



**HAL**  
open science

## Selective Inhibition of Mirror Invariance for Letters Consolidated by Sleep Doubles Reading Fluency

Ana Raquel Torres, Natália B Mota, Nery Adamy, Angela Naschold, Thiago Z Lima, Mauro Copelli, Janaina Weissheimer, Felipe Pegado, Sidarta Ribeiro

► **To cite this version:**

Ana Raquel Torres, Natália B Mota, Nery Adamy, Angela Naschold, Thiago Z Lima, et al.. Selective Inhibition of Mirror Invariance for Letters Consolidated by Sleep Doubles Reading Fluency. *Current Biology - CB*, 2021, 31, pp.742 - 752.e8. 10.1016/j.cub.2020.11.031 . hal-03226367

**HAL Id: hal-03226367**

**<https://amu.hal.science/hal-03226367>**

Submitted on 14 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

## Selective Inhibition of Mirror Invariance for Letters Consolidated by Sleep Doubles Reading Fluency

### Highlights

- Targeted training prevents mirror confusion for letters (b = d) in first graders
- Sleep boosts the magnitude, automaticity, and duration of this learning
- Training followed by sleep doubles reading fluency in first graders

### Authors

Ana Raquel Torres, Natália B. Mota, Nery Adamy, ..., Janaina Weissheimer, Felipe Pegado, Sidarta Ribeiro

### Correspondence

felipe.pegado@univ-amu.fr (F.P.),  
sidartaribeiro@neuro.ufrn.br (S.R.)

### In Brief

Torres et al. show that mirror confusion for letters (e.g., b = d) represents a major leash for reading fluency in first graders, and how it can be efficiently solved with a targeted training. The study also demonstrates the tremendous impact of post-training naps in the magnitude, automaticity, and duration of the targeted learning. The consolidated learning increased reading fluency by a factor of two.



Article

# Selective Inhibition of Mirror Invariance for Letters Consolidated by Sleep Doubles Reading Fluency

Ana Raquel Torres,<sup>1</sup> Natália B. Mota,<sup>1,2</sup> Nery Adamy,<sup>1</sup> Angela Naschold,<sup>4</sup> Thiago Z. Lima,<sup>1</sup> Mauro Copelli,<sup>2</sup> Janaina Weissheimer,<sup>1,3</sup> Felipe Pegado,<sup>5,6,7,\*</sup> and Sidarta Ribeiro<sup>1,7,8,\*</sup>

<sup>1</sup>Laboratory of Memory, Sleep and Dreams, Brain Institute, Federal University of Rio Grande do Norte, Av. Nascimento de Castro 2155, Natal 59056-450, Brazil

<sup>2</sup>Department of Physics, Federal University of Pernambuco, Av. Jorn. Aníbal Fernandes s/n, Recife 50740-540, Brazil

<sup>3</sup>Department of Modern Foreign Languages and Literatures, Federal University of Rio Grande do Norte, Av. Sen. Salgado Filho s/n, Natal 59078-970, Brazil

<sup>4</sup>Department of Education, CERES, Federal University of Rio Grande do Norte, Rua Joaquim Gregório 296, Caicó 59300-000, Brazil

<sup>5</sup>Laboratory of Cognitive Psychology, Institute for Language Communication and the Brain, CNRS and Aix-Marseille University- 3 place Victor Hugo, Marseille 13331, France

<sup>6</sup>Department of Brain and Cognition, KU Leuven, Tiensestraat 102, Leuven 3000, Belgium

<sup>7</sup>Senior author

<sup>8</sup>Lead Contact

\*Correspondence: [felipe.pegado@univ-amu.fr](mailto:felipe.pegado@univ-amu.fr) (F.P.), [sidartaribeiro@neuro.ufrn.br](mailto:sidartaribeiro@neuro.ufrn.br) (S.R.)

<https://doi.org/10.1016/j.cub.2020.11.031>

## SUMMARY

Mirror invariance is a visual mechanism that enables a prompt recognition of mirror images. This visual capacity emerges early in human development, is useful to recognize objects, faces, and places from both left and right perspectives, and is also present in primates, pigeons, and cephalopods. Notwithstanding, the same visual mechanism has been suspected to be the source of a specific difficulty for a relatively recent human invention—reading—by creating confusion between mirror letters (e.g., b-d in the Latin alphabet). Using an ecologically valid school-based design, we show here that mirror invariance represents indeed a major leash for reading fluency acquisition in first graders. Our causal approach, which specifically targeted mirror invariance inhibition for letters, in a synergic combination with post-training sleep to increase learning consolidation, revealed unprecedented improvement in reading fluency, which became two-times faster. This gain was obtained with as little as 7.5 h of multisensory-motor training to distinguish mirror letters, such as “b” versus “d.” The magnitude, automaticity, and duration of this mirror discrimination learning were greatly enhanced by sleep, which keeps the gains perfectly intact even after 4 months. The results were consistently replicated in three randomized controlled trials. They not only reveal an extreme case of *cognitive plasticity* in humans (i.e., the inhibition in just 3 weeks of a ~25-million-year-old visual mechanism), that allows adaptation to a cultural activity (reading), but at the same time also show a simple and cost-effective way to unleash the reading fluency potential of millions of children worldwide.

## INTRODUCTION

Every year, millions of children across the world engage in learning to read and write, but more than half of them never become fluent readers, remaining unable to fully comprehend textual communication.<sup>1</sup> The problem is further complicated by dyslexia, which affects 5%–10% of the world population,<sup>2,3</sup> and by the slow reading speed achieved when literacy is acquired later in life,<sup>4</sup> impacting reading comprehension.<sup>5</sup> Thus, the development of new efficient strategies to improve reading fluency in the initial stage of literacy acquisition is timely.

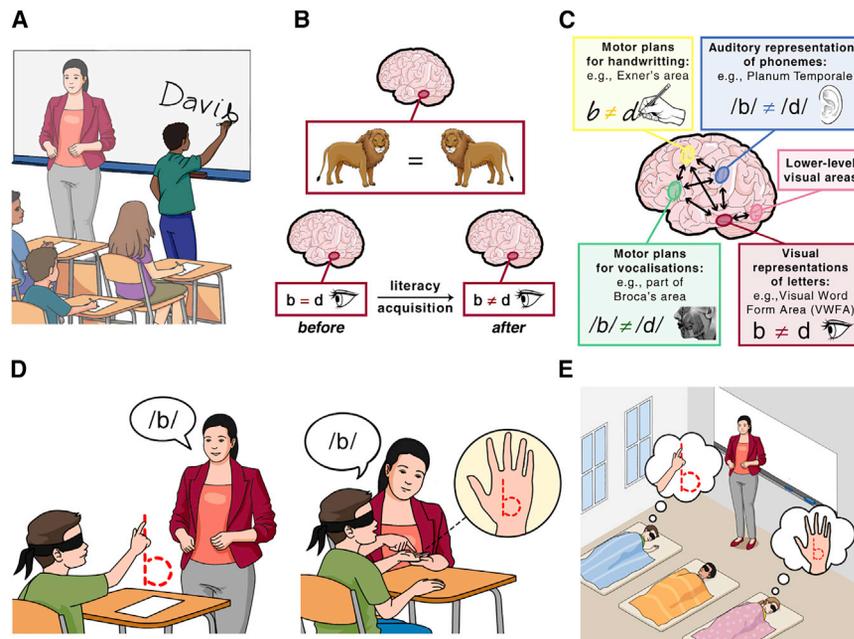
Decades of research on the link between letter discrimination and reading<sup>6,7</sup> have established that letter discrimination is enhanced by teaching children to attend to critical features of the stimuli during discrimination,<sup>8,9</sup> and that letter discrimination

training on highly confusable letters increases learning relative to the same training on less confusable letters.<sup>10</sup> Indeed, training children to discriminate visually similar letters and map them to corresponding phonemes is the proposed explanation for the efficacy of the highly replicated work of GraphoGame.<sup>11,12</sup>

While the biological and cognitive basis of literacy acquisition has been progressively understood and applied,<sup>2,4,13–19</sup> its translation to efficient educational practices worldwide remains “a bridge too far”<sup>20</sup> in most educational contexts.<sup>21</sup> To address this challenge, we focused here on one visual mechanism that could interfere with reading fluency acquisition: mirror invariance.

Mirror invariance enables the visual recognition of images from either the left or right profiles (mirror images). This capacity emerges early in human development<sup>22</sup> and is shared with other





**Figure 1. Study Rationale**

(A) Confusion with the correct orientation of letters is common among first graders,<sup>26</sup> with writing errors being paralleled by visual perception errors.<sup>29</sup> (B) Mirror invariance in high-level visual cortex is suspected to be the main source of mirror confusion for letters<sup>27,28,30–32</sup> and could hamper reading fluency acquisition. (C) Our hypothetical learning model of how mirror invariance for letters is slowly overcome in school in a non-systematic way (via multisystem mappings) (modified from Pegado et al.<sup>33</sup>). (D) Three RCTs were performed here to probe the impact of a brief targeted intervention (30 min/day for 3 weeks) on mirror invariance for letters and on reading fluency. Multisensory-motor activities were used, such as “air-writing” (left) and tactile perception of letter traces (right), while listening to and/or producing letter sounds. (E) To test whether sleep could improve learning consolidation<sup>34–37</sup> of the targeted training, one group of participants (T+S group) took post-training naps (up to 2 h).

animals, from monkeys<sup>23</sup> and pigeons<sup>24</sup> to cephalopods.<sup>25</sup> While mirror invariance is useful in the natural world, it is suspected to be the source of “mirror confusion” for letters at school,<sup>26–29</sup> a frequent and pervasive difficulty in the beginning of literacy acquisition (Figure 1A).

In the past decades, studies have shed light on the putative cognitive and neural basis of mirror confusion.<sup>38</sup> They show the existence of equivalent responses in the human visual cortex for mirror versions of the same images, be they objects,<sup>30</sup> faces,<sup>31</sup> places,<sup>32</sup> or iconic images<sup>27,28</sup> (Figure 1B). Confusion due to mirror invariance can persist even after 2–3 years of literacy practice,<sup>26,29</sup> whereas for functional illiterates it may last throughout life,<sup>39,40</sup> showing how deeply mirror invariance is rooted in the human visual system.

Despite this, skilled readers overcome mirror invariance, exhibiting fast discrimination between left/right orientation of letters, both at the perceptual level<sup>41</sup> and at the brain responses’ level.<sup>15,27,28</sup> It is now well established that literates show selective responses for orthographic stimuli in a restricted region of the left cerebral hemisphere, in the ventral part of the temporal visual cortex—the so-called Visual Word Form Area (VWFA)<sup>42</sup>—where mirror discrimination for letters<sup>28</sup> and letter strings<sup>27</sup> is observed, while preserving mirror invariance for other types of images. In fact, the VWFA represents a key node in the orthographic processing flow, functioning as a bottleneck in the communication between early visual areas with language areas.<sup>43</sup> While beginner readers exhibit a letter-by-letter processing style, expert readers can process several letters-in-words in parallel, with the VWFA processing one whole word at a time<sup>44</sup> and possibly more,<sup>45,46</sup> depending on the task.<sup>47</sup> Literacy acquisition also induces early (<200 ms) left-hemisphere lateralization of neural responses to visual presentations of the learned script<sup>48</sup> and also its mirror discrimination,<sup>15</sup> suggesting that these are quite automatic processes, in accordance to the fast responses observed at the behavioral level.<sup>41</sup> By learning grapheme-phoneme correspondences

(i.e., mapping visual and auditory representations), literates activate the auditory cortex when processing orthographic visual stimuli, but they also exhibit VWFA activation during an auditory lexical decision task.<sup>4</sup> Other mappings across neural systems are also promoted by literacy, such as that between the visual representation of letters and the corresponding writing gestures,<sup>49,50</sup> with the behavioral consequence that mirror confusion in the writing domain is paralleled with mirror confusion in the visual domain.<sup>29</sup>

We have previously postulated that readers typically overcome the mirror invariance of the visual system through multisensory-motor mappings of letter representations.<sup>33</sup> We had highlighted that these mappings could be particularly important to deal with the most problematic letters for the visual system, i.e., mirror letters (in the Latin alphabet, the lowercase letters b-d or p-q),<sup>51,52</sup> especially knowing that top-down information from other systems could disentangle mirror-letter pairs with discriminative information: different sounds, different writing gestures, and different vocalizations for each of the two letters in a mirror-letter pair.<sup>33</sup> (Figure 1C).

