Supporting Information

Supporting Information Corrected November 24, 2014

Brem et al. 10.1073/pnas.0904402107

SI Methods

Subjects. Thirty-eight German-speaking children aged between 5.7 and 7.1 years (20 female) participated, finishing both the behavioral and event-related potential (ERP) parts of the study. All children were in kindergarten during the training, where reading is not formally taught in Switzerland (reading training starts in grade 1, usually at the age of 7 years). Only "nonreading" children who were able to read no more than six out of a list of 30 short, high-frequency German words according to the word reading subtest of the Salzburger Lesetest (SLT) (1) were included in the study. The children remained nonreaders or displayed only very rudimentary reading skills by the end of the study (only three children were able to "decode" more than 10 words, even with no time limit for their reading attempts).

Twenty-eight of these children (18 female) also performed all functional MRI (fMRI) recordings, but two children were excluded because they fell asleep during fMRI recordings, and five children were excluded because of excessive head motion in at least one of three recording sessions. One child was excluded from EEG and behavioral analyses because of poor EEG data quality. Four left-handed children and one child with severe language retardation problems were excluded from all (behavioral, ERP, and fMRI) analyses. These exclusions left 32 healthy children (17 female, mean age = 6.5 ± 0.27 years, all right handed) for behavioral and ERP analysis and 16 children for the additional fMRI analysis; all children had an estimated IQ ≥80. Both parents of each child reported in a questionnaire whether any first-degree relative of the child was dyslexic (no formal diagnosis was required), and children with a dyslexic first-degree relative were considered at-risk of dyslexia. According to this information, 14 children came from families with a history of dyslexia, and 18 children came from families without a history of dyslexia. All children had normal or corrected-to-normal vision. Children and their families were contacted by handouts distributed at preschools. One parent of each child gave informed consent to the study, which was approved by the local ethics commission. Children received small presents after each recording.

Children were assigned randomly to groups starting either with the Graphogame (GG-first) or the control game(NC-first), and the groups were matched for age, gender, IQ, letter knowledge, receptive vocabulary and phonological skills, familial risk status, and an estimate of the socioeconomic background (Tables 1 and Table S1).

Study Design. ***Graphogame.*** The noncommercial, computerized Graphogame (GG) developed at the University of Jyväskylä, Finland (2, 22), introduced the association of graphemes and phonemes according to the frequency and consistency of a grapheme in a given language, starting with most frequent and consistent letter–sound (grapheme–phoneme) pairs (upper and lowercase) and moving to less consistent/less frequent and more complex graphemes, monosyllabic words, and pseudowords. The consistency of the introduced single grapheme–phoneme had been determined using the German CELEX database (3), based on the whole monosyllabic corpus.

The computerized game started with five introductory levels, in which the children learned the grapheme–phoneme association of four consonants (c) and one vowel (v) in upper and lowercase (N, R, T, S, O) and the subsequent blending of these phonemes in syllables, short (cvc) words (e.g., TOR, ROT), and pseudowords (e.g., SOT, NOS). After these introductory levels, all remaining graphemes were introduced, first in upper and then in lowercase (seven different phonemes are introduced per level). The aim of the game was for children to learn the association between each phoneme (which was presented audibly at the beginning of each trial) and the corresponding grapheme. The graphemes were presented within balls, which fell downward from the top of the screen. The child had to "catch" the correct grapheme for the current phoneme. The game was adaptive: The choices of available graphemes depended on the child's performance. Each phoneme was repeated five times per level. Children could proceed to the next level only when they had learned the set of phonemes of the present level and made no more than one incorrect response per phoneme within the present level and no more than five incorrect responses in total. The game also provided some support levels for the phonemes for which the child exhibited most problems. After introduction of the simple phonemes, diphthongs (e.g., "au," "ei") and complex phonemes (e.g., "ch" and "sch") were introduced. Then the game moved to syllables (cv, vc), short words (cvc) and pseudowords (cvc), consonant clusters (cc) such as ccv, vc, and ccvc, to ccvcc words and pseudowords, and finally to short exception words.

By gaming, the children could collect virtual animal stickers and put these stickers into their own zoo. The children played at their own speed and did not have to reach a specific level within the training period. The play occurred at home; the parents were responsible for ensuring that the child played regularly but only when the child was not tired and wanted to play. Parents were instructed to allow their children play the game for about 10 min per day, with a target of 45 min of active training time per week. (Training with Graphogame is very intensive, with much repetition of the items to be learned. Long training sessions are inadvisable of young children.) However, because of individual circumstances (sickness, holidays, tiredness, busy schedules with other activities, and other causes), there was some variation in how many minutes children played per day and per week.

Control game. The nonlinguistic, noncommercial computerized control game (NG) also was developed in the University of Jyväskylä (4). It teaches number knowledge, size and quantity estimations, ordering of numerals, and basic addition and subtraction (Fig. S1). The game is similar in design to the GG and is divided into different levels through which the child can advance only when responding with high accuracy in the present level.

Training Procedure. Before the first imaging session (T1) and before starting the training, all children were tested with an extended behavioral test battery. The midtest imaging sessions (T2) were conducted after the first training period, and the posttest imaging sessions (T3) took place after the second training period. Subjects always performed the same tasks during ERP and fMRI sessions, (for differences in the ERP and fMRI tasks, see "ERP and fMRI Stimulation and Task" in *Methods* and Fig. 1 in the main text). The order of the two experimental parts was the same at all test times and for both ERP and fMRI recordings of a given subject but was counterbalanced across subjects. The average interval (mean ± SD) between ERP and fMRI sessions (when both were performed) was 5.46 ± 2.8 days.

Behavioral Screening Battery. The behavioral screening test battery (Table 1) performed before starting training included assessments of IQ [Raven's colored matrices (5)], phonological skills (Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten, BISC) (6), and receptive vocabulary (a subset of for vocabulary and word comprehension test from the Marburger

Sprachverständnistest für Kinder, MSVT) (7). Further, letter knowledge and reading skills were tested (word-reading subtest of the SLT) (1) at T1, T2, and T3 to track training-related improvements (Fig. S2).

For the BISC, a cumulative score, computed from four subtests that test phonological awareness in a broad sense (6) (rhyming of word pairs, syllable segmentation, phoneme association, and phoneme extraction), was used for the supplementary analyses discussed in later sections. Parents completed a questionnaire of their own reading history (adult reading history questionnaire, ARHQ) (8). For two children, only the ARHQ of the mother was available. For all other children, the mean ARHQ score of both parents was computed and used for supplemental analyses to assess a child's familial risk of dyslexia.

