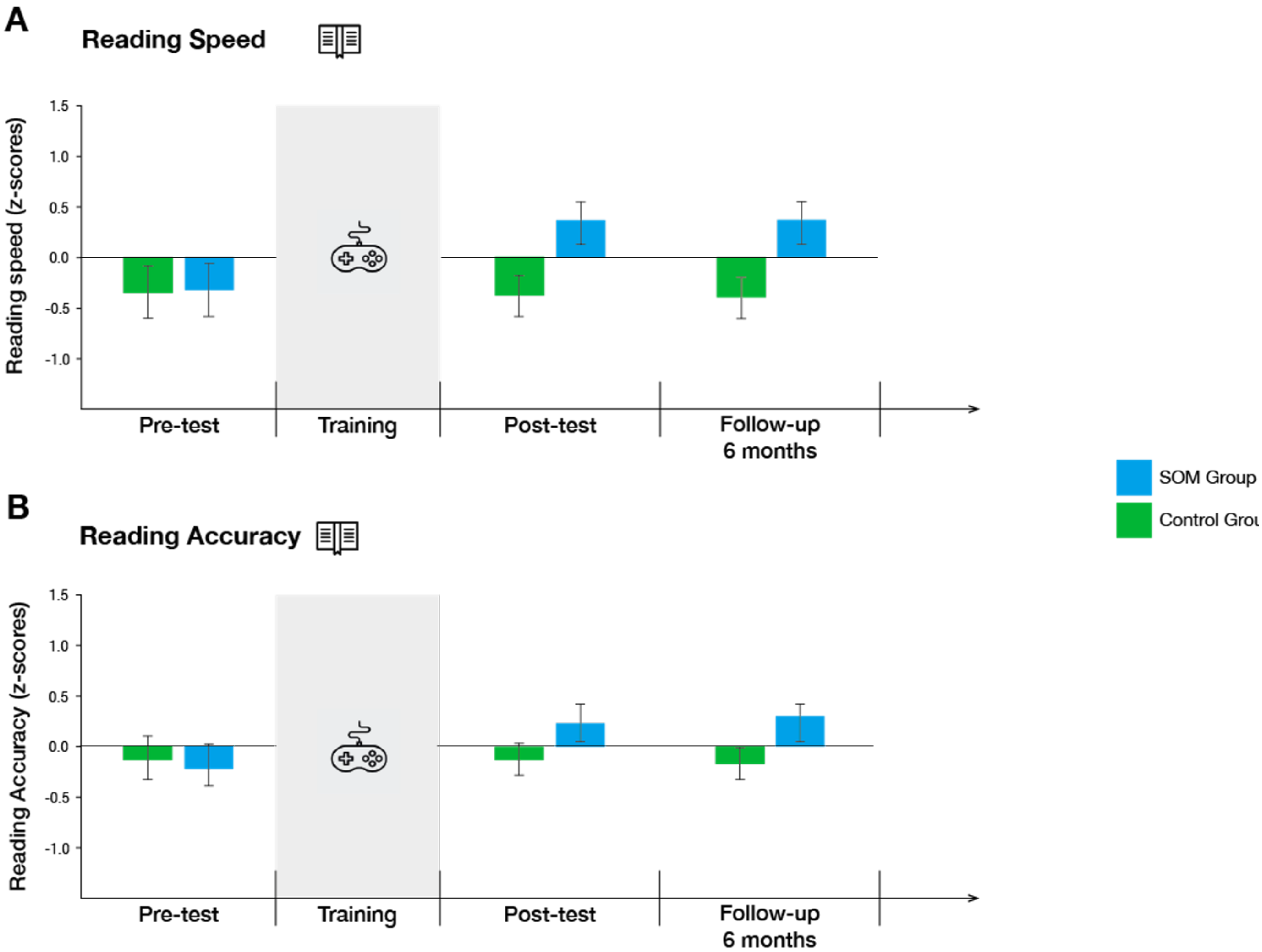

Figure 2| The mean of word, non-word and text reading speed (**A**) and accuracy (**B**), as measured before training (T1), after training (T2), and at 6 months follow-up (T3) in the SOM-trained (N = 79) and the Scratch-trained (N = 72) groups. Reading performance is expressed in z-scores. Error bars represent confidence intervals at 95% (CI).

3. Skies of Manawak differentially reduces reading errors

Reading errors made in the text reading fluency task[43,44] were categorized as either Sounding-out behaviours (i.e., sounding out parts of the word before producing the whole word) or Word-substitution errors, following the proposal of Hendriks and Kolk[45], which has been applied to Italian by Trenta and collaborators[46]. The categorization of reading errors applied here aims at differentiating between decoding capabilities and word unitization (i.e., the process by which single linguistic units are consolidated into whole-word units)[47] through the use of separate scores. Sounding-out behaviours are diagnostic of fluency weaknesses in the sub-lexical, orthographic-to-phoneme mapping processes, whereas word-substitution errors (in the presence of normal fluency) rather point to weaknesses in lexical, orthographic and phonological to semantic access processes[46–49].

As above, a 3x3x2 MANCOVA with Time (T1, T2, T3) and Error Type (Sounding-out, Word-substitution) as within-subject factors, Group and Sex as between-subject factors and Age and IQ as covariates were carried out on error rates. A significant Time X Group interaction [$F(2, 144) = 15.940$, $p < .001$, $\eta^2 p = .181$] indicated greater reduction of reading errors in SOM group compared to the control group, a reduction that was largely maintained at T3 (see Fig. 3 and Supplementary Table 4). Moreover, a triple Time X Group X Type of Errors interaction [$F(2,144) = 11.635$, $p = .008$, $\eta^2 p = .074$] highlighted a greater reduction in sounding-out behaviours compared to the Word Substitution errors in the SOM group than in the control group.