However, no systematic training to differentiate mirror letters is typically promoted at school. Under these conditions, the audio-visual-writing mappings for mirror letters are probably only built sparsely and slowly, as suggested by the long persistence (2–3 years) of mirror confusion for letters after the beginning of literacy training.<sup>26,29,39,41</sup> We have thus predicted that by systematically targeting mirror-letter distinction using a “multisystem mappings” approach, we would improve mirror discrimination for letters, accelerating their visual identification, and as a consequence, increasing reading fluency.<sup>33</sup>

Since mirror invariance is deeply rooted in our visual system, even an efficient targeted learning would require sufficient consolidation. Intensive or prolonged training regimes can produce learning consolidation but are not cost-effective. One physiological factor known to improve learning consolidation is sleep,<sup>34,35</sup> and there is mounting evidence that even post-

training naps can improve the short-term retention of declarative contents learned at school.<sup>36,37,53</sup>

Therefore, we designed a combined intervention associating target training—to address mirror invariance for letters using multisensory-motor activities (Figure 1D) with post-training naps, to enhance memory consolidation (Figure 1E). To test the impact of the intervention on mirror invariance for letters and reading fluency, we carried out three randomized controlled trials (RCTs), one pre-registered,<sup>54</sup> with 5- to 7-year-old children in Natal (Brazil), learning to read and write in Brazilian Portuguese. We evaluated both short- and long-term effects of the 3-week intervention, including the impact of post-training naps.

In each RCT replica, we applied our same-different *visual* task, as used in previous behavioral<sup>41</sup> and neuroimaging studies<sup>15,27,28</sup>: the participants were requested to distinguish the mirror presentations of stimuli (Figure 2, top). We also applied a writing task, performed on paper with pencil, where participants were asked to reproduce letters (Figure 5, top). Last, we employed the same reading fluency task as previously,<sup>4</sup> measuring the number of stimuli (words or pseudowords) read in 1 min (Figure 6, top).

For each task, three data points were collected: baseline, short-term (immediately after the intervention), and long-term (~120 days after the end of the intervention), except for the reading fluency task that was only accessed at the long-term measure, since the participants were not able to read yet at the initial assessments. In each RCT, participants were randomly assigned to one of the two training groups, i.e., Training (T) or Training + Sleep (T+S), or one of the two control groups, i.e., Control (C) and Active Control (AC) groups. In the C group, participants did not receive any training, and in the Active Control (AC) group participants received similar multisensory-motor activities but played with the symmetrical letters such as A and X, therefore avoiding training on mirror discrimination for letters but controlling for non-specific socio-emotional factors that were present in the two training groups.

## RESULTS

The results in all three RCT replicas consistently show *short-term* improvement in mirror discrimination for letters in the two training groups but not in the two control groups, revealing extreme cognitive plasticity in humans, quickly mitigating the expression of a visual mechanism that is putatively active in humans for millions of years. Indeed, in the visual task, participants in the T and T+S groups made very few errors when judging two subsequent letters in different orientations, and responded *faster* (Figures 2A and 2B; Data S1A and S1B), which was not the case in the control groups (C and AC). Note the strong impact of sleep further increasing the speed of visual discrimination (T+S versus T group). For the writing task, almost all trained participants (T and T+S groups) performed the task without errors, again when assessed just after the intervention (Figure 5, bottom; Data S1A and S1B; detailed statistics in Data S1C, [normality test], Data S1D [samples sizes at each time point, means, and standard deviations], and Table S1 [effect sizes]. These short-term results reveal an example of major cognitive plasticity in humans induced by culture, quickly mitigating the behavioral

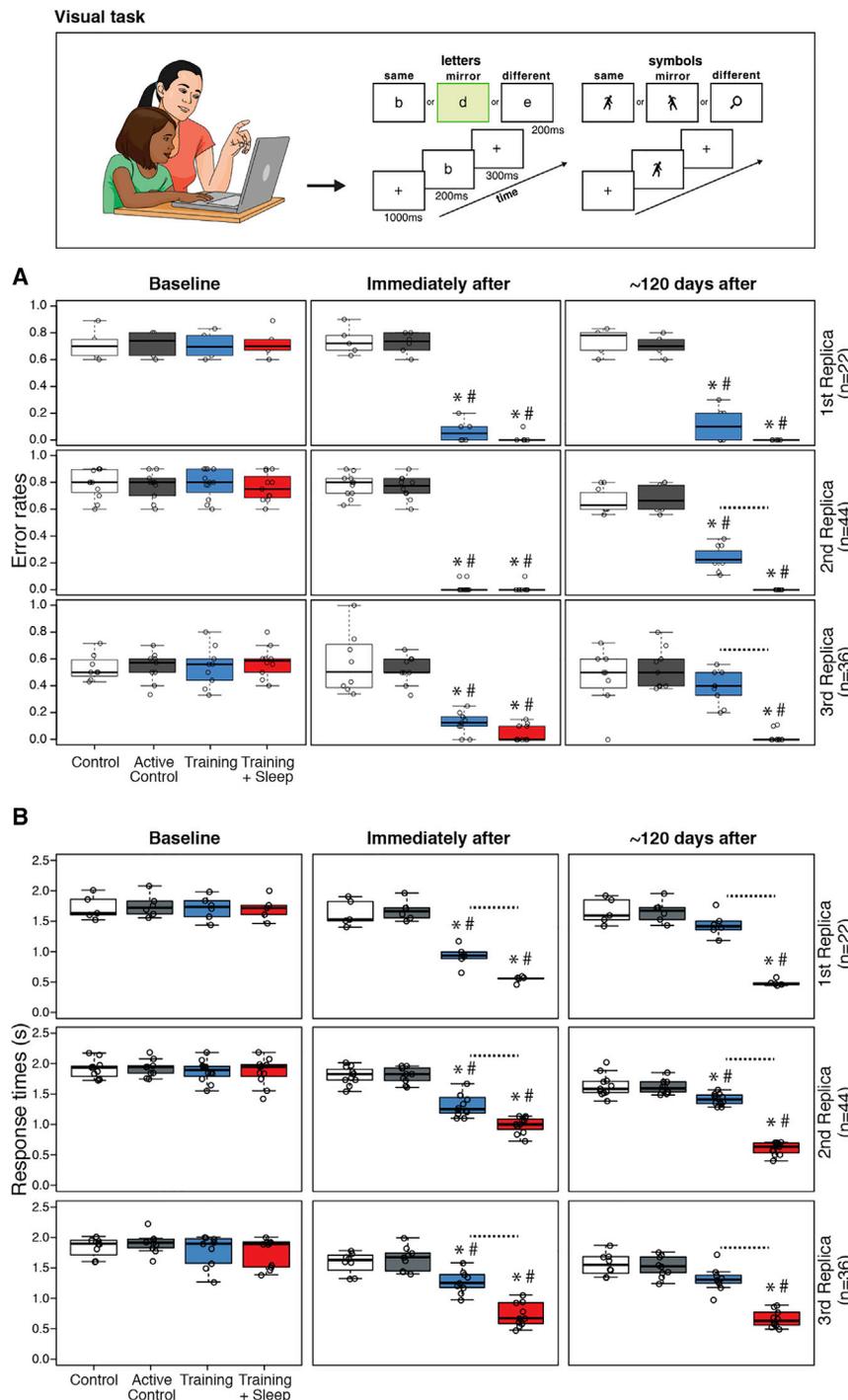
expression of a visual mechanism that is putatively active in humans for ~25 millions of years<sup>55</sup> or more.<sup>24,25</sup>

Concerning *long-term* effects of the intervention, we observed in all three replicas a strong impact of sleep on learning, as the T+S group show perfectly preserved improvements obtained at short term on both the visual (preserved speed and absence of errors) and writing tasks. Note in contrast the approximate 2-fold loss on long-term performance (relative to short-term) in the T group for both visual errors and writing errors.

The impact of the intervention was selective for letters, since no significant change was observed for non-alphabetic control images (Figure 3; Data S1A). Further, to rule out the possibility the observed effects were driven by general letter discrimination training rather than the highly specific influence of learning to actively suppress the intrusion of mirror invariance bias, we conducted signal detection theory (SDT) analyses ( $d'$  and bias), and we also present separate results for “same,” “different,” and “mirror” conditions (see “Extended analysis with data collapsed” in the STAR Methods). All the results pointed to a specific learning effect of mirror discrimination for letters but not for symbols, which took place after the intervention, only for the target groups (T and T+S). Results suggest a bias in mirror trials, perceived as “same” at the baseline, but after the intervention this bias was disrupted for the target training groups (T and T+S) specifically for letters, not for symbols (Figures S5 and S6).

To investigate whether such letter specificity was restricted to mirror letters, next we performed an exploratory “generalization analysis” by sorting the data for mirror letters versus non-mirror letters, as three mirror letters (b, s, z) used in the training of target groups (T and T+S) were present in the set of 10 letters of the visual task (b, f, h, k, s, z, c, e, a, g). As can be seen in Figure 4, there was a training-related improvement also for the other letters (non-mirror letters), which suggests that the specific learning of mirror discrimination for letters generalizes to other letters. This result converges with our previous findings of literacy effects on mirror discrimination that generalizes for visually similar pseudofonts,<sup>41</sup> showing that mirror discrimination learning for letters does not stay encapsulated for the mirror letters used during the training but generalizes to other letters.

Importantly, however, it was only when the targeted training was combined with post-training naps that it boosted *reading fluency*, as participants in the T+S group read approximately two times faster than controls (C and AC) (Figure 6A). Crucially, in all three RCTs, reading speed was proportional to improvements in mirror discrimination speed for letters (Figure 6B). It should be noted that training not followed by sleep (T group) was sufficient to produce clear short-term improvements on mirror discrimination relative to controls, both on the visual (Figure 2) and writing tasks (Figure 5), and even long-term improvements, reducing mirror *errors* for letters in the visual task (Figure 2A). However, the lack of sleep consolidation hampered the magnitude of short and long-term benefits that could be obtained by the training, especially concerning the *speed* of mirror discrimination for letters (Figure 2B). Note also that participants in the T group had regular sleep consolidation at night, but the delay between the morning training and the night sleep is that of several hours. In contrast, participants in the T+S group had sleep consolidation (in the form of post-lunch naps) much closer



**Figure 2. Impact of the Intervention on Mirror Discrimination for Letters**

Top (task): in this visual task, participants decided whether two stimuli sequentially presented on a laptop were the “same” or “different,” pressing corresponding buttons. Two categories of stimuli (letters or iconic symbols) and three trial types (same, different, or mirror) were present. Note that mirror trials should be assigned as “different.” Stimulus duration is indicated below each frame. Bottom (results): results concern the “mirror trials” for letters category (indicated here by a green frame in the top image).