ERP and fMRI Tasks. All word stimuli (W) were concrete nouns with an average occurrence of 16.1 per million words (range: 0.23–473.5) in contemporary publications (University of Leipzig: www.wortschatz.uni-leipzig.de). Nouns and false font stimuli (FFs) were matched for the main visual characteristics: character size, string length (3–5 characters), and number of ascenders/descenders. The first letter of each noun was capitalized according to German orthographic rules. Each word was seen once and was heard once in the course of the experiment. Auditory stimuli were presented through headphones. As a control condition for auditory system activation, the same words were spectrally rotated (4 kHz, plus 20 dB Hanning Stop Band filter implemented in the PRAAT program), resulting in nonintelligible sounds. The stimuli were presented either unimodally [visually (V) or auditorially (A)], or bimodally (visual and auditory stimuli presented simultaneously) whereby the visual and auditory stimuli could be congruent (same words, AVc) or incongruent (different words, AVi). Bimodal congruent and incongruent trials both required the same behavioral response—namely, pressing both response buttons; unimodal trials required pressing only either the left or the right button. Hand–modality pairings were counterbalanced across subjects but were constant for all recordings of the same subject.

For the fMRI analyses, all event types (A, V, AVc, and AVi) were modeled for both parts of the experiment (W and FF). The ERP and fMRI analyses were restricted to the unimodal visual trials, for which we had clear, a priori hypotheses.

fMRI Recording and Processing. Before the first fMRI testing, the scanning procedure was demonstrated to the children using a teddy bear. To protect children's hearing from scanner noise, we used a noise-insulation mat, as well as having the children wear earplugs and headphones. Involuntary head movements were minimized using head padding.

Selection of region of interest. Five consecutive spherical regions of interest (ROIs) (radius $r = 6$ mm) within the putative VWFS of the occipito-temporal cortex were chosen as in previous studies (9–13) but were adjusted for the brain anatomy of a young child as given by the pediatric (5–9.5 y) brain template (Cincinnati Children's Hospital Medical Center: CCHMC2_y, available at https://irc.cchmc.org). The nonoverlapping ROIs followed the fusiform gyrus over a slightly bent course with an anterior declination.

The MNI coordinates (x, y, z) of the ROI centers (R) were located at R1: ±52, −42, −18; R2: ±50, −54, −16; R3: ±48, −66, −14; R4: ±46, −78, −12; R5: ±38, −90, −10.

The present ROI approach, based on the anatomy of the fusiform gyrus, was chosen to circumvent circularities that may bias results from functional ROI selection, as discussed recently (14). However, in the present study functionally defined ROIs [functional clusters in the left and right occipito-temporal cortex from whole-brain ANOVA analyses at $P < 0.001$, $k \geq 15$; Fig. 2*B* and Table 2] yielded a pattern of activation similar to that obtained with the spherical ROIs. The MANOVA including the left and right clusters showed a main effect of hemisphere [$F(1,14) = 12.43$, $P = 0.003$, $\eta_p^2 = 0.47$] with more pronounced activity in the left hemisphere, an interaction of *condition* and *group* factors [$F(1,14) = 6.33$, $P = 0.025$, $\eta_p^2 = 0.31$] as well as the core interaction of *group* × *condition* × *test time* factors [$F(2,13) = 5.05$, $P = 0.024$, $\eta_p^2 = 0.44$].

fMRI statistics. We acknowledge that choosing liberal statistical thresholds may induce false positives (15). We justify the threshold selected here ($P < 0.001$, $k \geq 15$, uncorrected) because of (*i*) our clear a priori hypotheses of emerging activity within the visual word-form system (VWFS), (*ii*) the need to present the results of interactions between *condition* × *time* factors (double difference: W–FF Post vs. W–FF Pre) and *condition* × *time* × *training* (GG, NC) in a relatively small group of children; (*iii*) the use of a short-duration task with limited statistical power appropriate to the age of the children; and (*iv*) reference to other studies with children (e.g., refs. 16–18) that reported activations at similar or even lower thresholds. The core training effects seen after GG training in the left fusiform gyrus survived the stringent family-wise error correction at $P < 0.05$ (Table 2).

Activated brain structures were identified on the mean structural image of the group with the help of neuroanatomical atlases (http://www9.biostr.washington.edu/da.html, http://www.med.harvard.edu/AANLIB/home.html).

ERP Recording and Processing. The ERPs were recorded from 64 channels with impedances below 15 kOhm at a sampling rate of 500 Hz and filters set to 0.1–70 Hz. For EEG recordings, Fz served as the recording reference, AFz as ground. Children were seated in front of a computer screen (distance from screen, 120 cm), wore headphones, and used a mouse for responding. The electrode positions on the individual scalp were measured using a 3D digitizer.

For ocular artifact correction, an independent component analysis (ICA) (19) on 0.1- to 30-Hz filtered data was used. The ICA separates the input data into a sum of temporally independent components of fixed topography and allows components dominated by blink or lateral/vertical eye artifact distributions to be removed without topographic distortion. A minimum of 14 artifact-free epochs/segment (mean number of epochs: 35.4 ± 3.8; range: 15–42) was required to compute averages for each stimulus type. The number of epochs included in ERP analyses did not differ between conditions or test times. ERPs were transformed to the average reference (20) for all subsequent analyses.

The global field power (GFP) [i.e., the time-varying measure of map amplitude computed as the spatial root mean square] sinks in the waveform were used to define the N1 (195–289 ms) interval.

EEG montage. Caps used for the montage included all 10–20 system electrodes plus the following additional electrodes: FPz, FCz, CPz, POz, Oz, Iz, AF1/2, F5/6, FC1/2, FC3/4, FC5/6, FT7/8, FT9/10, C1/2, C5/6, CP1/2, CP3/4, CP5/6, P5/6, TP7/8, TP9/10, PO1/2, PO9/10, OI1/2, PPO9h/10h, and two EOG electrodes below the outer cantus of the each eye. O1′/2′ and Fp1′/2′ were placed 15% more laterally to Oz/Fpz for more even coverage. OI1 and OI2 were placed to the left and to the right of the midline halfway between Oz and Iz to provide better coverage of the occipital scalp distributions.

The topographic analysis of variance (TANOVA) was used in the present study to compare ERP map topographies. The TANOVA computes the exact probability of dissimilarity between two maps (21) using bootstrapping statistics and protects against possible errors caused by parametric statistics or nonrepresentative summary measures.