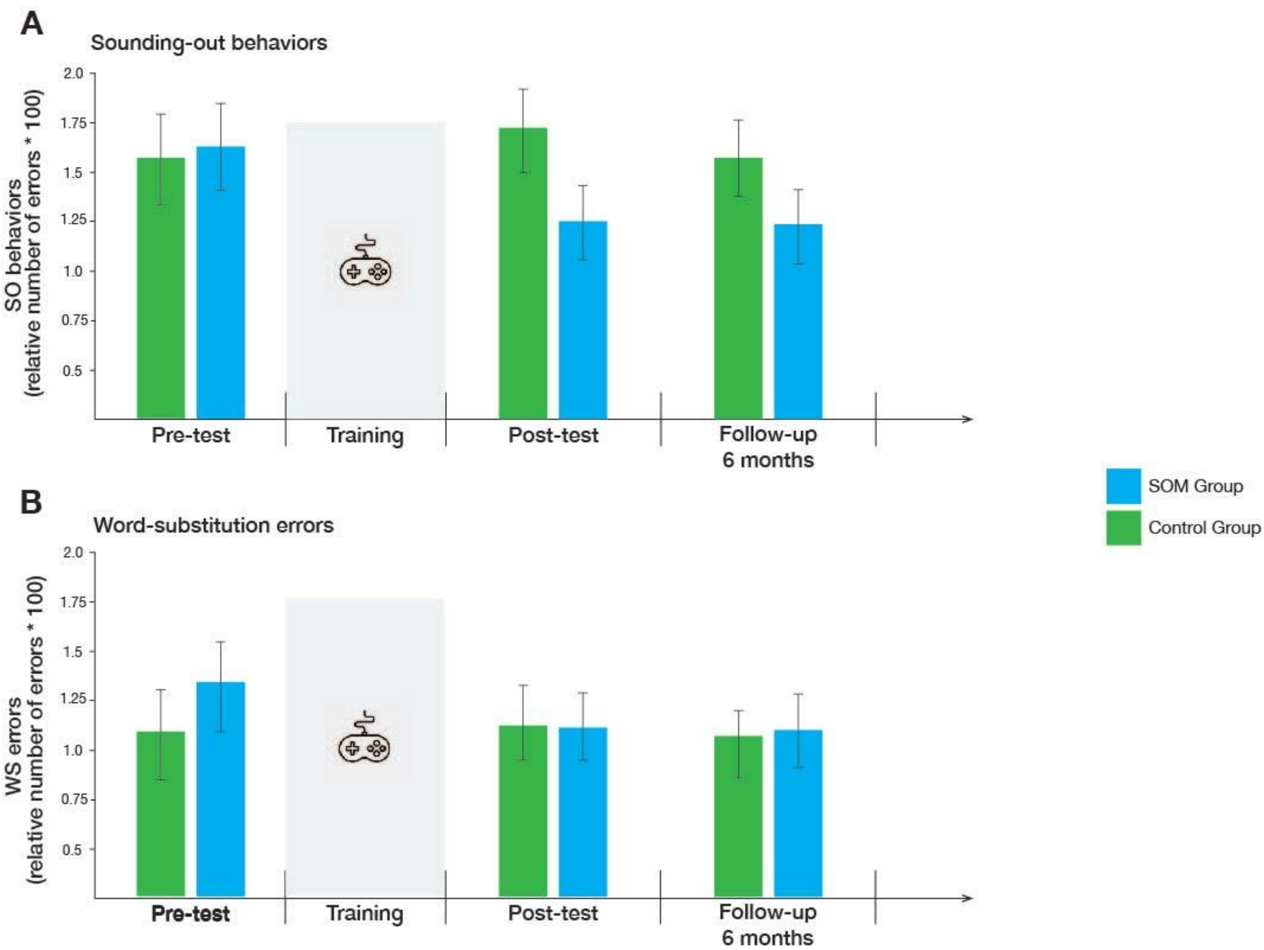

Figure 3| Errors - Sounding-out behaviours (**A**) and Word-substitutions (**B**), as measured before training (T1), after training (T2), and at 6 months follow-up (T3) in the SOM-trained (N = 79) and the Scratch-trained (N = 72) groups. It should be noted that compared to children reading opaque languages (e.g., English), Italian-speaking children make considerably fewer errors. Thus, errors -

expressed as the number of errors over the total number of enunciated words (*100) – are relatively rare. Error bars represent confidence intervals at 95% (CI).

4. Skies of Manawak enhances visuo-spatial attention

We examined attention after training by means of a classic barrage task[50], where children are asked to find as many target objects (bells) amidst distractors as possible. The mean accuracy score between the 'fast' score (number of targets found within 30 seconds) and the 'slow' score (total number of targets identified in 120 seconds) was estimated and used as a composite dependent variable.

The 3x2 ANCOVA with Time (T1, T2, T3) and Group (SOM vs Control) and Sex (Age and IQ as covariates) revealed a significant Time X Group interaction [$F(2,144) = 63.764$, $p < .001$, $\eta^2_p = .47$]. Figure 4A highlights greater improvements in attention in SOM than in the Control group from T1 to T2, with the effect being largely maintained at T3. No other significant effects were observed but for a main effect of Group [$F(1,145)=52.365$, $p = .001$, $\eta^2 p = .265$].

In addition, we carried exploratory analyses on planning, which we had assessed at all three time points through the Tower of London task[51]. A similar ANCOVA with the accuracy of planning as the dependent variable revealed a significant Time X Group interaction [$F(2,144) = 15.546$, $p = .002$, $\eta^2 p = .178$], indicating greater improvement in the SOM than the control group from T1 to T2, with this effect being sustained at T3 (see Fig. 4B and Supplementary Table 4). The only other significant effect was a triple interaction Time X Group X Sex [$F(1, 145) = 7.852$, $p = .012$, $\eta^2 p = .051$] discussed in the SI (Section 2 – Results: Covariates effects).

Thus, as expected given the specific focus of SOM in training attention and higher executive function skills, twelve hours of SOM play led to greater improvement in these skills, as compared to playing the same amount of the control game, Scratch.