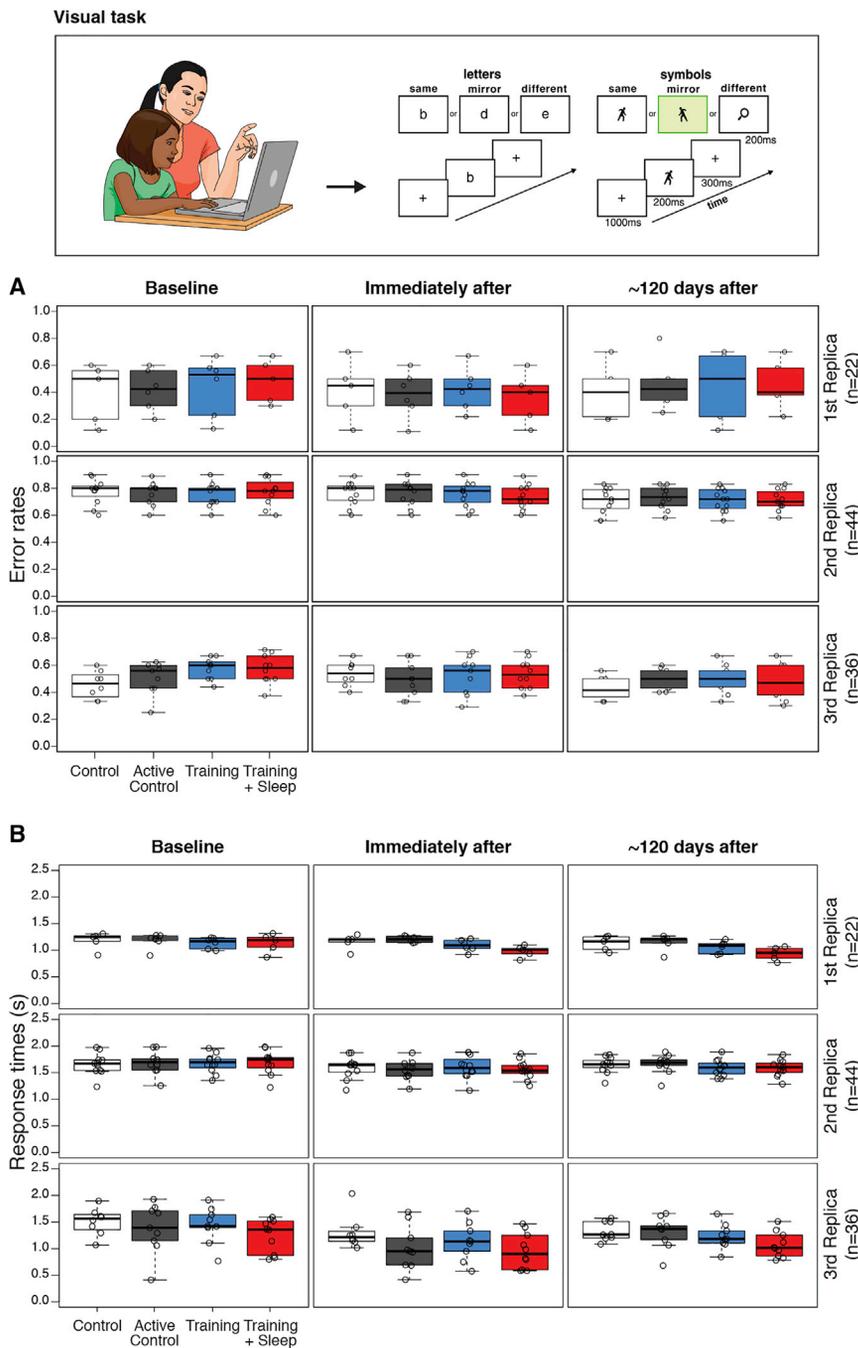
(A) Error rates and (B) response times with only correctly responded trials considered (see Figure S2). Note that these results concern all 10 letters of the experiment (see STAR Methods) and not only mirror letters. Columns: pre-intervention baseline (left), immediately after the 3-week intervention (middle), and ~120 days after the end of the intervention (right). Each dot represents one participant. Symbols indicate statistically significant differences in pairwise group comparisons (Wilcoxon rank-sum test;  $p < 0.05$  Bonferroni-corrected) from the C group (\*), from the AC group (#), and between the T and T+S (–) groups (see Data S1, A and B). The AC group underwent the same activities as the training group but played only with symmetrical letters (e.g., A and X), thus with no mirror discrimination learning. Boxplots: central horizontal line, median; box, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, minimum and maximum; outlier = box limits  $\pm 1.5$  interquartile range. Also see Figures S4–S6, Table S1, Data S2, and Data S3.

for letters and future reading fluency development. To quantify the unique variance of future reading speed development explained by the improvement in mirror discrimination speed, and control for age as a possible confound variable, we pooled all participants ( $n = 106$ ) to generate a linear regression model. Table S2 shows that, irrespective of age, the variance explained is ~43%. An extended description of the data collapsed across the three replicas confirmed the results (STAR Methods).

To rule out the possibility that the within-group correlations could contradict the across-group correlations (Simpson’s paradox), reading fluency data across the three replicas were pooled by group and analyzed. As shown in Figure S1, the within-group correlations did not show Simpson’s paradox, since none of them ( $Rho: -0.037, 0.2, 0.29, 0.33$ ) effectively reversed the direction of the across-group correlations ( $Rho: 0.55, 0.64, 0.70$ ). Please note that none of the within-group correlations were significant, while all the across-group correlations were significant. This was to be expected, since the dynamic range is narrow within groups but

to the morning training. These results support the notion that the delay between the training and the beginning of sleep consolidation is a critical factor, as shown by previous research in children.<sup>56,57</sup> Taken together, the results suggest that sleep consolidation was particularly important for the automaticity of the acquired mirror discrimination, being critical to double reading fluency (Figure 6).

The individual data reported in Figure 6B suggest a possible causal link between the selective inhibition of mirror invariance



**Figure 3. The Intervention Had No Impact on the Control Condition (“Mirror Trials for Symbols”)**

(A) Error rates and (B) response times for the control condition “mirror trials for symbols” (this condition is indicated by a green frame in the top image). In all measurements, replicas, and time points, there is no significant group effect (except for a small intervention effect on RTs in replica 1 immediately after the intervention; see [Data S1A](#)). Columns: measures at the pre-intervention baseline (left), at the end of the 3-week intervention (middle), and at ~120 days after the end of the intervention (right). Each dot represents one participant. Boxplots: central horizontal line, median; box, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, minimum and maximum; outlier = box limits  $\pm 1.5$  interquartile range. Also see [Figures S4–S6](#), [Table S1](#), [Data S2](#), and [Data S3](#).

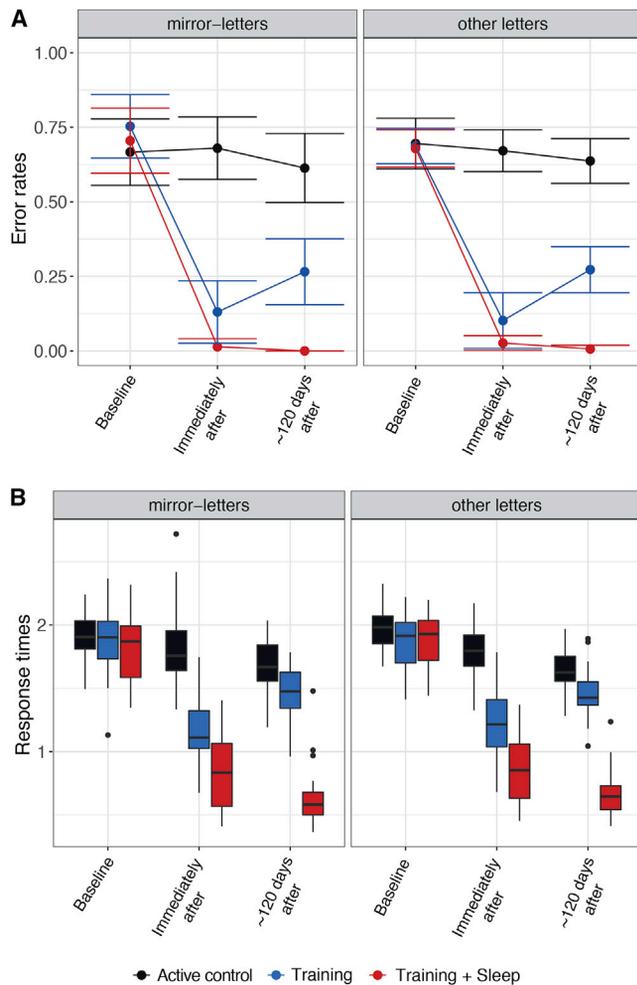
the T group after ~120 days, we should also expect Spearman’s Rho to be larger for T (some ceiling) than for T+S (full ceiling). Indeed, in the T+S group there should be no significant correlation whatsoever between measurements taken before and after training, since all students reached maximum learning. To test these hypotheses, we calculated separately for each group the correlations between performance at baseline and performance after ~120 days for each task. As shown in [Figure S2](#), our hypothesis held. In the control groups, the best-performing students kept their superior performance in relation to the other students, but in the trained groups, this correlation was much weaker, with zero correlation for the T+S group.

It is important to note that the three main measures (errors in the writing task, errors and speed in the visual task) presented equivalent baselines across all groups and replicas. The impact of the sleep-consolidated intervention (T+S) in all three replicas improved all

wide across groups after the intervention, with a causal impact on the performance.

Could it be that within-group differences were preserved or even amplified over time, leading to the large and significant Spearman correlations between performance at baseline and performance after ~120 days for each task? In principle, this could be the case for the control groups but should not occur for the trained groups; i.e., if the effects of the intervention are large enough they should mitigate small differences in the pre-training ranking due to a ceiling effect. If this is correct, given that the T+S group has better overall performance than

the measures relative to baseline. Across groups, participants presented equivalent capacities before the intervention, when reading was not achieved yet, and the differences observed in reading speed several months later are related to the synergic combination of the target training and sleep consolidation. Note that out of 15 baseline measures in total, comprising 5 measures (letters RTs, symbols RTs, letters ERs, symbols ERs, writing) and 3 replicas, only 3 out of 15 comparisons did not meet the normality criterion after Bonferroni correction (letter ERs and writing in the 2nd replica, letter RTs in the 3rd replica; [Data S1C](#);  $\alpha = 0.0033$ ), so there was no floor effect by which



**Figure 4. Generalization of Mirror Discrimination Learning to Other Letters**

(A) Error rates and (B) response times for mirror letters versus other letters in the critical “mirror condition for letters” (collapsed data across the three replicas).

This exploratory analysis suggests a generalization of mirror discrimination learning for other letters. Vertical bars illustrate pointwise 95% confidence intervals.

differences between participants could have been minimized or skewed as they approached some lower limit.

It has been suggested that the kind of training used in the present study could allow strong placebo and Hawthorne effects on student motivation.<sup>58</sup> It is quite unlikely that the results could be explained by potential teacher bias, because the teachers were blind to the experimental design and to group assignment. As to the children, those in the T, T+S, and AC groups indeed went outside the regular classes for 30 min every weekday, for 3 weeks, and this could in fact have generated a Hawthorne effect favoring these groups relative to the C group. Importantly, however, this “special feeling” should also affect the AC group who also went outside the regular classes to play the “nice multi-sensory games” with letters. The exact same training procedures were applied to the T+S, T, and AC groups, the only difference being the letters used in the training for the AC group

(symmetric letters, to avoid mirror discrimination learning). Crucially, however, equivalent results were found between the two control groups (AC and C) across all tasks, measures, and replicas, making the hypothesis of a systematic Hawthorne effect only for the target groups (T and T+S) and not for the AC group very unlikely to explain the overall pattern of results. The same can be concluded for the specific long-term gains in the T+S group. Another caveat worth considering is tester bias, because the tester in the reading test was the same tester who trained the groups. Note, however, that the computer version of the task was not subjected to human bias, and the results were quite similar to those of the pencil-and-paper tests. Furthermore, please note that for the correction of the writing task, where the group knowledge could indeed significantly bias the correction decision (“Was the letter copied correctly?”), we used a blind correction, performed by a different experimenter blind to the grouping of participants.

Despite the limitations of our study, which was performed in a school environment with a high number of dropouts, a small number of participants per group, and restrictions to short and few tests in order to minimize the disturbance on the school dynamics, the three replicas consistently show that the sleep-consolidated training is able to quickly mitigate mirror invariance for letters with very little training (7.5 h in total) and been tremendously amplified by sleep consolidation, generating a long-lasting mirror discrimination learning for letters, unleashing reading fluency in beginner readers.

Overall, the pattern of results is highly suggestive of specific effects of “mirror discrimination learning for letters”: (1) the changes in RTs, errors, *d*, and bias after the intervention are specific for the mirror condition in letters category; (2) the control category (symbols) and groups (C and AC groups) did not present changes in these measures after the intervention, despite equivalent baselines with the target training groups (note that the AC group received the exact same training but instead of mirror letters the children in the AC group were exposed with symmetrical letters, which means that no mirror discrimination learning was possible during their training); (3) the high positive correlation of improvements in mirror discrimination for letters (i.e., specifically in the mirror condition for letters category) and improvements in reading speed (Figure 6B) suggest that it was specifically mirror discrimination improvements for letters that was related to increases in reading fluency; (4) results for same and different conditions also converge to suggest a specific effect of mirror discrimination learning for letters. It is not excluded however that “mirror discrimination learning” can lead to a “general improvement” of letter representation (discrimination/identification); i.e., they are not mutually exclusive.