Task Performance, ERP, and fMRI. For analyses of task performance, we considered accuracy (% correct responses), reaction time, and a sensitivity index d' computed as the difference of the z-transform of the hit rate minus the z-transform of the false-alarm rate: $d'(\mathrm{W}) = z(\text{hit rate W}) - z(\text{false-alarm rate W})$, and $d'(\mathrm{FF}) = z(\text{hit rate FF}) - z(\text{false-alarm rate FF})$. The d' values for the detection of visual words (W) and false fonts (FF) in the implicit print and false font processing task were computed separately for ERP and fMRI task performance and are broken down by group, method, and test times in Table S2. For children who had an accuracy of 100% (hit rate =1), or a false-alarm rate of 0%, the values of 0.99 (hit rate) and (0.01 false-alarm rate) were inserted for d' computation. The higher the d', the better is the sensitivity. Because of technical problems, the behavioral fMRI data of two children at T1 and one child at T2 were not logged, leaving the data for 14 children at T1 and 15 children at T2 for fMRI performance analyses (*Results* and *SI Results*).

Task performance across test times and conditions indicated that, on average, children performed well despite the rapid sequence of stimuli requiring responses (Table S2).

SI Results

The following sections report the results of supporting analyses for the core effects. In summary, all additional analyses confirmed the core training effects in behavioral, ERP, and fMRI measures, as can be seen in Table S4 and the following sections. Supplementary results are not discussed further if they corroborated the main analyses; however, interactions with the covariates are reported and discussed.

Supplemental Analysis A: ERP and Behavioral Data of the fMRI Subgroup. To compare the ERP and behavioral training effects of the fMRI subgroup ($n = 16$) with the effects reported for the whole group ($n = 32$), the following analyses include only those children that formed the fMRI subgroup.

Letter knowledge, fMRI subgroup. As consistent with the main analyses, children named or sounded out significantly more uppercase than lowercase letters [$F(1,14) = 71.3$, $P < 0.001$, $\eta_p^2 = 0.84$], and their letter-naming performance increased with *test time* [$F(2,13) = 40.19$, $P < 0.001$, $\eta_p^2 = 0.86$]. The improvements in letter knowledge were driven primarily by practicing grapheme–phoneme correspondences [$F(2,13) = 7.02$, $P = 0.009$, $\eta_p^2 = 0.52$]. In the fMRI subgroup, no significant improvements in reading skills were found.

ERP results, fMRI subgroup. The additional analysis of ERP results in the fMRI subgroup confirmed the findings of the larger ERP sample reported in the main text: The main effect of *condition* [both hemispheres: $F(1,14) = 9.52$, $P = 0.008$, $\eta_p^2 = 0.41$; LOT: $F(1,14) = 12.31$, $P = 0.003$, $\eta_p^2 = 0.47$; ROT: $F(1,14) = 6.61$, $P = 0.022$, $\eta_p^2 = 0.32$] and the triple interaction of *time*, *condition*, and *group* [for both hemispheres: $F(2,13) = 5.92$, $P = 0.015$, $\eta_p^2 = 0.48$; for LOT: $F(2,13) = 11.4$, $P = 0.001$; $\eta_p^2 = 0.64$; for ROT: ns] remained significant for the smaller sample.

Supplemental Analysis B: Behavioral, ERP, and fMRI Data of Good Performers. The main analyses of the article included children with poor ERP/fMRI task performance as long as the children did not fall asleep but watched the stimuli on the screen and listened to the sounds. To examine whether poor performance, which might be caused by lack of attention, affected our results critically, we recomputed the main statistics with those children that performed well ($d' \geq 1$) during ERP or fMRI tasks. The summary of the core statistics in Table S4 shows that the core training effects in letter knowledge, N1 amplitude, and percent signal change (trend) in the VWFS remain, despite the small samples (ERP: $n = 19$; fMRI $n = 9$). ERP analyses also revealed a significant four-way interaction of *group* × *condition* × *time* × *hemisphere* that pointed to a more pronounced training effect in the left hemisphere than in the right hemisphere. The more pronounced N1 negativity to W than to FF stimuli and the interaction with *hemisphere* is explained by a the stronger N1 *condition* difference at LOT [$F(1,17) = 6.37$, $P = 0.022$, $\eta_p^2 = 0.27$], especially for the GG-first group [*hemisphere* × *condition* × *group*: $F(1,17) = 4.80$, $P = 0.043$, $\eta_p^2 = 0.22$], because in general the N1 amplitude was more pronounced in the left hemisphere for the GG-first group and in the right hemisphere for the NC-first group [*group* × *hemisphere*: $F(1,17) = 4.69$, $P = 0.045$, $\eta_p^2 = 0.22$].

Supplemental Analysis C: Behavioral, ERP, and fMRI Training Effects and Familial Dyslexia Risk. Children in this training study came from families with and without a history of developmental dyslexia. Even though the number of at-risk and no-risk children was similar in the different groups, we recomputed our main statistics and included the mean score of the parents' ARHQ, a continuous variable, as a covariate to control for the familial dyslexia risk of each child. The supplemental analyses corroborated our main results and showed that the familial risk status of the children did not affect the core effects. Furthermore, no main effect or interaction with the covariate was found.

Supplemental Analysis D: Behavioral, ERP, and fMRI Training Effects and Phonological Dyslexia Risk. Before starting the training procedures, all children were tested with a phonological screening test (BISC). The cumulative score of four subtests of the BISC that measure phonological awareness in a broad sense formed the continuous phonological risk score. This score was included in the supplemental analyses as a covariate to account for each child's phonological risk of dyslexia as measured before the start of training. As in the previous analyses, this covariate had little impact on our core effects. However, children with good phonological awareness exhibited a more pronounced N1 [main effect, *phonological awareness*; ERP: $F(1,29) = 6.2$, $P = 0.019$, $\eta_p^2 = 0.18$]. Furthermore, *phonological awareness* also modulated the interaction of *test time* and *condition* [$F(2,28) = 9.46$, $P < 0.001$, $\eta_p^2 = 0.4$]. In contrast, no impact of phonological awareness was detected for behavioral or fMRI data.

Supplemental Analysis E: Behavioral, ERP, and fMRI Training Effects and Differences in GG Training Time. Even though there was no significant group difference in the time the children in either the ERP group (trend) or the fMRI group spent playing GG (Table S1), the playing times of individual children differed. We therefore included an analysis with the *GG playing time* (in minutes) as a covariate to account for the differences in children's training times. In addition to our core effects, these analyses clarified that the increase in letter knowledge with *test time* depended on the time the children spent playing the GG [*time* × *GG playing time*: $F(2,27) = 6.54$, $P = 0.005$, $\eta_p^2 = 0.33$].

For the fMRI analyses of ROI 4 (R4), there was an interaction of the *GG playing time* with *condition* [*condition* × *GG playing time*: $F(1,13) = 6.68$, $P = 0.023$, $\eta_p^2 = 0.34$], which indicated that the significant condition difference in percent signal change [$F(1,13) = 5.93$, $P = 0.030$, $\eta_p^2 = 0.31$] was modulated by the playing time.