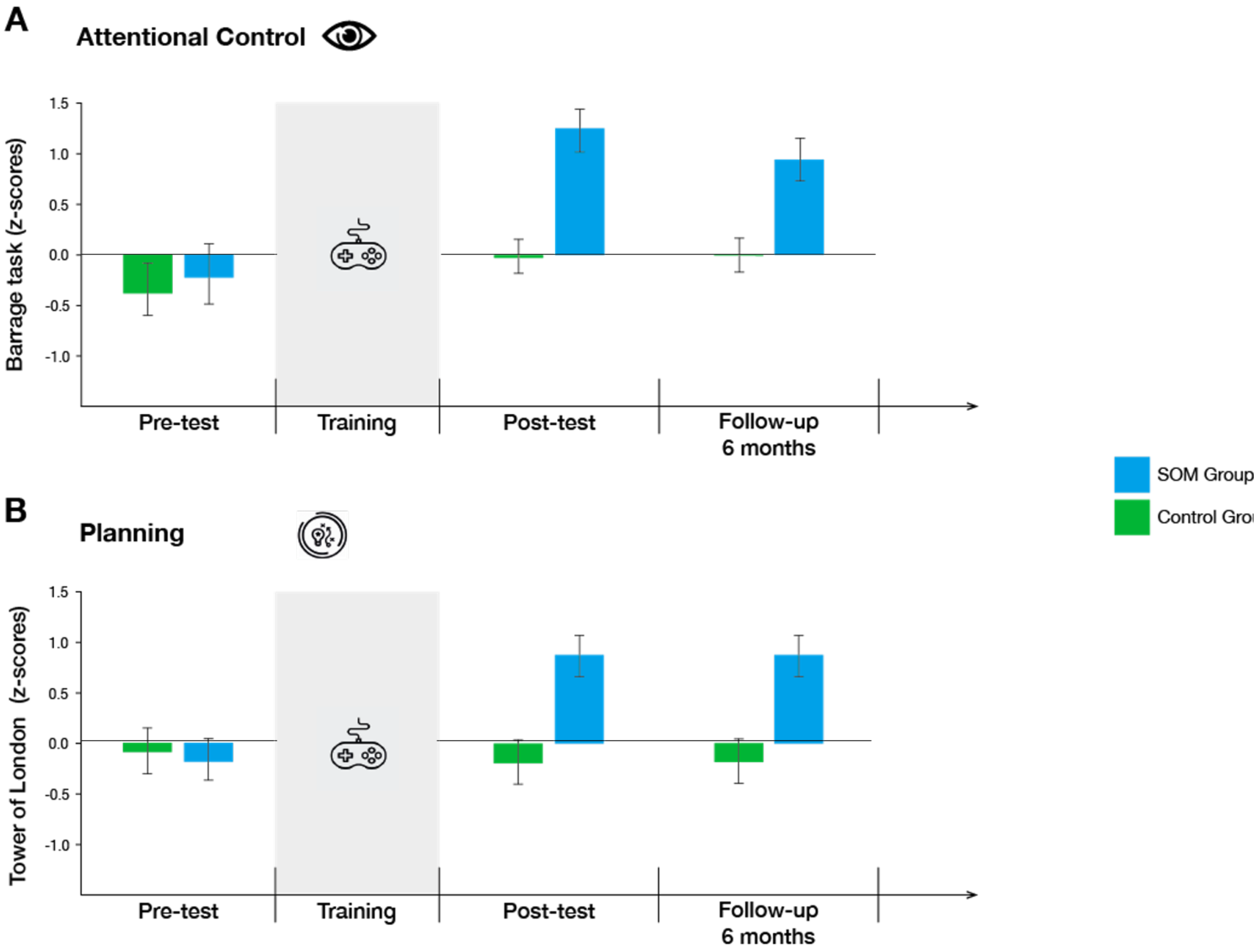

Figure 4| Performance (accuracy score) in the barrage task (**A**) and the Tower of London task (**B**), expressed in z-scores, as measured before training (T1), after training (T2), and at 6 months follow-up (T3) in the SOM-trained (N = 79) and the Control-trained (N = 72) groups. Error bars represent confidence intervals at 95% (CI).

In summary, the SOM-trained group exhibited greater improvement in reading speed and in reading accuracy as well as in our measures of attentional control and planning than the control-trained group. These differential gains were largely maintained 6 months later at follow-up.

5. Controlling for possible test-retest learning confounds

The same tests (namely all reading tests, as well as the barrage and planning tasks) were used through the three assessment sessions, meaning that improvements could be due to better memorization of the test items per se rather than enhanced reading or attentional skills. To check for such a possible test-retest learning confound, two novel tests were administered at the 6-months follow-up: a new text-reading task[52] and a supplementary barrage task[52] (see SI Section 3 – Test Batteries). A Pearson correlation was run to determine the relationship between these novel tests and the original text-reading task and barrage task as measured at follow-up. We then checked the differences in these novel tests at follow-up between SOM and Control group.

Strong, positive correlations were observed between the two reading tests scores (r =0.653, n = 151, p <.001) as well as between the two attention tasks scores (r = 0.970, n = 151, p < .001), establishing that the novel tasks capture similar constructs as the original ones.

Separate one-way ANCOVAs were conducted to compare the effect of training regimens (SOM vs Control) on the novel text reading and the novel attentional control tasks at T3. For each of the novel tasks, a main group effect confirmed an advantage in reading and in attention for the SOM-trained group over the control trained group: new text reading [SOM: M = 0.46, SD = 0.50; Control: M = -0.025, SD = 0.49; main effect of Group: F(1, 150) = 36.575, p < .001, η^2_p = .201,] and new barrage task [SOM: M = 1.25, SD = 0.85; Control: M = -0.01, SD = 0.63; main effect of Group: F(1, 150) = 112.629, p < .001, η^2_p = .437]. Thus, the advantages of the SOM trained group in reading and in attention are not limited to the tests used at T1-T2-T3 but rather generalize to similar tasks with novel stimuli at T3.

6. Role of attentional control in fostering literacy skills

A link has been proposed between training attention and enhancing reading abilities, especially as regards reading speed. A three steps, fixed entry, multiple regression analysis was performed on the entire sample of children (n = 151). The dependent variables were reading speed or reading accuracy improvements between T1 and T2, and the predictors were: (1) age, IQ and sex; (2) general reading speed respectively reading accuracy at T1, (3) attentional control changes (T2-T1), and as an exploratory variable, changes in planning (T2-T1).

The hierarchical multiple regression (for details, see Table 6 in SI) revealed that when all three demographic variables were included in Model 1, only Age significantly explained some of the variance in reading speed changes (4.4%); this model, which not only included age but also sex and IQ, did not provide a significantly better fit to the data compared to Model 1 [F(3,147) =2.282, p= .082]. Reading speed at T1 explained an additional 11% of variance [F(4, 146) = 6.533, p < .001, ΔR^2= .107]; this model provided a better fit of the data [F(1,146)= 18.474, $p < .001$]. In particular, children who improved most were those who started with lower levels of reading speed at T1. Finally, the addition of changes in attention and planning between T1 and T2 to the regression model explained an additional 15% of variance [F(6,144) =10.523, $p < .001$, $\Delta R^2 = .153$]. This change in R^2 was significant [F(2,144) =15.846, $p < .001$]. Importantly, only improvements in attention made a significant contribution to the prediction (p < .001), whereas improvements in planning did not (p = 0.132).

Results of the same hierarchical multiple regression with reading accuracy as dependent variable showed that none of the demographic variables significantly predicted the changes in reading accuracy [F(3,147)= 0.655, $p = .581$, $r^2 = .013$]. Reading accuracy score at T1 uniquely explained 39% of the variance in reading accuracy changes [F(4,146) = 23.436, $p < .001$, $\Delta R^2 = .378$]; this change in R^2 was significant [F(1,146)= 90,58, $p < .001$]. Again, children with worse accuracy improved more overall. Adding improvements in attention and in planning to the regression model explained an additional 4% of the variance [F(6,144) =18.135, p < .001, $\Delta R^2 = .039$] and this change in R^2 was significant [F(2,144)= 4.979; p=.008]. Again, only improvements in attention made a significant contribution to the prediction (p = .011). For details, see Table 7 in SI.

Furthermore, in a further, more exploratory analysis, we investigated the link between the reduction in reading errors and the attentional improvements. We performed a three-step, fixed entry, multiple regression analysis on the entire sample of children. The dependent variable was the changes in sounding-out behaviors, and as above, the predictors were: (1) age, IQ and sex; (2) sounding-out behaviors at T1, (3) attentional control changes (Δ T2-T1) and, as an exploratory variable, planning changes (Δ T2-T1). The delta scores in sounding-out behaviors were indexed by calculating the difference between sounding-out behaviors at T2 and T1. Analogously, to measure the unique effect of attentional improvements on word-substitution errors reduction, we performed a similar separate analysis, this time using the changes in word-substitution errors as dependent variable and pre-test word-substitution errors as predictor in step 2 of the regression model. For details, see SI Table 8.