Could the time spent in post-training naps be proportional to the gains in reading fluency? We added the daily measures of naptime over the 3 weeks of intervention to establish a total amount of sleep per student, akin to a total “dose” of sleep. Interestingly, there were significant correlations across all replicas between total sleep amount and improvement in mirror discrimination speed, as well as between total sleep amount and reading speed (Data S3A). Furthermore, the significant correlations between reading speed and the improvement in mirror discrimination speed (Figure 6B) turned non-significant after adjustment for total sleep



**Figure 5. Impact of the Intervention on Writing Errors**

Top (task): participants observed single letters for 3 s, then with eye-masks wrote the letters with pencil on paper. Bottom (results): error rates for writing letters (including mirror errors). Columns: pre-intervention baseline (left), immediately after the 3-week intervention (middle), and ~120 days after the end of the intervention (right). Each dot represents one participant. Symbols indicate statistically significant differences in pairwise group comparisons (Wilcoxon rank-sum test;  $p < 0.05$  Bonferroni corrected) from the C group (\*), from the AC group (#), and between the T and T+S (–) groups (see [Data S1A](#) and [S1B](#)). Boxplots: central horizontal line, median; box, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, minimum and maximum; outlier = box limits  $\pm 1.5$  interquartile range. In replica 3, the number of letters to copy was extended, including five additional letters (j, q, r, t, y).

amount. These results strongly suggest that total sleep amount mediates the relationship between improvement in mirror discrimination speed and reading speed. We confirmed this mediation effect of sleep with a Sobel test (see “Extended analysis with collapsed” data in the [STAR Methods](#)). To investigate whether sleep per se could improve reading fluency, we applied the reading fluency task in additional control groups who also had post-lunch naps but no training in the morning ([Figure S3](#)). The results show that sleep alone did not impact the reading speed ([Figure 6A](#), see yellow limits).

## DISCUSSION

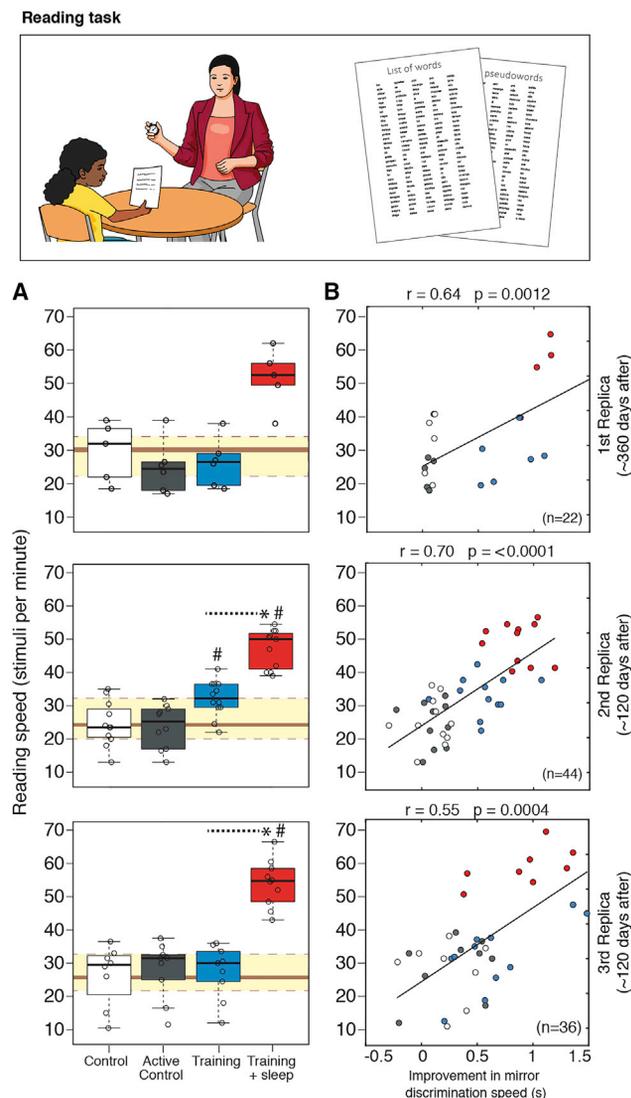
We demonstrated in the present study that mirror invariance is a major physiological barrier for the acquisition of reading fluency in beginner readers. We also showed that the initial bias caused by mirror invariance can be successfully overcome with effective training, specifically designed to improve mirror discrimination for letters. Furthermore, we showed that post-training sleep has a major positive impact on the long-term consolidation of this learning, so as to greatly boost reading fluency.

The accuracy results were particularly meaningful, since they are transparently comparable across conditions and the effects are quite clear. While these effects are very specific for mirror-letter training, with the current data we cannot entirely rule out the possibility that the two training conditions with asymmetric (T, T+S groups) or symmetric letters (AC group) possibly differ in the degree to which they support the discrimination of letters.<sup>6–10</sup> Also note that the same condition has the highest visual similarity between first and second stimulus, the different

condition has the lowest visual similarity, and the mirror condition is in between the two extremes. Thus, it is plausible that only mirror condition had sufficient sensibility to the training effects. A potential lack of sensibility in the same and different conditions could eventually prevent the revelation of a more global improvement in letter processing. Thus, we cannot fully disentangle here between a strict mirror discrimination effect and a more global improvement in visual letter discriminability.

Despite these specific doubts, the results contribute to a perspective *shift* on how humans can learn to read more fluently, targeting mirror invariance in the initial phase of literacy acquisition, by using a multisystem mapping perspective<sup>33</sup> that paradoxically excludes visual inputs during most of the training. However, we cannot rule out the possibility that this much-improved performance is only due to repeated exposure and training with mirror letters, and not with the multisensory component of the training. The importance of bypassing the visual system<sup>59</sup> should be formally tested in future studies, so as to disentangle the contribution of the other systems mappings beyond audio-visual mapping (i.e., grapheme-phoneme correspondences).

The present work also underscores the utility of post-training naps as a powerful and physiological learning enhancer, improving the automaticity, the magnitude, and the duration of the specific pre-sleep learning. Our study represents a concrete example of interdisciplinary *synergy* between the cognitive neuroscience of reading and the science of sleep, enabling an integrated view of how such specific learning *interacts* with sleep at short and long terms. It also shows the interest of using a causal approach (RCTs interventions) combined with ecologically valid designs for the school context, in order to provide evidence that can be rapidly translated into school practice. In fact, the present study can be classified in the Pasteur’s quadrant,<sup>60</sup> a type of research that is conducted both to answer important theoretical questions while having immediate use, in this case



**Figure 6. Impact of the Intervention on Reading Speed**

(A) Reading speed (average between the number of words and pseudo-words read in one minute) measured several months after the end of the intervention (indicated on the right vertical axis in B). The bottom and top boundaries of the yellow stripe in A correspond to the 1<sup>st</sup> and 3<sup>rd</sup> interquartiles (and the brown line to the median) of extra control groups, controlling for the impact of afternoon naps with no training (see Figure S3).

(B) Reading speed as a function of improvements (relative to baseline) in the speed of mirror discrimination for letters in the visual task (see Figure S2). Each dot represents one participant. Symbols indicate statistically significant differences in pairwise group comparisons (Wilcoxon rank-sum test;  $p < 0.05$  Bonferroni-corrected) from the C group (\*), from the AC group (#), and between the T and T+S (–) groups (–) see Data S1A, S1B, and S3 and Table S2. Boxplots: central horizontal line, median; box, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, minimum and maximum; outlier = box limits  $\pm 1.5$  interquartile range. Note that the low number of remaining participants in replica 1 (see Dropouts, Remaining participants, and Data Analyses sections in STAR Methods) precluded significantly different results for T+S group under this stringent correction.

at school (the present work encourages open-science practices such as pre-registration, replication, open data, open analysis, and open access). Further neuroimaging research should test

whether the results found using behavioral measures in our study would also translate to brain changes in the neural systems involved in such learning, as predicted in our model.<sup>33</sup> Further, the contribution of fronto-parietal attention networks during the learning process should be determined.<sup>61,62</sup> In particular, future studies should clarify the potential role played by inhibitory mechanisms<sup>63</sup> linked to the cognitive control network related to the prefrontal cortex (PFC), probing the dynamic aspect of learning, notably related to the switch from the effortful/conscious mirror discrimination to an easy/automatic discrimination by studying the differential contributions of sub-systems of this network: e.g., initially with more involvement of the dorso-lateral PFC (effortful), and later more dependent on the anterior cingulate cortex (automatic).<sup>64</sup>

Follow-up tests of the participants over the next years will determine whether the boost to reading fluency due to the intervention would eventually vanish, as the other students resolve their mirror confusion on their own, or whether the gains will hold on. It could be that the intervention provides only an initial advantage—which nevertheless would probably be helpful for those with more difficulties. However, the evolution trajectories of reading performance could also show parallel lines between the groups over time, with constant superiority of the T+S group. In the best-case scenario, there would be an increasing distance between the groups over the years, with increasing superiority in reading comprehension and pleasure to read for those with a faster departure, with potential gains for academic achievement overall.

Finally, our study adds to the effort to chart just how much improvement could be obtained in all other aspects of academic schooling by the addition of naptime after learning. In our particular case, it would be interesting to further assess (1) whether a sleep-consolidated intervention that just worked on phonics would also boost reading fluency, and (2) whether the two learning types would interact for phonics and mirror discrimination learning as they are supposed to tap on different aspects of reading acquisition, and there is potential for a synergic effect. Major gains are to be expected if sleep-dependent memory consolidation has a cumulative effect over time, as previously hypothesized.<sup>36</sup>

Future studies should determine whether dyslexic populations,<sup>2</sup> who are known to present prolonged mirror confusion with letters,<sup>38</sup> could eventually show improvements in reading outcomes using a similar intervention. Depending on the results, the current clinical view of dyslexia based solely on phonological or audio-visual mapping deficits could be challenged, with potential alternative routes.<sup>61,65,66</sup> It also remains to be determined how well the intervention might apply cross-linguistically to languages with different types of orthography. We predict it to be valid for all languages using the Latin alphabet. For other script systems, especially those without mirror-letters, the impact of such intervention is difficult to predict.

In summary, our study shows the extreme plasticity in human cognition, able to overcome within a very short period of time the “negative” impact of an ancient mechanism deeply rooted in our visual system (mirror invariance), in a selective way, to adapt for processing a new cultural visual object (letters). We reveal a simple, low-tech, and accessible method that can efficiently unleash the reading fluency potential of first graders. Finally, our work

also shows the tremendous potential of using sleep to improve learning at school. Both approaches can benefit millions of children worldwide.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Data collection
  - Groups
  - Extra control groups (sleep with no training)
  - Socioeconomic status (SES)
  - Calendar
  - Ecologically valid RCT design
  - Dropouts
  - Remaining participants
- **METHOD DETAILS**
  - Ethics statement
  - Tasks
  - Visual task
  - Writing task
  - Reading speed task
  - Training
  - Below we describe the activities during a typical session
  - Sleep
  - Pre-registration
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Extended analysis with data collapsed
  - Response times
  - Error rate
  - D primes
  - D prime for mirrored stimuli
  - D prime for different stimuli
  - Betas
  - Beta for mirrored stimuli
  - Beta for different stimuli
  - Indirect effect of mirror discrimination improvement through sleep

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cub.2020.11.031>.

## ACKNOWLEDGMENTS

We thank the Neves School for facilitating the experiments; the parents and children for permission to conduct them; the Latin American School for Education, Cognitive and Neural Sciences (LA School) for providing training and inspiration to the authors; J. Ziegler, S. Dehaene, L. Cohen, J. Morais, R. Kolinsky, J.R. King, J. Sitt, J.F. Araújo, F. Louzada, M. Triunfol, H. Abadzi, and R. Lent for fruitful discussions or manuscript revision; S. Freitas for illustrations, D. Costa Araújo and I. Pereira for library support; and K. Rocha, G. Santana,

and A.E. Oliveira for miscellaneous support. This work was supported by a Google Latin America Research Award (2017) and graduate student fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for A.R.T.; a FWO (Fonds Wetenschappelijk Onderzoek) postdoc fellowship (#12Q4615N) and ERC grant (#742141) for F.P.; resources from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) as grants #308775/2015-5 and #408145/2016-1 for S.R., grants #301744/2018-1 and 425329/2018-6 for M.C., and grant 306659/2019-0 for J.W.; grant #2013/07699-0 from the São Paulo Research Foundation (FAPESP) Center for Neuro-mathematics for S.R. and M.C.; and grant #0642-1.05/18 from Pernambuco Research Foundation (FACEPE) for M.C.