For the ERP, a significant interaction of *hemisphere*, *condition*, and *GG playing time* [*hemisphere* × *condition* × *GG playing time*: $F(1,28) = 4.61$, $P = 0.041$, $\eta_p^2 = 0.14$] pointed to an impact of *playing time* on *hemispheric* and *condition* differences. Furthermore, *hemisphere* reached significance with more pronounced N1 amplitudes in the left than in the right hemisphere [$F(1,28) = 4.39$, $P = 0.045$, $\eta_p^2 = 0.14$].

Supplemental Analysis F: Behavioral, ERP and fMRI Training Effects and Differences in the GG Training Period. Because the training periods differed for each child and because there was a significant difference in GG playing periods (days) in the fMRI sample (Table S1), our last analyses included the *GG playing period* as a covariate. Despite slightly weaker ERP and fMRI training effects than seen in our main analyses (statistical trends, Table S4), the core analyses again were confirmed. The differences in playing periods resulted from individual circumstances (sickness, holidays, tiredness, busy schedules with other activities, and other such causes). Despite the differences in GG playing periods and in the number of NC playing sessions (Table S1) for the fMRI group, the actual playing time in minutes (for GG and NC) did not differ between the GG-first and NC-first groups.

Supplemental Analysis G: Supplemental Analyses to Account for Group Differences in Print Sensitivity at T1. Because of the group differences in the print sensitivity of three posterior ROIs before training start (at T1), additional analyses were performed that included the differential percent signal change of W–FF at T1 (means over the left and right hemispheres of R3, R4, and R5) as covariate(s). These additional analyses confirming our core training effect are reported in the main text.

ANCOVAs for the fMRI signal change in R4 (left) with between-subject factor *group* (GG-first, NC-first) and within-subject factors *test time* (T1, T2, T3) and *condition* (W, FF) were repeated for all supplemental analyses and always included the *print sensitivity difference* at T1 as a second covariate. Here, only the core training effect (*condition* × *time* × *group*) in left R4 is listed for analyses B–F: (B) $F(2,5) = 4.31$, $P = 0.082$, $\eta_p^2 = 0.63$; (C) $F(2,11) = 4.63$, $P = 0.035$, $\eta_p^2 = 0.46$; (D) $F(2,11) = 4.93$, $P = 0.030$, $\eta_p^2 = 0.47$; (E) $F(2,11) = 4.52$, $P = 0.037$, $\eta_p^2 = 0.45$; (F) $F(2,11) = 3.36$, $P = 0.073$, $\eta_p^2 = 0.38$.

To summarize, the additional analyses with the baseline covariate supported our core findings.

1. Landerl K, Wimmer H, Moser E (1997) *Der Salzburger Lese- und Rechtschreibtest (SLRT)* (Huber, Bern, Switzerland).
2. Lyytinen H, Ronimus M, Alanko A, Poikkeus AM, Taanila M (2007) Early identification of dyslexia and the use of computer game-based practice to support reading acquisition. *Nord Psykol* 59:109–126.
3. Baayen RH, Piepenbrock R, van Rijn H (1993) *The CELEX Lexical Database* (Linguistic Data Consortium, University of Pennsylvania, Philadelphia).
4. Räsänen P, Salminen J, Wilson AJ, Aunio P, Dehaene S (2009) Computer-assisted intervention for children with low numeracy skills. *Cogn Dev* 24:450–472.
5. Raven JC (2002) *Coloured Progressive Matrices (CPM)* (Harcourt Test Services, Frankfurt).
6. Jansen A, Mannhaupt G, Marx H, Skowronek H (1999) *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC)* (Hogrefe, Göttingen, Germany).
7. Elben CE, Lohaus A (2000) *Marburger Sprachverständnistest für Kinder* (Hogrefe, Göttingen, Germany).
8. Lefly DL, Pennington BF (2000) Reliability and validity of the adult reading history questionnaire. *J Learn Disabil* 33:286–296.
9. Brem S, et al. (2006) Evidence for developmental changes in the visual word processing network beyond adolescence. *Neuroimage* 29:822–837.
10. Brem S, et al. (2009) Tuning of the visual word processing system: Distinct developmental ERP and fMRI effects. *Hum Brain Mapp* 30:1833–1844.
11. James KH, James TW, Jobard G, Wong AC, Gauthier I (2005) Letter processing in the visual system: Different activation patterns for single letters and strings. *Cogn Affect Behav Neurosci* 5:452–466.
12. van der Mark S, et al. (2009) Children with dyslexia lack multiple specializations along the visual word-form (VWF) system. *Neuroimage* 47:1940–1949.
13. Vinckier F, et al. (2007) Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron* 55:143–156.
14. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: The dangers of double dipping. *Nat Neurosci* 12:535–540.
15. Wright ND, et al. (2008) Selective activation around the left occipito-temporal sulcus for words relative to pictures: Individual variability or false positives? *Hum Brain Mapp* 29:986–1000.
16. Booth JR, Cho S, Burman DD, Bitan T (2007) Neural correlates of mapping from phonology to orthography in children performing an auditory spelling task. *Dev Sci* 10:441–451.
17. Brambati SM, et al. (2006) Neuropsychological deficits and neural dysfunction in familial dyslexia. *Brain Res* 1113:174–185.
18. Cao F, Bitan T, Chou TL, Burman DD, Booth JR (2006) Deficient orthographic and phonological representations in children with dyslexia revealed by brain activation patterns. *J Child Psychol Psychiatry* 47:1041–1050.
19. Jung TP, et al. (2000) Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol* 111:1745–1758.
20. Lehmann D, Skrandies W (1980) Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr Clin Neurophysiol* 48:609–621.
21. Strik WK, Fallgatter AJ, Brandeis D, Pascual-Marqui RD (1998) Three-dimensional tomography of event-related potentials during response inhibition: Evidence for phasic frontal lobe activation. *Electroencephalogr Clin Neurophysiol* 108:406–413.
22. Lyytinen H, Erskine J, Kujala J, Ojanen E, Richardson U (2009) In search of a science-based application: A learning tool for reading acquisition. *Scand J Psychol* 50(6):668–675.