The results showed that none of the demographic variables significantly predicted the changes in either sounding-out or word-substitution errors [SO: F(3,147)= .614, p = .607, r^2 = .012; WS: F(3,147)= 1.295, p = .28, r^2 = .026]. Errors at T1 uniquely explained 30% of the variance in sounding-out behaviours [SO: F(4,146) = 16.812, p < .001, ΔR^2= .303] and 36% in word-substitution [WS: F(4,146) = 23.140, p < .001, ΔR^2= .362]. Specifically, children expressing such behaviours more frequently at T1 were those who improved the most overall. Adding improvements in attention and in planning to the regression models explained an additional 9% of the variance for sounding-out behaviours [F(6,144) =15.894, p < .001, ΔR^2= .08], whereas they explained only an additional 0.2% of the variance for word-substitution errors [F(6,144) =15.374, p < .001, ΔR^2= .002]. Interestingly, improvements in attention significantly accounted for a portion of variance (p < .001) for changes in sounding-out behaviours but not in word-substitution errors (p=.403), linking attentional control to the reduction in sounding-out behaviours and thus more efficient orthographic to phoneme mapping processes (i.e., sequencing skills). For details, see Table 8 and 9 in SI.

7. Links with academic performance

In this study, six classes (two 3rd grade, two 4th grade, two 6th grade and two 7th grade) were involved. For each grade, one class was randomly allocated to the SOM group and the other one to the Control group. The two pairs of classes in elementary school (3rd, 4th grade) shared the exact same teams of teachers, while the middle school classes had in common Italian, Mathematics & Science, and English teachers (for a total of 19 teaching hours out of 30 hours per week). Such matching at the teacher level was preferred to limit possible confounds due to teaching styles when considering the impact of training on academic achievements.

Grades in Italian and in Mathematics were entered into two separate 5x2 ANCOVA with Time (T1, T2, T3, T4 and T5) and Groups (SOM, control). For Italian, a significant Time X Group interaction [F(4, 142) = 3.448, p = .016, η^2_p = .023] was observed, highlighting growing group differences over time, due to a small advantage of the SOM group over the Control group over time (see Fig. 5).

For Mathematics, no significant Time X Group interaction [F(4, 142) = .778, p = 1.08, η^2_p = 0.005] was observed.

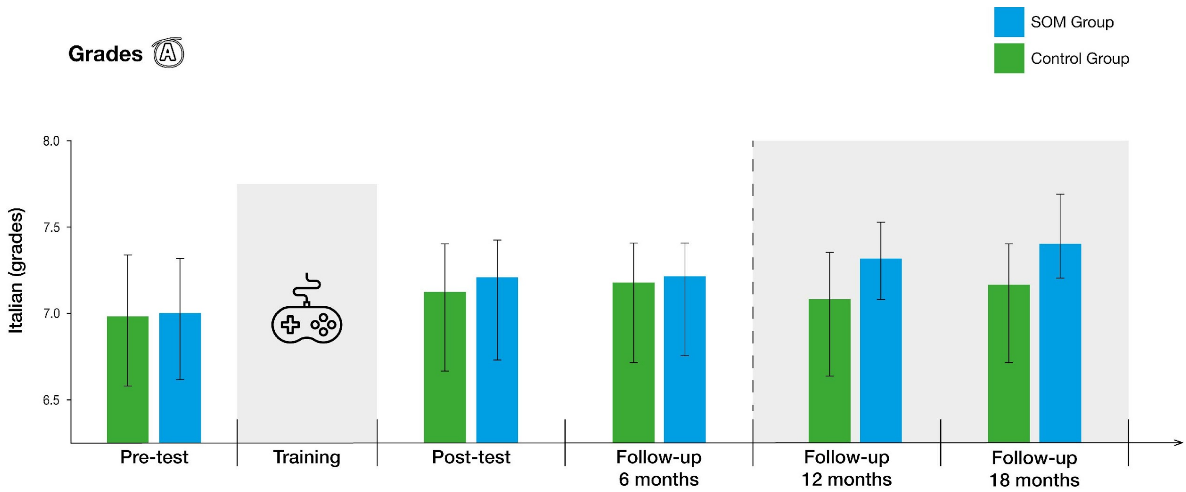

Figure 5| Academic grades (out of a maximum of 10) in Italian for children of the two groups (SOM-Control) at all five time points: pre-test (T1, at the beginning of the first school year), post-test (T2, at the end of the first semester), follow-up 6 months after the end of the training (T3, at the end of the first school year), 12 months after the end of the training (T4, at the end of the first semester of the second school year), and 18 months after the end of the training (T5, at the end of the second school year). Error bars represent confidence intervals (CI).

Discussion

The current study presents a custom-designed video game for children targeting attentional control training and executive functions relevant to reading, with the goal of facilitating some aspects of reading acquisition. The SOM game includes several mini-games loading on executive functions, each accessed through a central game encompassing action video game mechanics, hypothesized to enhance attentional control[37]. Overall, our results demonstrate both near- and far-transfer benefits following such training as compared to training on Scratch, a block-based visual programming game. Children in the experimental group showed greater enhancements in attentional control as compared to the control-trained group, as well as greater improvements in reading speed and accuracy. These enhancements were maintained 6 months after the end of training. Importantly, none of the training activities required reading nor searching for static target objects among distractors, as in the paper-and-pencil barrage tasks we used in the assessments.

Moreover, benefits – though of a small effect size – were also observed in school grades in Italian – a core school subject – 18 months after training, thus showing further transfer to academic performance. Crucially, no comparable improvements were observed as an outcome of the active control intervention, using Scratch.

These results extend effects previously reported in the context of developmental dyslexia remediation to a large sample of *typically developing children*. A handful of studies have used non-conventional training tools such as a child-friendly, commercial video game (Rayman's Raving Rabbids) to yield improvements in both attentional control and reading speed in Italian[32,53] and English-speaking[33] dyslexic children (but see[35], for a lack of replication). Importantly, our results expand on those previous findings by highlighting benefits in not only reading *speed* but also reading *accuracy*, the latter being crucially aligned with the general aim of increasing literacy outcomes as broadly as possible, and potentially supporting the development of text comprehension later in time[54]. Moreover, the gains in reading accuracy can be further appreciated by considering changes in reading behaviours. Indeed, the scoring performed on reading behaviours during text reading allowed us to separately assess orthographic to phonological decoding from recognizing words as whole units. Results showed that children who trained with the experimental video game displayed overall fewer errors after training than those trained with the control intervention, an effect driven possibly by an even greater reduction in Sounding-out behaviours at post-test in the experimentally trained group. Importantly, this advantage was maintained at a follow-up test 6-months later.