## AUTHOR CONTRIBUTIONS

Conceptualization, F.P., S.R., A.R.T., and N.B.M.; Data Curation, A.R.T. and S.R.; Funding Acquisition, S.R., F.P., A.R.T., N.B.M., and M.C.; Investigation, A.R.T., N.A., S.R., F.P., N.B.M., A.N., and J.W.; Methodology, F.P., S.R., and A.R.T.; Formal Analysis, A.R.T., F.P., S.R., T.Z.L., and N.B.M.; Project Administration, S.R. and F.P.; Resources, S.R., F.P., A.N., and M.C.; Software, F.P.; Supervision, S.R., F.P., N.B.M., and J.W.; Validation, A.R.T., F.P., S.R., M.C., and N.B.M.; Visualization, A.R.T., F.P., S.R., T.Z.L., and N.B.M.; Writing, F.P., S.R., A.R.T., T.Z.L., N.B.M., M.C., and J.W.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 3, 2020

Revised: August 18, 2020

Accepted: November 12, 2020

Published: December 17, 2020; corrected online: February 1, 2021

## REFERENCES

1. UNESCO (2017). *Education & Literacy* (UNESCO Institute for Statistics).
2. Pugh, K. (2006). A neurocognitive overview of reading acquisition and dyslexia across languages. *Dev. Sci.* 9, 448–450, discussion 451–453.
3. Peterson, R.L., and Pennington, B.F. (2015). Developmental dyslexia. *Annu. Rev. Clin. Psychol.* 11, 283–307.
4. Dehaene, S., Pegado, F., Braga, L.W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., and Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science* 330, 1359–1364.
5. National Reading Panel (2000). *Report of the National Reading Panel: teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: reports of the subgroups* (National Institute of Child Health and Human Development, National Institutes of Health).
6. Gibson, E.J., Gibson, J.J., Pick, A.D., and Osser, H. (1962). A developmental study of the discrimination of letter-like forms. *J. Comp. Physiol. Psychol.* 55, 897–906.
7. Gibson, E.J., Pick, A., Osser, H., and Hammond, M. (1962). The role of grapheme-phoneme correspondence in the perception of words. *Am. J. Psychol.* 75, 554–570.
8. Pick, A.D. (1965). Improvement of visual and tactual form discrimination. *J. Exp. Psychol.* 69, 331–339.
9. Samuels, S.J. (1973). Effect of distinctive feature training on paired-associate learning. *J. Educ. Psychol.* 64, 164–170.
10. Nelson, R.O., and Wein, K.S. (1974). Training letter discrimination by presentation of high-confusion versus low-confusion alternatives. *J. Educ. Psychol.* 66, 926–931.
11. Lyytinen, H., Erskine, J., Kujala, J., Ojanen, E., and Richardson, U. (2009). In search of a science-based application: a learning tool for reading acquisition. *Scand. J. Psychol.* 50, 668–675.

12. Saine, N.L., Lerkanen, M.K., Ahonen, T., Tolvanen, A., and Lyytinen, H. (2011). Computer-assisted remedial reading intervention for school beginners at risk for reading disability. *Child Dev.* **82**, 1013–1028.
13. Carreiras, M., Seghier, M.L., Baquero, S., Estévez, A., Lozano, A., Devlin, J.T., and Price, C.J. (2009). An anatomical signature for literacy. *Nature* **461**, 983–986.
14. Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S.F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., et al. (2000). A cultural effect on brain function. *Nat. Neurosci.* **3**, 91–96.
15. Pegado, F., Comerlato, E., Ventura, F., Jobert, A., Nakamura, K., Buiatti, M., Ventura, P., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., et al. (2014). Timing the impact of literacy on visual processing. *Proc. Natl. Acad. Sci. USA* **111**, E5233–E5242.
16. Boets, B., Vandermosten, M., Cornelissen, P., Wouters, J., and Ghesquière, P. (2011). Coherent motion sensitivity and reading development in the transition from prereading to reading stage. *Child Dev.* **82**, 854–869.
17. Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K., and Facoetti, A. (2012). A causal link between visual spatial attention and reading acquisition. *Curr. Biol.* **22**, 814–819.
18. Carroll, J.M., Solity, J., and Shapiro, L.R. (2016). Predicting dyslexia using prereading skills: the role of sensorimotor and cognitive abilities. *J. Child Psychol. Psychiatry* **57**, 750–758.
19. Bertoni, S., Franceschini, S., Ronconi, L., Gori, S., and Facoetti, A. (2019). Is excessive visual crowding causally linked to developmental dyslexia? *Neuropsychologia* **130**, 107–117.
20. Bruer, J.T. (1997). Education and the brain: a bridge too far. *Educ. Res.* **26**, 4–16.
21. Sigman, M., Peña, M., Goldin, A.P., and Ribeiro, S. (2014). Neuroscience and education: prime time to build the bridge. *Nat. Neurosci.* **17**, 497–502.
22. Bornstein, M.H., Gross, C.G., and Wolf, J.Z. (1978). Perceptual similarity of mirror images in infancy. *Cognition* **6**, 89–116.
23. Rollenhagen, J.E., and Olson, C.R. (2000). Mirror-image confusion in single neurons of the macaque inferotemporal cortex. *Science* **287**, 1506–1508.
24. Hollard, V.D., and Delius, J.D. (1982). Rotational invariance in visual pattern recognition by pigeons and humans. *Science* **218**, 804–806.
25. Sutherland, N.S. (1957). Visual discrimination of orientation and shape by the octopus. *Nature* **179**, 11–13.
26. Cornell, J.M. (1985). Spontaneous mirror-writing in children. *Can. J. Psychol.* **39**, 174–179.
27. Dehaene, S., Nakamura, K., Jobert, A., Kuroki, C., Ogawa, S., and Cohen, L. (2010). Why do children make mirror errors in reading? Neural correlates of mirror invariance in the visual word form area. *Neuroimage* **49**, 1837–1848.
28. Pegado, F., Nakamura, K., Cohen, L., and Dehaene, S. (2011). Breaking the symmetry: mirror discrimination for single letters but not for pictures in the visual word form area. *Neuroimage* **55**, 742–749.
29. McIntosh, R.D., Hillary, K., Brennan, A., and Lechowicz, M. (2018). Developmental mirror-writing is paralleled by orientation recognition errors. *Laterality* **23**, 664–683.
30. Eger, E., Henson, R.N., Driver, J., and Dolan, R.J. (2004). BOLD repetition decreases in object-responsive ventral visual areas depend on spatial attention. *J. Neurophysiol.* **92**, 1241–1247.
31. Vuilleumier, P., Schwartz, S., Duhoux, S., Dolan, R.J., and Driver, J. (2005). Selective attention modulates neural substrates of repetition priming and “implicit” visual memory: suppressions and enhancements revealed by fMRI. *J. Cogn. Neurosci.* **17**, 1245–1260.
32. Dilks, D.D., Julian, J.B., Kubilius, J., Spelke, E.S., and Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *J. Neurosci.* **31**, 11305–11312.
33. Pegado, F., Nakamura, K., and Hannagan, T. (2014). How does literacy break mirror invariance in the visual system? *Front. Psychol.* **5**, 703.
34. Stickgold, R., James, L., and Hobson, J.A. (2000). Visual discrimination learning requires sleep after training. *Nat. Neurosci.* **3**, 1237–1238.
35. Marshall, L., Helgadóttir, H., Mölle, M., and Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature* **444**, 610–613.
36. Cabral, T., Mota, N.B., Fraga, L., Copelli, M., McDaniel, M.A., and Ribeiro, S. (2018). Post-class naps boost declarative learning in a naturalistic school setting. *NPJ Sci. Learn.* **3**, 14.
37. Cousins, J.N., van Rijn, E., Ong, J.L., Wong, K.F., and Chee, M.W.L. (2019). Does splitting sleep improve long-term memory in chronically sleep deprived adolescents? *npj Sci. Learn.* **4**, 8.
38. Corballis, M.C. (2018). Mirror-image equivalence and interhemispheric mirror-image reversal. *Front. Hum. Neurosci.* **12**, 140.
39. Kolinsky, R., Morais, J., Content, A., and Cary, L. (1987). Finding parts within figures: a developmental study. *Perception* **16**, 399–407.
40. Danziger, E., and Pederson, E. (1998). Through the looking glass: literacy, writing systems and mirror-image discrimination. *Writ. Lang. Lit.* **1**, 153–169.
41. Pegado, F., Nakamura, K., Braga, L.W., Ventura, P., Nunes Filho, G., Pallier, C., Jobert, A., Morais, J., Cohen, L., Kolinsky, R., and Dehaene, S. (2014). Literacy breaks mirror invariance for visual stimuli: a behavioral study with adult illiterates. *J. Exp. Psychol. Gen.* **143**, 887–894.
42. Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.A., and Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* **123**, 291–307.
43. Bouhali, F., Thiebaut de Schotten, M., Pinel, P., Poupon, C., Mangin, J.-F., Dehaene, S., and Cohen, L. (2014). Anatomical connections of the visual word form area. *J. Neurosci.* **34**, 15402–15414.
44. White, A.L., Palmer, J., Boynton, G.M., and Yeatman, J.D. (2019). Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. *Proc. Natl. Acad. Sci. USA* **116**, 10087–10096.
45. Pegado, F., and Grainger, J. (2020). A transposed-word effect in same-different judgments to sequences of words. *J. Exp. Psychol. Learn. Mem. Cogn.* **46**, 1364–1371.
46. Pegado, F., and Grainger, J. (2019). Dissociating lexical and sublexical contributions to transposed-word effects. *Acta Psychol. (Amst.)* **201**, 102943.
47. Snell, J., and Grainger, J. (2019). Consciousness is not key in the serial-versus-parallel debate. *Trends Cogn. Sci.* **23**, 814–815.
48. Yoncheva, Y.N., Wise, J., and McCandliss, B. (2015). Hemispheric specialization for visual words is shaped by attention to sublexical units during initial learning. *Brain Lang.* **145–146**, 23–33.
49. Nakamura, K., Kuo, W.J., Pegado, F., Cohen, L., Tzeng, O.J., and Dehaene, S. (2012). Universal brain systems for recognizing word shapes and handwriting gestures during reading. *Proc. Natl. Acad. Sci. USA* **109**, 20762–20767.
50. Longcamp, M., Boucard, C., Gilhodes, J.C., Anton, J.L., Roth, M., Nazarian, B., and Velay, J.L. (2008). Learning through hand- or typewriting influences visual recognition of new graphic shapes: behavioral and functional imaging evidence. *J. Cogn. Neurosci.* **20**, 802–815.
51. Perea, M., Moret-Tatay, C., and Panadero, V. (2011). Suppression of mirror generalization for reversible letters: evidence from masked priming. *J. Mem. Lang.* **65**, 237–246.
52. Duñabeitia, J.A., Molinaro, N., and Carreiras, M. (2011). Through the looking-glass: mirror reading. *Neuroimage* **54**, 3004–3009.
53. Lemos, N., Weissheimer, J., and Ribeiro, S. (2014). Naps in school can enhance the duration of declarative memories learned by adolescents. *Front. Syst. Neurosci.* **8**, 103.
54. Torres, A.R., Mota, N.B., Adami, N., Weissheimer, J., Naschold, A., Copelli, M., Pegado, F., and Ribeiro, S. (2018). Effects of multi-sensory training and post-training naps on breaking mirror invariance for letters during literacy acquisition. OSF. Published online September 2, 2018. <https://doi.org/10.17605/OSF.IO/TZNWE>.