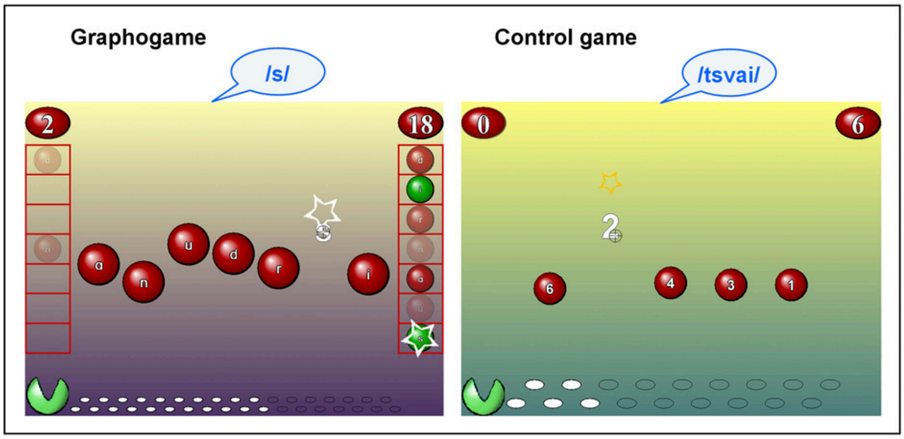

Fig. S1. Screen shots of the two training games used in this study. (*Left*) In the Graphogame, the child heard the phoneme /s/ at the beginning of the trial and then had to use the “catcher” to select the corresponding grapheme from the different choices given (on balls with graphemes). If the choice was incorrect, the correct ball was highlighted, and the phoneme sound was repeated. (*Right*) In the control game the child had to match the correct numeral “2” to the spoken number word “two.”

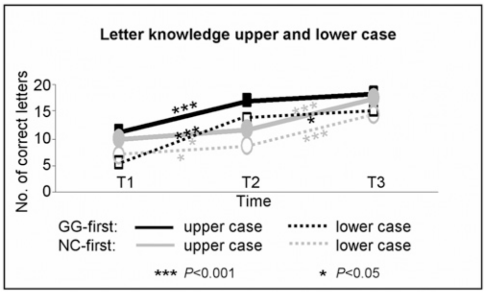

Fig. S2. Training effects for Graphogame (GG) and control game (NC) on letter knowledge, assessed by the number of correctly named or pronounced upper and lowercase letters.

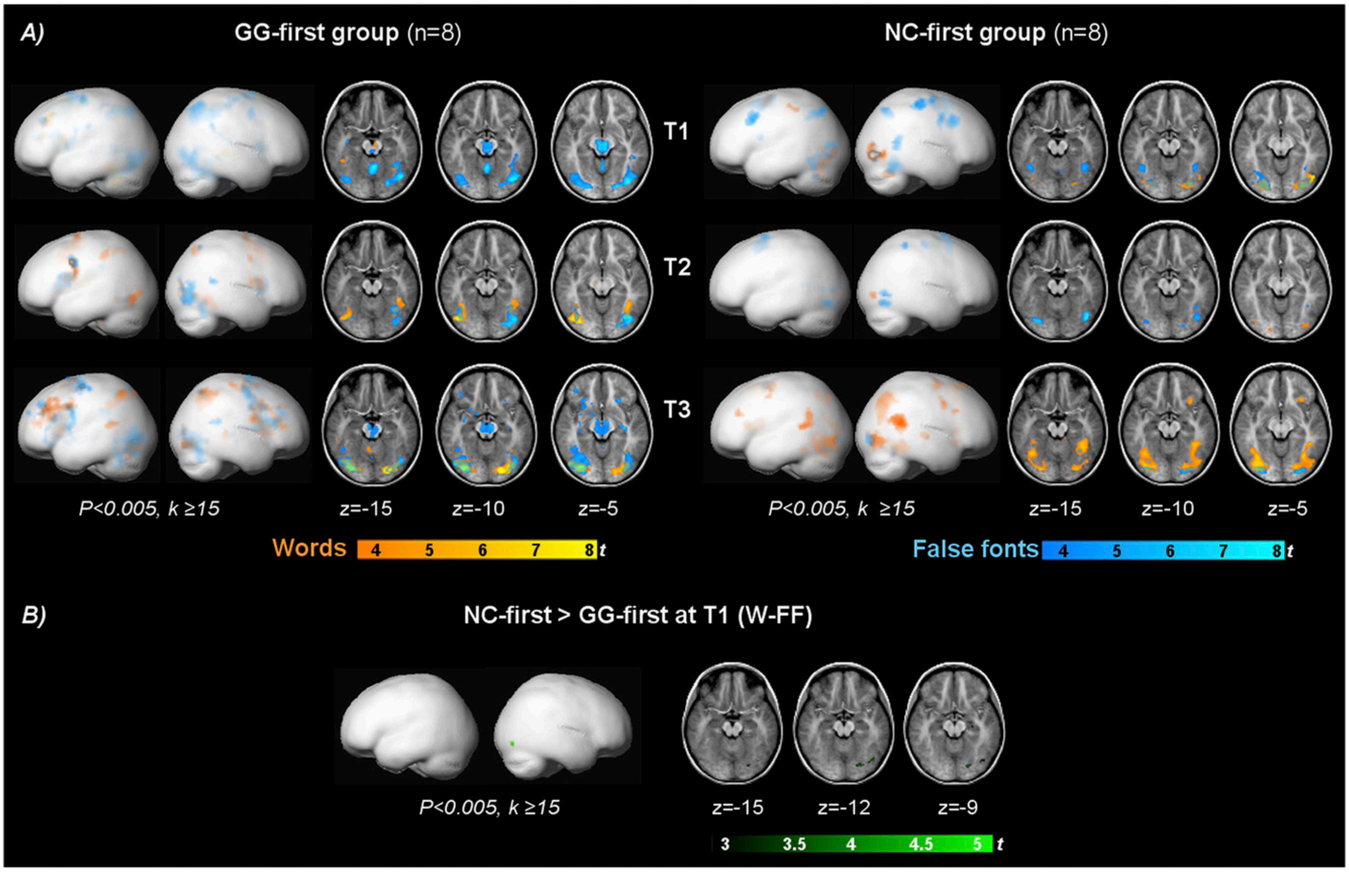

Fig. S3. (*A*) Whole-brain analyses for GG-first and NC-first groups. Illustrated are the activation to words (orange) and false fonts (blue) for both GG-first and NC-first groups at each test time (one-sample *t* tests, $P < 0.005$, $k \geq 15$). To the right of each group, axial slices with the activation thresholded at $t \geq 3.5$, $k \geq 0$ are shown at $z = -15$, -10, and -5. (*B*) Group differences in print sensitivity before training. The activation difference of the main contrast (W–FF) between groups (NC-first > GG-first) is shown in green ($P < 0.005$, $k \geq 15$). (*Left*) One cluster ($k = 16$) in the right inferior occipital gyrus exhibited more pronounced activity to W than to FF stimuli ($Z = 3.2$; MNI: $x = 24$, $y = -87$, $z = -12$) for the NC-first subgroup at T1. No region exhibited a more pronounced difference in the GG-first group. (*Right*) The group difference, thresholded at $t \geq 3$, $k \geq 0$, is shown on axial slices at $z = -15$, -12, and -9. The activation in *A* and *B* is projected onto a pediatric template (5–9.5 y) provided by the Cincinnati Children's Hospital Medical Center (*Left*) and also is overlaid on three axial slices of the mean structural image of the group (*Right*).