Additionally, our results add to previous literature by showing that the obtained benefits on attentional control and reading abilities are lasting, as they were found to be maintained 6 months after the end of training.

A distinctive feature of the present study was to simultaneously track academic achievement in Italian and to continue tracking performance in Italian classes for another full academic year. Italian grades comprise both written and oral comprehension; speaking and writing capacities (i.e., oral and written production), as well as metalinguistic skills (i.e., knowledge and application of grammar rules in Italian, which is participants' mother tongue). We show that our novel video game intervention provides transfer beyond lab-based assessment of reading, namely by positively affecting school grades in Italian.

Importantly, children, parents and all but one experimenter were blinded as to the aim of the study. Note that an analysis of only the children assessed by blind experimenters confirms all main results (see Supplementary Information 2 - Controlling for possible testing confounds). Although teachers could not be blinded to their class assignment, the finding that differences in Italian grades were only seen 18 months after the training speaks against expectation effects. A potential weakness of this work is that the cognitive tests administered at pre-, post and follow-up tests used the exact same items. At follow-up, however, two additional measures of reading ability and attention were added to ensure that greater gains in the SOM group as compared to the control-trained group were not due to greater memory of the very test items administered at all time points. This was indeed the case. By showing generalization to novel test items, these results provide further support to the observation that SOM brings benefits to reading ability and attentional control skills per se.

A critical issue concerns the link between reading performance and attentional control skills. As in Franceschini and collaborators[32], we tested for a link between reading improvement from pre- to post-test and attention improvements. We did confirm such a relationship. This is in line with a number of previous investigations that have suggested the existence of a link between attention and reading development. For instance, in typically-developing children, a longitudinal study[8] showed that attentional skills evaluated at pre-reading stages are predictive of reading abilities later in grade 1 and grade 2. Indeed, several attentional components have been related to reading proficiency. Among these, efficient attentional control appears to be essential for moving the spotlight of attention over the letters of a text, while inhibiting the processing of close phonological or orthographic competitors, and flexibly shifting from one line to the other[10,55]. Attentional control is also fundamental for visual word recognition, which consists of the analysis of different linguistic attributes of printed words (i.e., the orthographic, phonological, and semantic). Indeed, it has been shown that attention has the capacity of enhancing the processing of printed material, flexibly switching between different levels of information processing depending on task goals[12].

In addition, greater attentional control and more generally better ability to attend to task-relevant materials and ignore sources of noise or distractions have been linked to enhanced reading skills in typically-developing[8,12] and in dyslexic children[55–57].

In line with those previous results, our study also highlights a putative link between attentional control and reading abilities, as a multiple regression on the entire sample of children has shown that gains in attention explained up to 15% of the variance in reading speed.

This finding should be taken with a note of caution, however. We acknowledge that the existence of a *causal* link between gains in attentional control and gains in reading abilities remains to be established. Indeed, while changes in attentional control could explain part of the variance in reading gains across the whole sample of children (N=151), this relation was not confirmed when splitting the whole population into the two training groups (see SI Section 2 – Results: Role of attentional control on fostering reading skills - for details). Differences in attentional control gains and in reading gains between the groups are thus driving the observed relation, not allowing us to firmly conclude on the existence of a causal link between the two.

An interesting and important question concerns the impact of any training regimen on real-life competence – and in the case of school-aged children, on academic achievement. By collecting grades in both Italian and Mathematics at multiple time points, and up to 18 months after the end of training, we uncovered increasing group differences over time in Italian grades, with an advantage of the SOM-trained group over the Control-trained group. Few studies have managed to collect this type of information. Among those that have, Goldin and collaborators[58] did uncover the impact of multiple EFs training in 6-7 years old when considering a composite grade for school mathematics and literacy/language. It remains unclear given the use of such a composite whether both literacy and mathematics improved. While several experiments indicate enhanced school mathematics after EF training, grades for mathematics did not differ across training in our experiment. This null effect is difficult to interpret as the control game Scratch may have positively impacted school mathematics[59]. The greater improvement in school language grades with SOM underlines the far transfer effects of our intervention. Interestingly, the uncovered effects on school grades in Italian were delayed in time, being statistically significant only 18 months after the end of the intervention. This is in line with studies indicating that positive effects of training might be detectable in everyday life activities only after a significant delay (i.e. six months to one year)[60,61].

Overall, this study highlights how training with SOM not only improved attentional control and reading skills but also other fundamental components of the executive system. Following the footsteps of previous studies that trained attentional control and working memory capacity ([14,19,61–64], for two extensive meta-analyses of the benefits of working memory programs see[65,66]), SOM was purposefully designed to train multiple executive components in addition to attentional control. Interestingly, SOM consists of a variety of training tasks, in order to obtain overall improvement in domain-general abilities and higher chances of generalization to untrained tasks. Another design feature of SOM is to encompass challenging and variable activities integrated within a main story frame, requiring children to continuously adapt their behaviour, while creating a fair sense of appropriation and competition. This departs from many cognitive training programs that tend to use just a few, stand-alone mini-tasks[67,68]. In this regard, the overall appreciation for the training activities was high in terms of entertainment, engagement and motivation for the children. Importantly, SOM and the Control game were comparable in that regard (see Supplementary Information, Section 2 – Results: Feedback questionnaire). Finally, thanks to the high adaptability of the video game, each player pursues a different training path: for example, the children who performed worse at the beginning of the training in a specific mini-game that trains inhibition, will be exposed to more exercises that tap into that specific cognitive function. Vice versa, children that already possessed adequate levels of a specific skill will be trained less on that specific skill in favour of others.

There are a few limitations of our study we need to acknowledge. In the current version of SOM, data from the participants' game play were not recorded. A few studies report that game-play performance during training correlates with performance in standardized neuropsychological tests (e.g., NEPSY-II) assessing visuo-spatial, attentional and language skills[69]. In addition, Franceschini and Bertoni[53] showed that only the children who improved through the game-play obtained significant transfer effects in untrained, cognitively-related tasks such as reading skills. In the present experimental game, in-game data have the potential to disentangle the effects of the more action-like video game play (i.e., timed activities that involve response under time pressure, divided attention, timely shifts between focused and divided) from the effects of the mini-games that do not have any specific action game constraints. Further studies are thus required in order to fully exploit video game play as a way of assessing how changes in various cognitive skills relate to reading gains. These further investigations will shed critical light on the factors that most successfully predict the outcome of training for reading skills. It is also of note that, while the present results form a promising basis for more individually based training, future research should more carefully address the cognitive profile of each child beyond language and into the executive domain, with the aim of reaching the best training algorithm on the basis of the pattern of strengths and weaknesses of each child[70]. Finally, it will be important to assess SOM efficacy in children speaking other languages, such as English or French, to understand whether benefits are seen in an opaque orthography, in which phonological components are expected to play a larger role in reading acquisition and automatization[71,72].