55. Stevens, N.J., Seiffert, E.R., O'Connor, P.M., Roberts, E.M., Schmitz, M.D., Krause, C., Gorscak, E., Ngasala, S., Hieronymus, T.L., and Temu, J. (2013). Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes. *Nature* *497*, 611–614.
56. Diekelmann, S., Wilhelm, I., and Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Med. Rev.* *13*, 309–321.
57. Backhaus, J., Hoeckesfeld, R., Born, J., Hohagen, F., and Junghanns, K. (2008). Immediate as well as delayed post learning sleep but not wakefulness enhances declarative memory consolidation in children. *Neurobiol. Learn. Mem.* *89*, 76–80.
58. Adair, J.G., Sharpe, D., and Huynh, C.L. (1989). Hawthorne control procedures in educational experiments: a reconsideration of their use and effectiveness. *Rev. Educ. Res.* *59*, 215–228.
59. Rock, I., and Victor, J. (1964). Vision and touch: an experimentally created conflict between the two senses. *Science* *143*, 594–596.
60. Stokes, D.E. (1997). *Pasteur's Quadrant – Basic Science and Technological Innovation* (Brookings Institution Press).
61. Franceschini, S., Gori, S., Ruffino, M., Viola, S., Molteni, M., and Facoetti, A. (2013). Action video games make dyslexic children read better. *Curr. Biol.* *23*, 462–466.
62. Bavelier, D., and Green, C.S. (2019). Enhancing attentional control: lessons from action video games. *Neuron* *104*, 147–163.
63. Borst, G., Ahr, E., Roell, M., and Houdé, O. (2015). The cost of blocking the mirror generalization process in reading: evidence for the role of inhibitory control in discriminating letters with lateral mirror-image counterparts. *Psychon. Bull. Rev.* *22*, 228–234.
64. Jiang, J., Zhang, Q., and van Gaal, S. (2015). Conflict awareness dissociates theta-band neural dynamics of the medial frontal and lateral frontal cortex during trial-by-trial cognitive control. *Neuroimage* *116*, 102–111.
65. Witton, C., Talcott, J.B., Hansen, P.C., Richardson, A.J., Griffiths, T.D., Rees, A., Stein, J.F., and Green, G.G. (1998). Sensitivity to dynamic auditory and visual stimuli predicts nonword reading ability in both dyslexic and normal readers. *Curr. Biol.* *8*, 791–797.
66. Harrar, V., Tammam, J., Pérez-Bellido, A., Pitt, A., Stein, J., and Spence, C. (2014). Multisensory integration and attention in developmental dyslexia. *Curr. Biol.* *24*, 531–535.
67. Seabra, A.G., Dias, N.M., Mecca, T., and Macedo, E.C. (2017). Contribution of word reading speed to reading comprehension in Brazilian children: does speed matter to the comprehension model? *Front. Psychol.* *8*, 630.
68. Quinn, D.M., Cooc, N., McIntyre, J., and Gomez, C.J. (2016). Seasonal dynamics of academic achievement inequality by socioeconomic status and race/ethnicity: updating and extending past research with new national data. *Educ. Res.* *45*, 443–453.
69. Kolinsky, R., and Verhaeghe, A. (2017). Lace your mind: the impact of an extra-curricular activity on enantiomorphy. *Journal of Cultural Cognitive Science* *1*, 57–64.
70. Fernandes, T., and Kolinsky, R. (2013). From hand to eye: the role of literacy, familiarity, graspability, and vision-for-action on enantiomorphy. *Acta Psychol. (Amst.)* *142*, 51–61.
71. Ballarini, F., Martínez, M.C., Díaz Perez, M., Moncada, D., and Viola, H. (2013). Memory in elementary school children is improved by an unrelated novel experience. *PLoS ONE* *8*, e66875.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw and analyzed data plus code have been deposited in OSF	This paper	<a href="https://osf.io/643jh/">https://osf.io/643jh/</a>
Experimental Models: Organisms/Strains		
Humans	Children between 5 and 7 years old were assessed in the Colegio Nossa Senhora das Neves school in the city of Natal, Brazil.	<a href="https://osf.io/643jh/">https://osf.io/643jh/</a>
Software and Algorithms		
The R Project for Statistical Computing	N/A	<a href="https://www.r-project.org">https://www.r-project.org</a> software
PsychoPy	N/A	<a href="https://www.psychopy.org">https://www.psychopy.org</a> software
Photoshop and Illustrator	Adobe, San Jose, CA	N/A

### RESOURCE AVAILABILITY

#### Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by Sidarta Ribeiro ([sidartaribeiro@neuro.ufrn.br](mailto:sidartaribeiro@neuro.ufrn.br)).

#### Materials Availability

This study did not generate new unique reagents.

#### Data and Code Availability

Original data and code have been deposited in OSF and can be found on <https://osf.io/643jh/>.

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

#### Data collection

Data collection took place in the Colegio Nossa Senhora das Neves school in the city of Natal, Brazil (Av. Coronel Estevam, 21 – Alecrim), starting on 21 September 2016, and ending on 14 June 2019. Children between 5 and 7 years old (1<sup>st</sup> Replica (n = 32; 15 females), mean age = 6.0 ± 0.56 years; 2<sup>nd</sup> Replica (n = 60; 33 females), mean age = 5.95 ± 0.62 years; 3<sup>rd</sup> Replica (n = 48; 23 females), mean age = 6.02 ± 0.64 years) were initially assessed at the beginning of literacy acquisition, a few months after the start of the academic year. Legal guardians were invited to meet in the school to receive information about the project and to give written consent for the participation of their children in the research. Participants were not paid to take part in this study, since this is legally precluded in Brazil. Individuals with a previous history of neurologic or psychiatric symptoms were excluded. Each participant was informed that his/her parent(s) allowed his/her participation in the research in the coming weeks, and that they (children) could choose whether or not to participate. Children who agreed to participate were subjected to a previous rudimentary phonic lesson, so that they all received instruction on the spelling of letters. During the phonic lesson one of the researchers (ART) presented all the letters of the alphabet in their uppercase and lowercase forms. Next, participants were subjected to two baseline tests: the visual task (same/different matching) presented on a laptop and a letter-writing task using paper and pencil (details below). Teachers were blind to the assignment of groups. The researchers responsible for the intervention (ART, NA) were trained by way of verbal instructions based on the procedures described in the manuscript.

#### Groups

The average scores of these baseline tests were used to rank and randomly assign participants into four balanced groups for initial performance, with an equal number of children in each group. After this stratification the groups were differentially treated as following: a multisensory-motor training group using mirror-letter pairs such as b and d (T = Training); a group receiving the same training with mirror-letters been followed by post-training naps (T+S = Training + Sleep); a classical no-training control group (C = Control), and an additional active control group (AC = Active Control) that received the same multisensory-motor training as the previous one but playing only with symmetrical letters, thus with no mirror discrimination training.

The AC group was used as a control for several socio-emotional aspects introduced by the special situation of the intervention for participants in the training groups (T and T+S). Note that during these three weeks, children in the training groups and the AC group did not participate in the regular activities of the school during the 30-minute duration of the training but instead played new multi-sensory-motor games in restricted groups (including colleagues from different classes). Thus, while controlling for unspecific factors, the training in this group did not involve mirror discrimination learning.

The C group only took part in the tests but did not participate in any activity during the intervention period. The other groups (T, T+S, and AC) were invited to participate daily (weekdays) according to their respective training/nap sessions.

Baseline performance across replicas showed a large degree of inter-individual variability, which reflects the variability in the pool of participants from one academic year to the next. Importantly, however, within any given year (i.e., replica), the performance at baseline was guaranteed to be equivalent across groups, due to the stratified randomization. As a consequence, the medians varied very little across replicas.

### Extra control groups (sleep with no training)

The extra control groups were obtained from a convenience sample of non-trained children undergoing four prevalent regimes of class/sleep combinations in their daily routines. Typically, Brazilian children spend either all day or only the morning at school. Based on a questionnaire filled by the parents and regime enrolled in school, the children were sorted into four naturalistic groups: the “no nap at home” group with children that stayed in school only in the morning and usually had no afternoon naps at home, the “nap at home” group with children that stayed in school only in the morning and usually had afternoon naps at home; the “no nap at school” group with children that stayed in school the whole day but did not have afternoon naps; and the “nap at school” group with children that stayed in school the whole day and had afternoon naps. The only difference between this last group and the T+S group is that the “nap at school” participants did not undergo any training before taking the nap. The results show that sleep alone does not impact reading speed at any level (Figure S3) and that all four extra groups display the same reading fluency as the main control groups. Note in Figure 6A that the yellow limits corresponds to the 1<sup>st</sup> and 3<sup>rd</sup> inter-quartiles of these extra control groups averaged together; and the brown line represents the median). The extra controls were not laboratory-oriented but rather school-oriented, and were used to obtain extra control on the expected reading performance from that population, and to test the impact of naps without training. Randomization of the extra control groups was performed with a 5-digit random number table.

### Socioeconomic status (SES)

Socio-Economic Status was assessed via the declared salary range of the legal guardians. The minimum monthly wage (salary) in Brazil in 2019 is 998 Brazilian Reais (~262 USD), and is paid 13 times per year. We found the following SES distributions for our samples: 1<sup>st</sup> Replica: 6.25% up to 1 salary, 12.5% 1–2 salaries, 34.4% 2–3 salaries, 46.9% above 4 salaries; 2<sup>nd</sup> Replica: 6.68% up to 1 salary, 8.33% 1–2 salaries, 21.7% 2–3 salaries, 63.3% above 4 salaries; 3<sup>rd</sup> Replica: 4.25% up to 1 salary, 12.8% 1–2 salaries, 17.0% 2–3 salaries, 65.9% above 4 salaries.

### Calendar

The academic year in Brazil typically starts in February. The baseline measurements took place in September. The “immediately after” measurements took place in October (maximum 1 week after the end of the intervention) and the long-term measurements (~120 days after the end of the intervention) in February of the next academic year. The reading scores of replica 1 were measured ~360 days after the end of the intervention, i.e., in September of the following academic year, and for replicas 2 and 3 they were measured ~120 days after the end of the intervention.

### Ecologically valid RCT design

In order to have direct relevance for education, well-controlled RCTs were designed in a school-based manner. First, the training, the sleep, and the testing were performed inside the school. Second, not only “laboratory-style” tasks were used but also those normally practiced at school (writing and reading on paper). Third, stimuli in all three tasks (visual, writing, and reading) respected the natural prevalence of letters in Brazilian Portuguese, instead of over-representing mirror-letters for instance. The naturalistic design is thus a plus, not a caveat. Our intervention dealt with a translation to real school education, beyond the idealized laboratory situation, with direct relevance to educational practices, i.e., with societal impact.

Our design choices of course imposed some limitations. In order to reduce the impact of the research on the school dynamics, we had to restrict the number of participants, tests, and trials. Despite these limitations, our measures presented sufficient stability, including equivalent results between the two control groups (C and AC) across data points, tasks, and replicas, and sufficient sensitivity to learning changes in the trained groups (T and T+S), in order to satisfactorily answer our research questions.

The RCTs we conducted adhered to the most important aspects of the CONSORT statement (established for clinical trials) as a randomized (for groups) and controlled (two control groups) study, with description of participants, study settings, outcomes, background, objectives, pre-registration, changes in trial design (relative to pre-registration, as described here), changes in the outcomes (dropouts were described for each task), sample size (the pre-registration describes how it was determined), similarity of interventions (clearly stated, e.g., see description of group AC), statistical methods, participant flow (N.A.), losses and exclusions (N.A.), reasons for stopping (the dropouts are described below), recruitment, baseline data (except for reading speed because participants were not reading at baseline), blinding (the experimenter was not blind to the group when applied the tasks but for the “writing

task,” which carries a potential of human bias we used a blind correction; the computer task was of course blind), the limitations of the study and the funding information. We did not adhere to minor points in the checklist of CONSORT concerning the format of the report (identification as a randomized trial in the title and the use of structured summary, i.e., design, methods, results, conclusions).

### Dropouts

Since the studies were performed not in a laboratory but in a school, and especially because long-term retesting was required, a high rate of dropouts was expected. Indeed, several participants left the school in the following academic years, hindering long-term retesting (three participants in the 1<sup>st</sup> replica, six in the 2<sup>nd</sup> replica with two additional participants who declined to participate after the intervention had begun, and 12 in the 3<sup>rd</sup> replica). Further reduction in the number of participants for the visual task was produced by the exclusion criteria for quality control (see Data Analysis section)

### Remaining participants

Note that the small number of remaining participants in replica 1 (5 or 6 per group) precluded significant long-term group differences between T versus T+S in both the visual (for Error Rates), writing and reading tasks, when using the stringent Bonferroni-correction. In contrast, replicas 2 and 3 with more participants did not presented this issue. Further, replica 2 (the most powered one), show significant long-term effects on the T group (relative to the other control groups) in all three tasks (visual, writing and reading). Thus, these results points to a *threshold* of minimum number of participants required in future replications when using stringent corrections as here: it corresponds to the number of participants in replica 2, i.e., 11 participants per group (as observed in the visual task). Note also that this is the number of *remaining* participants, so a higher number of *initial* participants is necessary to take into account the high prevalence of dropouts in a school-based study with several data points.