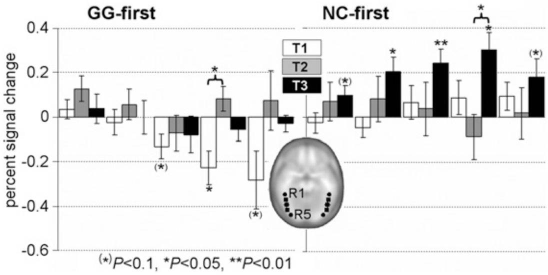

Fig. S4. Print sensitivity within ROIs along the occipito-temporal cortex. Mean percent signal change of the W–FF contrast (error bars, ±1 SEM) for GG-first (*Left*) and NC-first (*Right*) groups in five consecutive spherical ROIs (from left to right: R1–R5) plotted for the left hemisphere at each test time.

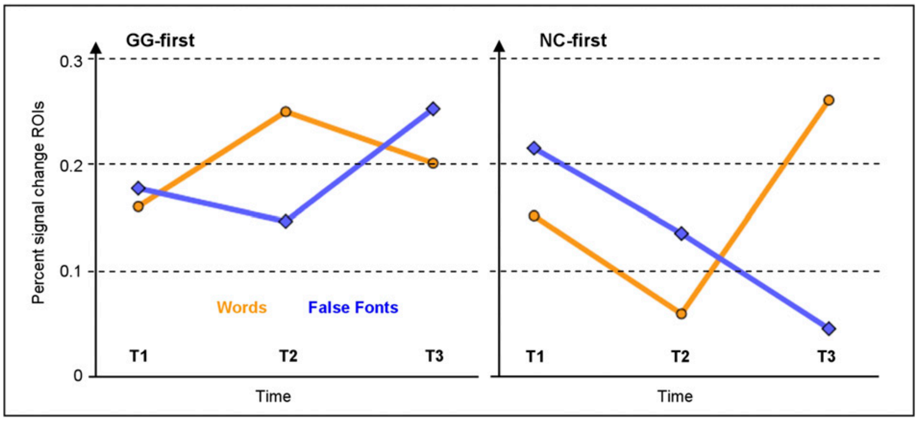

Fig. S5. Interaction plot showing the training effect within the occipito-temporal cortex. Interaction plot of the MANCOVA on percent signal change including all five bilateral ROIs and the three baseline covariates [*time* x *condition* x *group*, $F(2,10) = 6.11$, $P = 0.018$, $\eta_p^2 = 0.55$] to account for T1 *condition* differences between the GG-first (*Left*) and the NC-first (*Right*) subgroups. The changes over the three test times (T1, T2, T3) are illustrated for W (orange) and FF (blue) stimuli.

Table S1. Summary of training intensities with Graphogame (GG) and control training (NC) for the ERP and the fMRI groups

	Graphogame (GG) mean ± SD			Control game (NC) mean ± SD			Training statistics (*t* tests) (*P* value)		
Group	Time (min)	No. of sessions	Training period (days)	Time (min)	No. of sessions	Training period (days)	Time	No. of sessions	Training period
ERP group									
All (*n* = 32)	224.4 ± 95.6*	24.9 ± 9.6*	58.1 ± 12.5	217.8 ± 91.1	21.9 ± 8.1	56.9 ± 14.1	0.68	0.089	0.68
GG-first (*n* = 15)	255.2 ± 113.5	25.9 ± 11.0	59.8 ± 7.1	201.8 ± 85.6	18.6 ± 7.9	59.8 ± 18.7	0.04[†]	0.001[‡]	1
NC-first (*n* = 17)	195.5 ± 66.4*	23.9 ± 8.3*	56.6 ± 15.9	231.8 ± 96.1	24.8 ± 7.3	54.3 ± 7.9	0.09	0.54	0.55
P (groups)	0.082	0.56	0.48	0.36	0.027[†]	0.28			
fMRI group									
All (*n* = 16)	233.2 ± 80.6	26.1 ± 9.9	54.4 ± 8.2	225.6 ± 85.6	23.1 ± 7.6	54.7 ± 8.7	0.77	0.21	0.94
GG-first (*n* = 8)	250.8 ± 95.2	27.4 ± 11.8	58.8 ± 7.2	199.6 ± 71.6	19.0 ± 6.7	53.8 ± 9.9	0.22	0.018[†]	0.30
NC-first (*n* = 8)	215.5 ± 64.4	24.9 ± 8.3	50.1 ± 7.1	251.5 ± 95.0	27.3 ± 6.3	55.6 ± 7.9	0.22	0.39	0.17
P (groups)	0.4	0.63	0.030[†]	0.24	0.024[†]	0.68			

*Because of technical problems, the accurate total playing time and number of sessions for Graphogame were missing for one child in the NC-first group.

[†] $P < 0.05$.

[‡] $P < 0.001$.

Table S2. Performance in the implicit word and false font processing task for ERP and fMRI recordings

Group	Performance parameter	Words (mean ± SD)			False fonts (mean ± SD)		
		T1	T2	T3	T1	T2	T3
ERP group							
All (*n* = 32)	%	76 ± 17	78 ± 17	76 ± 15	72 ± 20	73 ± 18	69 ± 17
	d'	1.89 ± 1.08	2.57 ± 0.83	2.38 ± 0.84	2.0 ± 1.16	2.28 ± 1.07	2.18 ± 0.92
	RT	1369 ± 314	1437 ± 325	1429 ± 283	1363 ± 265	1435 ± 405	1480 ± 323
GG-first (*n* = 15)	%	77 ± 18	75 ± 18	74 ± 15	72 ± 24	71 ± 21	69 ± 19
	d'	1.97 ± 1.0	2.47 ± 0.90	2.5 ± 0.93	1.98 ± 1.21	2.28 ± 1.18	2.31 ± 1.03
	RT	1384 ± 372	1477 ± 389	1518 ± 302	1380 ± 271	1459 ± 450	1542 ± 317
NC-first (*n* = 17)	%	75 ± 17	81 ± 15	78 ± 15	72 ± 16	75 ± 15	70 ± 15
	d'	1.82 ± 1.17	2.66 ± 0.78	2.28 ± 0.77	2.03 ± 1.16	2.28 ± 1.0	2.06 ± 0.83
	RT	1356 ± 263	1403 ± 265	1350 ± 248	1348 ± 268	1414 ± 374	1425 ± 328
P (groups)	%	ns	ns	ns	ns	ns	ns
	d'	ns	ns	ns	ns	ns	ns
	RT	ns	ns	0.095	ns	ns	ns
fMRI group							
All (*n* = 16)	%	76 ± 12*	72 ± 18	78 ± 17	77 ± 16*	69 ± 16[†]	71 ± 17
	d'	2.23 ± 0.61*	2.39 ± 1.07	2.35 ± 0.91	2.17 ± 0.62*	2.12 ± 0.81[†]	2.13 ± 1.23
	RT	1473 ± 301*	1459 ± 295	1428 ± 326	1401 ± 270*	1518 ± 356[†]	1420 ± 324
GG-first (*n* = 8)	%	76 ± 13[‡]	80 ± 15	78 ± 17	80 ± 17[‡]	73 ± 16[‡]	72 ± 22
	d'	2.1 ± 0.83[‡]	2.71 ± 1.18	2.41 ± 1.09	2.3 ± 0.66[‡]	2.03 ± 0.85[‡]	2.21 ± 1.61
	RT	1456 ± 287[‡]	1458 ± 372	1403 ± 320	1412 ± 256[‡]	1489 ± 467[‡]	1493 ± 389
NC-first (*n* = 8)	%	77 ± 11[‡]	64 ± 19	78 ± 18	74 ± 15[‡]	65 ± 17	71 ± 13
	d'	2.37 ± 0.28[‡]	2.07 ± 0.91	2.28 ± 0.76	2.04 ± 0.61[‡]	2.2 ± 0.82	2.05 ± 0.78
	RT	1490 ± 337[‡]	1459 ± 220	1452 ± 353	1392 ± 304[‡]	1543 ± 255	1348 ± 248
P (groups)	%	ns	0.087	ns	ns	ns	ns
	d'	ns	ns	ns	ns	ns	ns
	RT	ns	ns	ns	ns	ns	ns