Method

Participants

151 typically developing Italian children were recruited from eight classes from 3rd, 4th, 6th or 7th grades (8-12 years old) within a northern Italy public school. One of the two classes included at each grade were randomly assigned to either the training or the control group. Although all children within each class were treated similarly, criteria for data analyses inclusion were: (i) normal or corrected to normal visual acuity; (ii) no Attention Deficit Hyperactivity Disorder diagnosis; (iii) no diagnosis of Learning Disorder or no reading delay in word and non-word reading tests; (iv) no diagnosis of Intellectual Disability. Children with learning disorders (n=21) or with borderline intellectual functioning (n=10) participated in the study like any other classmate, however their data were not included in the analysis sample leading to a final sample of N=79, 39 females in the experimental group and N = 72, 41 females in the control group.

All the children's parents gave written informed consent after a description of the research study, in accordance with the principles of the Declaration of Helsinki; the University of Trento ethic committee approved the research protocol (p. n. 2019/048). At the end of the school year, participants were thanked for their time and effort with a small gift.

Study Design

The study included five phases, a first data and grade collection time point, followed by the intervention and then two other data and grades collection time points; finally, grades were collected at two later time points.

Procedures

Testing: Children were tested at school. Baseline (T1) assessments were completed one week prior to the start of treatment, and outcome assessments (T2 and T3) one week and 6-months after the end of the treatment. Tasks were administered in a pseudorandom manner; IQ (Raven's CPM[41]) and comprehension abilities (MT Reading Task[43,44]) were always assessed in a collective session, whereas the other tests were administered individually in a quiet room. Other tasks included attentional control (Modified Bells Test[50]), verbal and visual working memory (Digit and Visuo-spatial span[73], planning (Tower of London[51]) and literacy skills (Word, Non-word and Text reading[42–44]; Word and Non-word writing[42]). A detailed description of the measures is presented in Supplementary Information 3 - Tools. Note that IQ was only assessed once at T1; otherwise, T1 and T2 assessment used identical test materials. The assessment battery at T3 consisted of a selection of tests previously used (for reading, attention and planning) and two additional tests (text-reading task that measures reading abilities of meaningful material and a novel barrage task that evaluates visuo-spatial attention[52]). Testing was conducted by three different clinicians, two of them blinded to children assignments and one of them (AP) being the experimental game expert. Note that although we report data over the whole sample, results from the sample tested by blinded experimenters only did not differ from the whole sample (See SI Section 2 – Results: Controlling for possible testing confounds).

Classification of the reading errors.

Reading performance was further analysed adopting the error classification proposed by Hendricks and Kolk[45] and applied to Italian by Trenta and colleagues[46]. Reading behaviours of each student at the three time points when reading the MT text passage were analysed by focusing on the syllabic level. The passages are normed for each grade and thus contain different numbers of syllables (3° grade: 305 syllables; 4° grade syllables: 297; 6° grade: 592 syllables; 7° grade: 596 syllables).

The classification focused on the contrast between two types of reading behaviours: sounding-out and word-substitution behaviours. The former refers to a particular behaviour in which the child tries to read the target word through several attempts (e.g. the child has to read "crocodile" and instead he/she reads "cr.cro.crocodile"). In contrast, the latter indicates a substitution of the target word with another word (e.g.

the child has to read “mood” and he/she reads “moon”). The classification includes a third category, i.e., Residuals, combining reading errors or dysfluencies that are not comprised in the other two main categories. A brief description of each subcategory of the classification used from Trenta and colleagues[46] (along with examples of errors made in our sample of Italian typically reading children) can be found in the Supplementary Information Section 4 – Errors classification. For examples in English and Dutch, please refer to Hendriks & Kolk[45].

At pre-test, the proportion of reading errors were rather equally distributed between sounding-out behaviours (SOM: 48.6% and Control: 44.6%), and word-substitutions (SOM: 41.9% and Control: 34.1%) with less than 1% of residual errors (given this small percentage, this type of errors was not analyzed further). Fleiss’ kappa procedure[74] was used to determine reliability. Scoring was carried out by AP and two additional independent raters, who analyzed 20% of the data, randomly selected. All the raters were blind to group allocation and to the time points in which data were collected. The reading errors classification showed very high inter-rater reliability, with kappa values (95% CI) of 0.81 (0.80, 0.81), 0.85 (0.84, 0.85), 0.45 (0.44, 0.45) for sounding-out, word-substitution, and residuals, respectively.

Grade Collection: School grades were obtained directly from school teachers. The same teacher provided grades for both the experimental and the control class they had been assigned. Maths and Italian grades corresponded to those in the report cards containing the average grade for each semester.

Intervention Treatment: Training was administered at school during school time with duration and frequency (1 hour per session, twice a week for 6 weeks) matched across the two intervention groups. All children trained for at least 11 hours out of the 12 hours of gameplay (M= 11.40; SD= 0.27) to be included in the analyses. The experimental group trained on Skies of Manawak (SOM); those in the control group received a computer-based training on coding using *Scratch*, a programming language developed by the Lifelong Kindergarten (MIT) Group, which allows children to create interactive stories, games, animations, and simulations[75]. Each game was administered under the supervision of an expert experimenter ensuring that participants were focused and that they remained motivated, encouraging them whenever needed. The SOM expert was AP; the Scratch expert was a computer science teacher also familiar with the school. It has to be noted that none of the activities required reading: in SOM, text-to-speech audio clips were created for each dialogue, providing a full dubbing of the game, while in Scratch all the explanations regarding the software's functionalities and the different task assignments were provided orally by the teacher. Furthermore, at the end of the training all children completed a questionnaire aimed at investigating their motivation, the perceived difficulty and appreciation of the video game, as well as their personal evaluation of the training. No statistically significant differences in the level of appreciation between the two training activities were found (p-value = .243; for further details, see SI 2 – feedback questionnaire).

Additional information regarding SOM and the control activity, Scratch, can be found in the Supplementary Information 5 – Training regimens.

Data availability

We are planning to upload the paper, SI and its corresponding data under the OpenScience Framework.

REFERENCES

1. Parsons, S. & Bynner, J. *Research report over the lifecourse*. (2007).
2. Pape, K., Bjørngaard, J. H., Westin, S., Holmen, T. L. & Krokstad, S. Reading and writing difficulties in adolescence and later risk of welfare dependence. A ten year follow-up, the HUNT Study, Norway. *BMC Public Health* **11**, (2011).
3. Carneiro, R. & Gordon, J. Warranting our Future: literacy and literacies. *Eur. J. Educ.* **48**, 476–497 (2013).
4. Snowling, M. Dyslexia: A language learning impairment. *J. Br. Acad.* **2**, 43–58 (2014).
5. Ziegler, J. C. & Goswami, U. Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychol. Bull.* **131**, 3–29 (2005).
6. Elhassan, Z., Crewther, S. G. & Bavin, E. L. The Contribution of Phonological Awareness to Reading Fluency and Its Individual Sub-skills in Readers Aged 9- to 12-years. *Front. Psychol.* **8**, 533–533 (2017).
7. Vellutino, F. R., Fletcher, J. M., Snowling, M. J. & Scanlon, D. M. Specific reading disability (dyslexia): What have we learned in the past four decades? *J.*

Child Psychol. Psychiatry **45**, 2–40 (2004).

8. Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K. & Facoetti, A. A causal link between visual spatial attention and reading acquisition. *Curr. Biol.* **22**, 814–819 (2012).
9. Grainger, J., Tydgat, I. & Isselé, J. Crowding affects letters and symbols differently. *J. Exp. Psychol. Hum. Percept. Perform.* **36**, 673–688 (2010).
10. Grainger, J., Dufau, S. & Ziegler, J. C. A Vision of Reading. *Trends Cogn. Sci.* **20**, 171–179 (2016).
11. Strijkers, K., Bertrand, D. & Grainger, J. Seeing the Same Words Differently : The Time Course of Automaticity and Top – Down Intention in Reading. *J. Cogn. Neurosci.* **27**, 1542–1551 (2015).
12. Ruz, M. & Nobre, A. C. Attention Modulates Initial Stages of Visual Word Processing. *J. Cogn. Neurosci.* **20**, 1727–1736 (2008).
13. Verhoeven, L., Reitsma, P. & Siegel, L. S. Cognitive and linguistic factors in reading acquisition. *Read. Writ.* **24**, 387–394 (2011).
14. Karbach, J., Strobach, T. & Schubert, T. Adaptive working-memory training benefits reading, but not mathematics in middle childhood. *Child Neuropsychol.* **21**, 285–301 (2015).
15. Loosli, S. V., Buschkuehl, M., Perrig, W. J. & Jaeggi, S. M. Working memory training improves reading processes in typically developing children. *Child Neuropsychol.* **18**, 62–78 (2012).
16. Siegel, L. S. & Ryan, E. B. The Development of Working Memory in Normally Achieving and Subtypes of Learning Disabled Children. *Child Dev.* **60**, 973–980 (1989).
17. Gathercole, S. E., Willis, C. S., Baddeley, A. D. & Emslie, H. The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory* **2**, 103–127 (1994).
18. Baddeley, A. Working memory and language: an overview. *J. Commun. Disord.* **36**, 189–208 (2003).
19. Henry, L. A., Messer, D. J. & Nash, G. Testing for Near and Far Transfer Effects with a Short, Face-to-Face Adaptive Working Memory Training Intervention in Typical Children. *Infant Child Dev.* **23**, 84–103 (2014).
20. Blair, C. & Razza, R. P. Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child*

Dev. **78**, 647–663 (2007).

21. Colè, P., Duncan, L. G. & Blaye, A. Cognitive flexibility predicts early reading skills. *Front. Psychol.* **5**, 1–8 (2014).
22. Cartwright, K. B. *et al.* Cognitive flexibility deficits in children with specific reading comprehension difficulties. *Contemp. Educ. Psychol.* **50**, 33–44 (2017).
23. Cartwright, K. B., Marshall, T. R., Huemer, C. M. & Payne, J. B. Executive function in the classroom: Cognitive flexibility supports reading fluency for typical readers and teacher-identified low-achieving readers. *Res. Dev. Disabil.* **88**, 42–52 (2019).
24. Christopher, M. E. *et al.* Predicting word reading and comprehension with executive function and speed measures across development: A latent variable analysis. *J. Exp. Psychol. Gen.* **141**, 470–488 (2012).
25. Reiter, A., Tucha, O. & Lange, K. W. Executive functions in children with dyslexia. *Dyslexia* **11**, 116–131 (2005).
26. Locascio, G., Mahone, M. E., Eason, S. & Cutting, L. E. Executive Dysfunction Among Children With Reading Comprehension Deficits. *J. Learn. Disabil.* **43**, 441–454 (2010).
27. Bejjanki, V. R. *et al.* Action video game play facilitates the development of better perceptual templates. *Proc. Natl. Acad. Sci.* **111**, 16961–6 (2014).
28. Cardoso-Leite, P. & Bavelier, D. Video game play, attention, and learning. *Curr. Opin. Neurol.* **27**, 185–191 (2014).
29. Bavelier, D., Green, C. S., Pouget, A. & Schrater, P. Brain plasticity through the life span: learning to learn and action video games. *Annu. Rev. Neurosci.* **35**, 391–416 (2012).
30. Bediou, B. *et al.* Meta-Analysis of Action Video Game Impact on Perceptual , Attentional , and Meta-Analysis of Action Video Game Impact on Perceptual , Attentional , and Cognitive Skills. *Psychol. Bull.* (2018) doi:10.1037/bul0000130.
31. Dye, M. W. G., Green, C. S. & Bavelier, D. The development of attention skills in action video game players. *Neuropsychologia* **47**, 1780–1789 (2009).
32. Franceschini, S. *et al.* Action video games make dyslexic children read better. *Curr. Biol.* **23**, 462–466 (2013).
33. Franceschini, S. *et al.* Action video games improve reading abilities and visual-to-auditory attentional shifting in English-speaking children with dyslexia. *Sci. Rep.* **7**, 1–12 (2017).

34. Bavelier, D., Green, C. S. & Seidenberg, M. S. Cognitive development: Gaming your way out of dyslexia? *Curr. Biol.* **23**, R282–R283 (2013).
35. Łuniewska, M. *et al.* Neither action nor phonological video games make dyslexic children read better. *Sci. Rep.* **8**, 1–11 (2018).
36. Peters, J. L., Losa, L. D., Bavin, E. L. & Crewther, S. G. Neuroscience and Biobehavioral Reviews Efficacy of dynamic visuo-attentional interventions for reading in dyslexic and neurotypical children : A systematic review. *Neurosci. Biobehav. Rev.* **100**, 58–76 (2019).
37. Cardoso-Leite, P., Joessel, A. & Bavelier, D. Games for enhancing cognitive abilities. in *Handbook of Game-based Learning* (eds. Plass, J., Mayer, R. & Homer, B.) (MIT Press, 2020).
38. Anguera, J. A. & Gazzaley, A. Video games, cognitive exercises, and the enhancement of cognitive abilities. *Curr. Opin. Behav. Sci.* **4**, 160–165 (2015).
39. Bavelier, D. & Green, C. S. Enhancing Attentional Control : Lessons from Action Video Games. *Neuron* **104**, 147–163 (2019).
40. Franceschini, S. *et al.* Shall We Play a Game ? : Improving Reading Through Action Video Games in Developmental Dyslexia. *Curr. Dev. Disord. Rep.* 318–329 (2015) doi:10.1007/s40474-015-0064-4.
41. Raven, J. J. Raven Progressive Matrices. *Handb. Nonverbal Assess.* 223–237 (2003) doi:10.1007/978-1-4615-0153-4_11.
42. Sartori, G., Remo, J. & Tressoldi, P. E. DDE-2,Batteria per la valutazione della dislessia e della disortografia evolutiva-2. (1995).
43. Cornoldi, C. & Colpo, G. Prove di Lettura MT per la Scuola Elementare -2. (1998).
44. Cornoldi, C. & Colpo, G. Prove di Lettura MT per la scuola media inferiore -2. (1995).
45. Hendriks, A. W. & Kolk, H. H. J. Strategic Control in Developmental Dyslexia. *Cogn. Neuropsychol.* **14**, 321–366 (1997).
46. Trenta, M., Benassi, M., Di Filippo, G., Pontillo, M. & Zoccolotti, P. Developmental dyslexia in a regular orthography: Can the reading profile be reduced to strategic control? *Cogn. Neuropsychol.* **30**, 147–171 (2013).
47. Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* **108**, 204–256 (2001).