## METHOD DETAILS

### Ethics statement

This research was approved by the UFRN local ethical committee (5537 – Universidade Federal do Rio Grande do Norte – Lagoa Nova Campus Central, Brazil), permit # 85580318.7.0000.5537.

### Tasks

Three tasks were used here. A “laboratory-style” *visual* discrimination task, and two “school-style” tasks: a paper and pencil *writing* task and *reading* on paper task. Note that for replica 3 we had planned to record an intermediate post-intervention time-point (60 days after the intervention) but for practical reasons we decided to cancel this extra recording.

### Visual task

In our standard test for mirror discrimination (Figure 2), pairs of images were presented on a computer screen (laptop), and participants were instructed to decide whether they were the “same” or “different” and press the corresponding buttons. In separate trials, single letters or visual icons were presented in a sequence similar to that used previously.<sup>15,28,41</sup> Each trial started with a fixation cross for 1000 ms, followed by the first stimulus for 200 ms, then a new fixation cross for 300 ms, then the second stimulus (the same, a mirror version, or a different one), which was 25% larger than the first stimulus, to avoid physical repetition. Two categories (letters versus symbols) and three types of repetition (same, mirror, different) were used intermixed in a single block, for a total of 60 trials per recording. All trials were automatically sorted, so as to generate a random sequence of stimulus presentation. We presented a slightly reduced subset of stimuli previously used in Pegado et al.<sup>28</sup> with asymmetric single letters (lower case): b, f, h, k, s, z, c, e, a, g and asymmetric visual icons:



### Writing task

In this “letter writing” task the participants had to copy the following letters (b, c, a, f, e, d, g, h, k, s, z, p) for replica 1 and 2, and an extended version for replica 3 with five additional letters (j, q, r, t, y). All letters were presented in Arial font, size 90. Each child received a blank sheet divided into squares (one for each letter copy). The researcher showed one letter at a time for 3 s, then immediately after that the participant had to write the letter while blindfolded.

### Reading speed task

Here, children were requested to read a list of words and another list of pseudo-words printed on separate sheets of paper. Participants had 1 minute to read each list. These lists were essentially the same used in our previous studies done with adults, but stimuli outside the typical vocabulary of young children were excluded. The reading speed scores (total number of stimuli read) were

calculated by averaging the number of words and pseudo-words read in one minute, as was done in our previous studies.<sup>4,15,41</sup> Results for word and pseudo-word reading fluency, respectively for C, AC, T, and T+S, were the following: *1<sup>st</sup> replica* (words: 32.0; 29.3; 31.5; 58.0 and pseudo-words: 27.2; 20.5; 21.2; 45.2); *2<sup>nd</sup> replica* (words: 26.5; 25.1; 35; 50.7 and pseudo-words: 22.5; 22.5; 29.3; 42.8); *3<sup>rd</sup> replica* (words: 28.9; 31.8; 32.2; 57.1 and pseudo-words: 24.0; 24.0; 22.8; 48.9).

The average fluency for word reading in all of our control groups (~25/30 words per minute) was in the expected range, similar to those obtained from equivalent populations in Brazil<sup>67</sup> despite differences in the testing material. Interestingly however, our scores can be readily compared with those obtained in our previous study using essentially the same material.<sup>4</sup> Brazilian adults who learned to read in adulthood (referred to as ex-illiterates from Brazil, i.e., “EXB” group) obtained average scores (~30 words and ~23 pseudo-words per minute) that were comparable to those of our present control participants. Note that EXB participants had on average 1.4 years of literacy training. However, another group of ex-illiterates (from Portugal-“EXP” group) only reached a very modest level of reading fluency (~15 words and ~11 pseudo-words per minute) despite 2.4 years of literacy training (see Table S1 and Figure S1 in Dehaene et al.<sup>4</sup>).

Further, note the equivalent levels of reading speed in replica 1 (measured ~360 days after the end of the training) and the other two replicas (measured ~120 days after) in Figure 4A. One could expect better reading performance in children with more literacy training (replica 1). However, as mentioned in the previous paragraph, this relationship is not always reliable. Additionally, in replica 1, children had an additional holiday period (~1 month), and the tests were performed immediately after this period of literacy training interruption, when the performance is expected to drop, a classical effect described more than a century ago and known as “summer learning loss.” This effect is particularly strong in less developed areas<sup>68</sup> such as in Natal, given that children are not highly stimulated to read during holidays, and watch TV instead.

### Training

Participants in the T, T+S, and AC groups received daily (weekdays) training sessions of 30 minutes at 09:00, 09:35, and 10:10 for 3 weeks. The order of the three groups for each time slot was counterbalanced across days using a Latin square procedure.

A typical training session comprised a sequence of short (a few minutes) multisensory-motor activities presented as games. They were performed in reduced groups (between 8 to 15 children; always with an equivalent initial number of participants across the T, T+S and AC groups *per replica*) by a single researcher (ART). Only one pair of letters was used in each session. T and T+S groups played with opposed mirror-letters while the AC group played in contrast with symmetrical letters.

The training was built based on our *multisystem mapping* hypothetical mechanism for mirror discrimination learning for letters.<sup>33</sup> We have postulated that this mechanism should be particularly important to discriminate mirror-letters. Thus, we created activities aiming to maximize mappings between systems representing letters and used pairs of mirror-letters in the training groups in order to improve the discrimination between them. Beyond the three systems traditionally involved in letter representations at school, i.e., auditory, visual, and writing gestures systems, we also included active mappings with the articulatory system (i.e., phonation of letter name) and, more unconventionally, we also included mappings with the somatosensory system. To reduce the influence of vision<sup>60</sup> and to increase attention to other sensory and motor representations (phonological, writing, articulatory, and tactile representations), around 70% of the time of the session children had their eyes closed. The training sessions typically started by blocking visual inputs with eye-masks. However, at the end of the session, eyes were opened and the same activities were performed again (for a shorter duration), aiming this time, to map the auditory, writing, articulatory, and tactile representations with visual representations. The final aim is to develop a robust multisystem letter representation. Three types of activities (described below) were proposed with eyes closed: 1) “**air-writing**”; 2) **writing on a paper**; and 3) “**perceiving letters on hands**”.

By the end of the session, one *additional* activity was introduced (with eyes opened): “**perception for action**,” inspired by the fact that illiterates that work as ‘lace makers’ develop mirror discrimination abilities despite the lack of literacy.<sup>69</sup>

### Below we describe the activities during a typical session

At the beginning of a session the participant’s hand were held by the experimenter so that the latter could guide the participant to write, either on a sheet of paper using a pencil, or in the air using larger writing gestures. After familiarization (producing basic shapes such as circles, semicircles, vertical and horizontal lines), the researcher began to guide the participant’s hand for the maneuver of each letter. Then, the same procedure was performed for the other letter (only two letters were used on each session). These activities were practiced both on paper and as air writing. Then, participants were asked to *actively* draw the same letters (Figure 1D, left). Variations in the size of letters were proposed, aiming to develop a more abstract *size invariant* letter representation, as well as making the training more interesting for the children. To improve an association between the sound of the letter and the maneuver to write it, the researcher repeated the sound of the letters during the exercises. The researcher also asked the participant to say aloud the sound of the letters, aiming to also create associations with articulatory representations.<sup>33</sup>

Another type of activity (“perceiving letters on hands”) consisted in tracing a letter in the hand of the participant and asked them to say out loud what the letter was (Figure 1D, right). This activity aimed to map tactile representations with articulatory and auditory representations of letters.

By the end of the session the same types of activities were shortly performed again but this time with eyes open. In addition, we also introduced at this point activities requiring “perception for action” where children were requested to indicate parts of letters. For the training groups (T and T+S) participants were requested to indicate the lateral round part of a ‘b’ shape for instance with the corresponding hand from the same side (i.e., with the right hand in this case) while for the AC group participants were requested to indicate

parts of the symmetrical letters (e.g., central point of the letter 'A') with either hand. The rationale to include this activity was that even illiterates (who typically make mirror errors) can distinguish left and right orientations when action upon the image is needed.<sup>70</sup> We aimed to take advantage of the discriminative capacity of the “perception for action” system for left versus right sides, to help the “perception for recognition” system to break its mirror invariance for letters.

For all these groups having multisensory-motor ‘games’ (T, T+S and AC), the experimenter was keen to keep the participants fully engaged, by making the games fun, of short duration, with a dynamic transition between them, and presenting a difficulty level suited to their developmental stage.

## Sleep

Participants took post-training nap sessions of up to 2 h a day (between 12:30–14:30) during the 3 weeks of the intervention. Participants were invited to nap after lunch in a quiet room of the school on mats and wearing eye masks (Figure 1E). While it would have been interesting to measure physiological correlates of sleep, our naturalistic experimental design did not include devices to gauge activity. Instead, the experimenter recorded the amount of time that each participant remained still during the nap opportunity, as a proxy of the amount of post-training sleep per day of intervention. As most schools in Brazil, the one where the experiment was carried out serves a community where actigraphs or portable EEG devices would most certainly be gossiped about, tampered with, bartered or otherwise cracked for fun and curiosity by the technology-craving students, as well as their relatives and friends. Given the evidence that school learning is improved by novelty,<sup>71</sup> we chose not to use external devices and thus circumvented a major source of confound to preserve the school environment as intact as possible.

## Pre-registration

Replica 3 was pre-registered at OSF Registries on 09/02/2018<sup>54</sup> (<https://osf.io/tznwe/>). The predicted results were obvious because of the rationale of the study, the previous publications of our model, our previous behavioral and neuroimaging results<sup>4,15,28,33,36,41,53</sup> and above all, the results already obtained for replicas 1 and 2. However, to contribute for a larger use of pre-registration in our field as a way to address the current issue of lack of reproducibility, we pre-registered replica 3, following the template proposed at OSF (Open Science Framework). In addition to the suggested descriptions of the tests, analysis and the predicted results, we also included a short theoretical background to clearly define the rationale of the study to understand the predicted results. We believe that a critical factor to increase the practice of pre-registration is to keep the procedure practical (i.e., not excessively time consuming) at this initial stage of the research process. Our pre-registration included the critical information to prevent p-hacking and post hoc interpretations but was not perfectly polished. Four imperfections were detected and are worth mentioning: 1) the use of “unilateral” instead of “bilateral” statistical tests (bilateral tests were used); 2) time of naps: we had initially planned to provide naps immediately after the training still in the morning but for practical and physiological reasons we postponed them for the after lunch period (for all replicas), as reported here; 3) an extra computerized reading test was run for some participants (28 in total) but due to technical difficulties with the voice recorder plug-in this test was abandoned, and the data were not analyzed; finally, 4) the ideal five testing points initially planned were abandoned when we were confronted with the fact that the children were not reading yet at baseline. These errors were caused by a missed update of information, because the pre-registration derived from a recycle of the initial proposal file. However, these imperfections in the pre-registration do not alter the quality, nor any of the conclusions of the present study.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses and plots were performed using R software (version 1.2.1335). First, we performed a normality test for the results of each task (Data S1C). As not all data turned out to be normally distributed, we used non-parametric statistical tests to compare the group scores. Kruskal-Wallis tests for group effects with significant p values were followed by Wilcoxon rank-sum tests, with  $\alpha = 0.0083$  after Bonferroni correction for the six possible pairwise comparisons across the four groups (C, AC, T, T+S).

Two exclusion criteria were used to guarantee the quality of the data for the visual task. First, we excluded participants who failed, in any of the sessions, to press each of the two computer keys used in the test at least once, a situation that could reflect for instance a lack of task understanding or a potential technical problem (1<sup>st</sup> Replica: 24%; 2<sup>nd</sup> Replica: 16%; 3<sup>rd</sup> Replica: 0%). Second, trials with no response were excluded.

Blind correction (for group identity) was used for the writing test to avoid potential subjective bias.