ns, nonsignificant *t* test between groups ($P > 0.1$; trends are reported); RT, reaction time in ms; %, percent correct responses.

*Data for 14 children.

[†] Data for 15 children.

[‡] Data for 7 children.

Table S3. **Activation maxima for words, false fonts, and the main contrast (W vs. FF) before and after graphogame and control training**

Condition		H	Area	*x*	*y*	*z*	Z	*k*
Graphogame								
W	Pre	L	Medial frontal gyrus	−3	9	60	4.52	361
		L	Fusiform gyrus	−45	−54	−21	4.01	55
		R	Fusiform gyrus	36	−60	−12	3.92	89
		L	Insula	−30	15	21	3.85	40
		L	Thalamus	−15	−15	12	3.61	45
		L	Cingulate gyrus	−6	−27	30	3.51	53
	Post	L	Fusiform gyrus	−39	−78	−6	5.34	392
		L	Medial frontal gyrus	−3	6	60	5.03	956
		R	Fusiform gyrus, inferior occipital gyrus	42	−51	−9	5.02	605
		L	Insula	−30	6	18	4.88	369
		R	Precuneus, superior parietal lobule	33	−75	30	4.61	288
		R	Precentral gyrus, middle frontal gyrus	48	6	42	4.3	152
		L	Superior parietal lobule, inferior parietal lobule	−30	−63	54	4.19	130
		R	Insula, inferior frontal gyrus	33	18	15	4.13	198
		R	Superior temporal gyrus, supramarginal gyrus	66	−51	15	4.06	126
		L	Supramarginal gyrus, inferior parietal lobule	−51	−42	51	3.99	42
		L	Thalamus	−15	−15	15	3.91	44
		R	Cerebellum	18	−57	−36	3.9	38
		L	Precentral gyrus, middle frontal gyrus	−51	3	42	3.85	119
		R	Thalamus	12	−15	9	3.84	68
FF	Pre	R	Fusiform gyrus, inferior occipital gyrus, cerebellum	45	−75	−12	5.87	1493
		L	Fusiform gyrus, inferior occipital gyrus, cerebellum	−39	−81	−15	5.06	
		L	Medial frontal gyrus, superior frontal gyrus	−3	12	57	5.35	919
		L	Insula	−33	21	18	4.3	153
		R	Precentral gyrus, postcentral gyrus	42	−21	66	4.22	105
		R	Precentral gyrus	48	0	45	4.1	84
		R	Inferior parietal lobule, superior parietal lobule	54	−48	60	4.03	181
		R	Thalamus	9	−18	−3	4.01	66
		R	Insula, inferior frontal gyrus	48	12	12	4.01	149
		L	Precentral gyrus	−30	−9	63	3.55	32
	Post	R	Inferior occipital gyrus, cerebellum, fusiform gyrus	36	−90	−9	4.99	408
		R	Medial frontal gyrus	9	12	63	4.15	337
		R	Superior temporal sulcus	57	−48	15	3.93	61
		L	Precentral gyrus	−60	0	51	3.49	36
		L	Insula	−36	15	21	3.47	53
W > FF*	Pre		No activation at $P < 0.005$, $k = 15$					
	Post	L	Inferior frontal gyrus	−54	27	0	3.33	18
		L	Fusiform gyrus	−45	−78	−15	3.32	19
		R	Fusiform gyrus	57	−63	−12	3.06	21

Table S3. Cont.

Condition		H	Area	*x*	*y*	*z*	Z	*k*
Control game								
W	Pre	L	Medial frontal gyrus	−6	0	66	5.51	583
		L	Inferior occipital gyrus, fusiform gyrus	−36	−87	−3	4.96	236
		R	Fusiform gyrus, inferior occipital gyrus	30	−90	−9	4.55	368
		L	Supramarginal gyrus, inferior parietal lobule	−54	−39	54	4.32	45
		R	Insula	30	18	15	4.01	75
		L	Insula	−33	15	18	3.99	79
		R	Cerebellum	9	−54	−33	3.94	66
	Post	L	Fusiform gyrus, inferior occipital gyrus	−39	−84	−12	4.6	405
		R	Inferior occipital gyrus, fusiform gyrus, cerebellum	36	−90	−6	4.5	277
		L	Medial frontal gyrus	−9	9	57	4.46	443
		R	Inferior parietal lobule, precuneus	36	−57	48	4.06	188
		L	Inferior parietal lobule	−39	−60	54	3.93	106
		R	Postcentral gyrus	42	−24	66	3.69	42
		L	Insula, inferior frontal gyrus	−33	15	21	3.64	87
		L	Cingulate gyrus	−6	−30	33	3.58	43
		R	Precentral gyrus, middle frontal gyrus	48	0	39	3.49	105
		L	Middle frontal gyrus	−45	39	36	3.44	31
FF	Pre	R	Fusiform gyrus, inferior occipital gyrus, cerebellum	33	−87	−9	5.29	632
		L	Insula	−36	9	18	4.63	168
		R	Medial frontal gyrus	6	12	63	4.62	655
		L	Fusiform gyrus, inferior occipital gyrus	−42	−60	−9	4.38	301
		R	Precentral gyrus, middle frontal gyrus	51	3	42	4.29	159
		R	Superior temporal sulcus	54	−51	15	4.17	90
		R	Inferior parietal lobule, precuneus	36	−72	51	4.09	239
		L	Inferior parietal lobule	−39	−60	51	4.05	156
		R	Insula, inferior frontal gyrus	33	15	15	4.04	135
		L	Precentral gyrus	−45	−3	36	3.8	42
	Post	L	Superior frontal gyrus, medial frontal gyrus, cingulate gyrus	−9	0	75	5.62	2180
		R	Fusiform gyrus, inferior occipital gyrus	48	−75	−15	5.62	556
		L	Fusiform gyrus, cerebellum, lingual gyrus	−36	−87	−12	5.25	604
		R	Thalamus	9	−15	−6	4.18	77
		L	Precentral gyrus	−30	−9	60	3.7	55
		L	Cerebellum	−6	−51	−18	3.45	54
W > FF*	Pre	L	Parahippocampal gyrus	−21	−27	−18	3.01	25
	Post		No activation at $P < 0.005$, $k = 15$					