48. Perry, C., Ziegler, J. C. & Zorzi, M. CDP++.Italian: Modelling Sublexical and Supralexical Inconsistency in a Shallow Orthography. *PLOS ONE* **9**, e94291–e94291 (2014).
49. Rakhlin, N. V., Mourgues, C., Cardoso-Martins, C., Kornev, A. N. & Grigorenko, E. L. Orthographic processing is a key predictor of reading fluency in good and poor readers in a transparent orthography. *Contemp. Educ. Psychol.* **56**, 250–261 (2019).
50. Biancardi, A. & Stoppa, E. Il Test delle Campanelle Modificato (TCM): una proposta di studio per l'attenzione in età evolutiva. *Psichiatria dell'infanzia e dell'Adolescenza* 73–84 (1997).
51. Sannio Fancello, G., Vio, C. & Cianchetti, C. Tower of London test. (2006).
52. Bisiacchi, P., Cendron, M., Gugliotta, M., Tressoldi, P. E. & Vio, C. BVN 5-11 - Batteria di valutazione neuropsicologica per l'età evolutiva. (2005).
53. Franceschini, S. & Bertoni, S. Improving action video games abilities increases the phonological decoding speed and phonological short-term memory in children with developmental dyslexia. *Neuropsychologia* **130**, 100–106 (2019).
54. Double, K. S., Mcgrane, J. A., Stiff, J. C. & Hopfenbeck, T. N. The importance of early phonics improvements for predicting later reading comprehension. *Br. Educ. Res. J.* (2019) doi:10.1002/berj.3559.
55. Vidyasagar, T. R. & Pammer, K. Dyslexia: a deficit in visuo-spatial attention, not in phonological processing. *Trends Cogn. Sci.* **14**, 57–63 (2010).
56. Facoetti, A. *et al.* The relationship between visuo-spatial attention and nonword reading in developmental dyslexia. (2006) doi:10.1080/02643290500483090.
57. Gabrieli, J. D. E. & Norton, E. S. Reading Abilities : Importance of Visual-Spatial Attention Speciation Genetics : Reinforcement by Shades and Hues. *Curr. Biol.* **22**, R298–R299 (2012).
58. Goldin, A. P. *et al.* Far transfer to language and math of a short software-based gaming intervention. *Proc. Natl. Acad. Sci.* **111**, 6443–6448 (2014).
59. Calao, L. A., Moreno-León, J., Correa, H. E. & Robles, G. Developing Mathematical Thinking with Scratch BT - Design for Teaching and Learning in a Networked World. in (eds. Conole, G., Klobučar, T., Rensing, C., Konert, J. & Lavoué, E.) 17–27 (Springer International Publishing, 2015).
60. Blair, C. & Raver, C. C. Closing the achievement gap through modification of neurocognitive and neuroendocrine function: Results from a cluster randomized

controlled trial of an innovative approach to the education of children in kindergarten. *PLoS ONE* **9**, (2014).

61. Holmes, J., Gathercole, S. E. & Dunning, D. L. Adaptive training leads to sustained enhancement of poor working memory in children. *Dev. Sci.* **12**, 9–15 (2009).
62. Bergman-Nutley, S. & Klingberg, T. Effect of working memory training on working memory, arithmetic and following instructions. *Psychol. Res.* **78**, 869–877 (2014).
63. Judd, N. & Klingberg, T. Training spatial cognition enhances mathematical learning in a randomized study of 17,000 children. *Nat. Hum. Behav.* (2021) doi:10.1038/s41562-021-01118-4.
64. Johann, V. E. & Karbach, J. Effects of game-based and standard executive control training on cognitive and academic abilities in elementary school children. *Dev. Sci.* **23**, e12866–e12866 (2020).
65. Spencer-Smith, M. & Klingberg, T. Benefits of a working memory training program for inattention in daily life: A systematic review and meta-analysis. *PLoS ONE* **10**, 1–18 (2015).
66. Berger, E., Fehr, E., Hermes, H., Schunk, D. & Winkel, K. The Impact of Working Memory Training on Children's Cognitive and Noncognitive Skills (2020).
67. Burgers, C., Eden, A., Van Engelenburg, M. D. & Buningh, S. How feedback boosts motivation and play in a brain-training game. *Comput. Hum. Behav.* **48**, 94–103 (2015).
68. Eichenbaum, A., Bavelier, D. & Green, C. S. Video Games: Play That Can Do Serious Good. *Am. J. Play* **7**, (2014).
69. Martinovic, D., Burgess, G. H., Pomerleau, C. M. & Marin, C. Comparison of children's gaming scores to NEPSY-II scores: Validation of computer games as cognitive tools. *Comput. Hum. Behav.* **49**, 487–498 (2015).
70. Mishra, J., Anguera, J. A. & Gazzaley, A. Video Games for Neuro-Cognitive Optimization. *Neuron* **90**, 214–218 (2016).
71. Goswami, U., Ziegler, J. C. & Richardson, U. The effects of spelling consistency on phonological awareness: A comparison of English and German. *J Exp. Child Psychol.* 2016–2020 (2005) doi:10.1016/j.jecp.2005.06.002.
72. Defior, S. Phonological Awareness and Learning to Read: A Cross-Linguistic

Perspective. in *Handbook of Children's Literacy* (eds. Nunes, T. & Bryant, P.) 631–649 (Springer Netherlands, 2004). doi:10.1007/978-94-017-1731-1_33.

73. Pazzaglia, F., Mammarella, I. C., Toso, C. & Cornoldi, C. BVS - Corsi - Batteria per la valutazione della memoria visiva e spaziale. (2008).
74. Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychol. Bull.* **76**, 378–382 (1971).
75. Resnick, M., Maloney, J. & Hernández, A. M. Scratch: Programming for Everyone Scratch : Programming for Everyone. *Commun. ACM* **52**, 60–67 (2009).