## Extended analysis with data collapsed

Here we present analyses with data from all three replicas collapsed together. The experimental design crossed four experimental factors with fixed effect, being three of them with repeated-measures, and a replica variable concerning the batch of data collection. Intervention is the whole-plot / between subjects factor, whose levels are control - C, active control - AC, trained - T, or trained followed by sleep - T+S. The time factor indicates when a trial was executed in relation to application of intervention: session 1 for baseline, session 2 for immediately after the treatment and session 3 for four months after the treatment (and this for all measures, except for reading scores in replica 1 that were obtained ~12 months after the intervention). In sequence, category represents an experimental factor informing if the stimuli for a trial were letters, or symbols. At last, condition relates to the relationship of the pair of stimuli in a given trial, as different, mirrored or same. For inferential purpose the subjects were considered as a random factor turning to

experimental design into a mixed type. In order to explore the generalization of learning beyond the letters used for the target training (mirror-letters), an additional split-plot / within subjects factor (called training) was occasionally considered for the subset of trials with mirrored letters under interventions other than control.

From 3 replicas, data were collected from 102 subjects. The statistical significance of factors' effects was tested under the scope of (generalized) linear mixed models. According to its binary nature, the error rate response was analyzed by a mixed logistic regression. The corresponding association with predictors (experimental factors) was measured by odds ratio, whose statistical significance was evaluated by Wald tests. For responses compatible with normality and variance homogeneity, factors' effects were tested by linear mixed models followed by Satterthwaite's method for ANOVA. For all these cases, model fitting was confirmed by residual analysis. At last, responses unsuited to parametric modeling were tested by nonparametric ANOVA type for longitudinal data. In this context, Wilcoxon Mann-Whitney tests were used for post hoc multiple comparisons after Bonferroni correction. As usual, the significance level was set at 0.05. All statistical analysis was carried out in RStudio.

### Response times

Intervention produced a statistically significant effect on response time, but this effect was dependent on the levels of all remaining factors (Data S1E; Satterthwaite's ANOVA for LMM:  $SS = 3.106$ ;  $DF = 12/1601.07$ ;  $F = 2.641$ ;  $p = 0.0017$ ). This four factors interaction is illustrated in Figure S4, where difference among groups only took shape at sessions 2 and 3 and when the paired stimuli were mirrored letters. By comparing (pointwise 95%) confidence intervals, the effect of training (T) stands out as reducing the response time immediately after such treatment had been applied (session 2), but returning closer to control groups after four months (session 3). In turn, the reduction in response time was even more intense among participants subjected to training plus sleep (T+S) intervention at session 2, yet this effect persisted when reevaluated four months later (session 3). It is worth noting that the statistical power of a post hoc test is lower than that of ANOVA and its precise value (along with the significance level) depends on the amount of multiple comparisons of interest.

The interaction effect described above endorses the hypothesis that some levels of intervention improved the speed to respond to the stimuli in a distinct time frame. However, such learning was restricted to mirrored letters, indicating that training with letters was not generalized to symbols. To explore whether there was a generalization to other letters not used for training, only the subset of trials from AC, T and T+S interventions with mirrored letters was considered. Under these circumstances, the performance (regarding response time) through the sessions was still dependent on the level of intervention (Data S1F; nonparametric ANOVA type: statistic = 29.248, numDF = 3.627,  $p = 2.538E-22$ ). However, we could not find evidence enough to endorse better results for the trained letters (mirror-letters) over the non-trained letters (other letters) (Data S1F; nonparametric ANOVA type, all  $p > 0.05$  for main and interactions effects). The similarity between response times from trained and not trained letters is depicted in Figure 4, showing that both situations present very comparable distributions. Thus, the generalization of learning to non-trained letters is a quite reasonable inference.

### Error rate

As demonstrated by the response times, the effect of experimental factors on the error rate took shape as an interaction of all four factors (Figure S4). As a result of such complexity and given the fitted model (Data S1G; individual odds ratios become a little misleading. For instance, the odds ratio for an incorrect classification of mirrored letters at session 2 can be estimated as 5.992 as compared to same symbols at session 1. Important however is the interaction of all factors manifesting as an (estimated) decrease of such odds ratio by 92.60% among participants subjected to training intervention (Data S1H; Wald test:  $p = 0.00033$ ). Under similar settings, reduction was even more prominent when intervention consisted of training followed by sleep: 98.64% decrease (Data S1H; Wald test:  $p = 1.51E-10$ ). On the other hand, the odds ratio for incorrect mirror letters identification at session 2 (as compared to same symbols at session 1) remained stable under control or active control interventions, once we lacked evidence to support that the multiplicative factor representing interaction is statistically different from 1 in both contexts (Data S1H).

At session 3, four months after the intervention had been applied, the effect of experimental factors remained relatively stable. The exception was training intervention, whose association with incorrect mirrored letters identification faded a little relative to the previous session. If the odds ratio for incorrect identification of mirrored letters at session 3 (as compared to same symbols at session 1) was conceived at 7.772, it decreased 86.28% (Data S1H; Wald test:  $p = 0.000143$ ) under training intervention. As before, the multiplicative adjustment for control and active control interventions were devoid of statistical significance, hence assumed as being equal to 1 (Data S1H). In turn, the odds ratio for the incorrect classification of mirrored letters was multiplied by 0.002673 (99.73% reduction) among participants subjected to training and sleep (Data S1H; Wald test:  $p = 1.67E-8$ ).

In short, only training and training followed by sleep levels of interventions are associated with improvement of the incorrect classification of the stimuli. Yet, this effect was restricted to mirror condition for letters category. However training with and without sleep diverge from each other regarding the stability of their effects: without sleep the effect weakens four months later, while it remains consolidated when sleep is combined with training.

In terms of error rate and response times, Figure S4 makes even clearer that the learning based on letters was not generalized to symbols. In order to investigate whether there was a generalization of the intervention effects to the processing of letters not used for training, only the subset of trials from AC, T and T+S interventions with mirrored-letters versus other letters was considered. In this context however there was no evidence to endorse an association of trained letters to the rate of incorrect classification of stimuli, nor an influence on intervention's effect (Data S1J). These results suggest a generalization of mirror discrimination learning to other letters (but not to symbols).

### D primes

We calculated individual  $d'$  values [ $Z_{\text{score hits}} - Z_{\text{score False alarms (FA)}}$ ] separately for mirror trials and different trials.

#### D prime for mirrored stimuli

$D'$  for mirror stimuli was defined in the following way: hit = mirror trial, answered different and false alarm = same trial, answered different.

For  $d'$  prime for mirrored stimuli, the effect of intervention through sessions depended on the category of stimuli, as indicated by the statistical significance of a three factors interaction (Data S2A; Satterthwaite's ANOVA for LMM:  $SS = 21.769$ ;  $DF = 6/490$ ;  $F = 4.961$ ;  $p = 6.042E-10$ ). By comparing (pointwise) confidence intervals, it becomes clear that the intervention's effect on  $d'$  prime for mirrored stimuli was limited to letters, indicating that a generalization to symbols did not occur. In addition, training increased  $d'$  prime immediately after the intervention had been applied. However the significance of such effect faded after four months, so that the average of  $d'$  prime for mirrored stimuli from participants subjected to training was comparable to control peers at session 3. In turn, applying sleep after training sustained  $d'$  prime for mirrored stimuli at a higher magnitude (even when compared to the training intervention) for the remainder of the experiment. Four months later, the average  $d'$  prime for the combination of training and sleep was still found to be greater than that for training without sleep.

#### D prime for different stimuli

$D'$  for different stimuli was defined in the following way: hit = different trial, answered different and false alarm = same trial, answered different.

Contrasting the intricate three factors interaction observed on  $d'$  prime for mirrored stimuli, only the time showed a statistical significant effect when different stimuli were taken in account (Data S2B; Satterthwaite's ANOVA for LMM:  $SS = 26.280$ ;  $DF = 2/490$ ;  $F = 13.675$ ;  $p = 1.66E-06$ ). The average for each group, along with corresponding (pointwise) confidence intervals, is presented in Figure S5. Accordingly, average  $d'$  prime seems to decrease in session 2, remaining stable until four months later. Yet, this pattern is independent of both intervention and category of stimuli. Any variability confronting this inference is likely to emerge from sampling errors.

### Betas

We calculated individual beta values [ $\exp(-Z_{\text{score hits}} * Z_{\text{score hits}} / 2 + Z_{\text{score FA}} * Z_{\text{score FA}} / 2)$ ] where beta = 1 reflects a lack of bias and beta > 1 reflects an overall tendency to answer "same."

#### Beta for mirrored stimuli

There was a significant effect of intervention on beta for mirrored stimuli, being such effect unfolded across sessions in a specific way according to the category of the stimuli (Data S2C; nonparametric ANOVA type: statistic = 9.926, numDF = 5.515,  $p = 2.869E-10$ ). This three factors interaction is illustrated in Figure S6. Accordingly, the influence of intervention through the sessions was limited to letters. For this category of mirrored stimuli, the distribution of beta under training with and without sleep concentrates at lower values than under active control at session 2 (Data S2D; Wilcoxon Mann-Whitney with Bonferroni correction:  $p = 1.688E-06$  and  $p = 1.776E-04$ , respectively). At this moment, the distribution of beta from participants subjected to training without sleep was indistinguishable from that of peers with sleep combined to training (Data S2D; Wilcoxon Mann-Whitney with Bonferroni correction:  $p = 1.000$ ). However, four months later, the distribution of beta for mirrored letters under training (without sleep) spread through a wider range of values, becoming comparable to the distribution of beta from subjects in the active control group (Data S2D; Wilcoxon Mann-Whitney with Bonferroni correction:  $p = 1.000$ ) and different from beta distribution for mirrored letters under training with sleep (Data S2D; Wilcoxon Mann-Whitney with Bonferroni correction:  $p = 1.421E-04$ ). Aside from this difference, beta (for mirrored letters) distribution among participants subjected to training and sleep remained largely dissociated from that of the active control group at session 3 (Data S2D; Wilcoxon Mann-Whitney with Bonferroni correction:  $p = 3.059E-06$ ).

#### Beta for different stimuli

Figure S6 summarizes the distribution of beta for different stimuli according to the groups as defined by the combination of factors' levels. Under the circumstances, beta changed mildly, although statistically significant, through the sessions (Data S2E; nonparametric ANOVA type: statistic = 3.920, numDF = 1.948,  $p = 0.021$ ). In addition, intervention also had a statistically significant effect on beta for different stimuli (Data S2E; nonparametric ANOVA type: statistic = 3.069, numDF = 2.691,  $p = 0.032$ ).

#### Indirect effect of mirror discrimination improvement through sleep

The Sobel test was used to address if the influence of improved mirror discrimination on reading speed was mediated by the duration of sleep that had occurred after training. When such improvement was measured in terms of response time, part of its effect did not provide enough evidence to support a direct component of its influence (Data S3B; Wald test for linear regression: coef =  $-0.691$ , SE = 3.165,  $t = -0.218$ ,  $p = 0.828$ ). In contrast, the indirect effect of response time improvement on reading speed through sleep was, in statistical terms, highly significant (Data S3B; Sobel test: coef =  $-17.909$ , SE = 2.906,  $Z = -6.164$ ,  $p = 7.104E-10$ ). As a result, the reading speed reduced, on average, 18.6 words per minute for every one second increase in response time improvement, an effect that was exclusively mediated by sleep duration.

The reading speed was also shown to be dependent on  $d'$  prime improvement for mirror letters (Data S3C; Wald test for linear regression: coef = 3.468, SE = 0.804,  $t = 4.315$ ,  $p = 3.820E-5$ ), so that the rate of read words (per minute) increased 3.468 for every unit increment on  $d'$  prime. When partitioned, such an influence resulted in both statistically significant direct (Data S3C; Wald test for linear regression: coef = 1.291, SE = 0.556,  $t = 2.320$ ,  $p = 0.022$ ) and indirect effect (Data S3C; Sobel test: coef = 2.177, SE = 0.640,  $z = 3.400$ ,  $p = 6.737E-4$ ). The indirect component exceeds the direct one, once 2.177 out of 3.468 increment on reading speed for a unit increase in  $d'$  prime was shown to be mediated by the duration of sleep.

Thus, both results for the potential most sensitive measure (RTs) and the potential most specific measure ( $d'$  mirror) of intervention effects converge to the conclusion that sleep duration mediated the long-term training impact on reading fluency.