Listed are MNI coordinates (x, y, z), Z-values and cluster sizes (k) of cluster maxima at the threshold $P < 0.001$, $k \geq 30$ (uncorrected). FF, false fonts; H, hemisphere; L, left; Pre, before training; Post, after training; R, right; W, words.

*Condition contrasts (W–FF) are listed for $P < 0.005$, $k \geq 15$ (uncorrected). Not listed are activations in brainstem.

Table S4. **Summary of *SI Results*, supplemental analyses B–F**

	(B) Good performers			(C) Familial risk			(D) Phonological risk			(E) GG playing time			(F) GG training period		
MAN(C)OVAs for the number of correctly named/ pronounced letters with between-subject factor *group* (GG-first, NC-first) and within-subject factors *test time* (T1, T2, T3), *letter case* (uppercase, lowercase)															
Behavior	F	*P*	η_p^2	F	*P*	η_p^2	F	*P*	η_p^2	F	*P*	η_p^2	F	*P*	η_p^2
LK (n = 32)															
T	(2,16) 37.89	<0.001***	0.83	(2,28) 7.62	0.002**	0.35	(2,28) 0.06	0.95	0.004	(2,27) 0.76	0.48	0.05	(2,28) 11.43	<0.001	0.45
LC	(1,17) 91.5	<0.001***	0.84	(1,29) 8.64	0.006**	0.23	(1,29) 2.04	0.16	0.07	(1,28) 21.32	<0.001***	0.43	(1,29) 4.0	0.055$^{(*)}$	0.12
G × T	(2,16) 7.98	0.004**	0.50	(2,28) 13.41	<0.001***	0.49	(2,28) 12.38	<0.001***	0.47	(2,27) 9.87	<0.001***	0.42	(2,28) 12.84	<0.001***	0.48
G × T × LC	(2,16) 0.60	0.56	0.07	(2,28) 1.65	0.211	0.11	(2,28) 1.58	0.22	0.10	(2,27) 1.43	0.26	0.10	(2,28) 1.89	0.17	0.12
G × LC	(1,17) 0.92	0.35	0.05	(1,29) 5.66	0.024*	0.16	(1,29) 4.98	0.033*	0.15	(1,28) 4.46	0.044*	0.14	(1,29) 4.6	0.041*	0.14
MAN(C)OVAs for the **ERP N1** (LOT/ ROT) and fMRI signal change R4 (left/ right) with between-subject factor *group* (GG-first, NC-first) and within-subject factors *test time* (T1, T2, T3), *condition* (W, FF), and *hemisphere*. Listed are interactions with test time															
ERP N1	F	*P*	η_p^2	F	*P*	η_p^2	F	*P*	η_p^2	F	*P*	η_p^2	F	*P*	η_p^2
(n = 32)															
G × T	(2,16) 1.14	0.34	0.13	(2,28) 1.23	0.31	0.08	(2,28) 1.48	0.24	0.10	(2,27) 2.76	0.081$^{(*)}$	0.17	(2,28) 0.92	0.41	0.06
T × C	(2,16) 1.54	0.25	0.16	(2,28) 2.34	0.12	0.14	(2,28) 9.64	<0.001***	0.41	(2,27) 2.13	0.14	0.14	(2,28) 2.77	0.080$^{(*)}$	0.17
G × T × C	(2,16) 7.56	0.005**	0.49	(2,28) 3.68	0.038*	0.21	(2,28) 7.72	0.002**	0.36	(2,27) 5.39	0.011*	0.29	(2,28) 3.2	0.056$^{(*)}$	0.19
G × T × C × H	(2,16) 4.36	0.031*	0.35	(2,28) 1.92	0.17	0.12	(2,28) 2.21	0.13	0.14	(2,27) 1.34	0.28	0.09	(2,28) 2.08	0.14	0.13
fMRI ROI 4	F	*P*	η_p^2	F	*P*	η_p^2	F	*P*	η_p^2	F	*P*	η_p^2	F	*P*	η_p^2
(n = 16)															
G × T	(2,6) 0.72	0.53	0.19	(2,12) 0.23	0.80	0.04	(2,12) 0.42	0.67	0.07	(2,12) 0.61	0.56	0.92	(2,12) 0.18	0.84	0.03
T × C	(2,6) 5.16	0.050$^{(*)}$	0.63	(2,12) 0.054	0.95	0.009	(2,12) 0.72	0.51	0.11	(2,12) 0.061	0.94	0.01	(2,12) 0.31	0.74	0.05
G × T × C	(2,6) 3.98	0.079$^{(*)}$	0.57	(2,12) 4.77	0.030*	0.44	(2,12) 4.45	0.036*	0.43	(2,12) 5.29	0.023*	0.47	(2,12) 2.89	0.095$^{(*)}$	0.33
G × T × C × H	(2,6) 1.31	0.34	0.30	(2,12) 0.98	0.40	0.14	(2,12) 0.44	0.66	0.07	(2,12) 0.56	0.59	0.09	(2,12) 0.063	0.94	0.01

Listed are interactions indicating training effects for all supporting analyses (top: behavioral analyses, bottom: ERP and fMRI analyses). The ROI analysis included the left and right R4. Analyses in column B summarize the results of good performers [ERP: n = 19 (8/11) fMRI: n = 9 (4/5)]. Analyses in columns C–F summarize the results when introducing different covariates (MANCOVAs). For analyses in column E, part ERP, only 31 children were included because the exact playing time for one child was unknown. C, condition (Words, False Fonts); G, group (GG-first or NC-first); H, hemisphere (left or right); LC, letter case (letter knowledge LK: uppercase, lowercase); T, test time (T1, T2, or T3).

$^{(*)}P < 0.1$; *, $P < 0.05$; **, $P < 0.01$; ***. $P < 0.001$.