# Far transfer to language and math of a short software-based gaming intervention

Andrea Paula Goldin[a,b,1], María Julia Hermida[c], Diego E. Shalom[a], Martín Elias Costa[a], Matías Lopez-Rosenfeld[a,d], María Soledad Segretin[c], Diego Fernández-Slezak[d], Sebastián J. Lipina[c,e], and Mariano Sigman[a,b]

[a]Laboratorio de Neurociencia Integrativa, Departamento de Física, Instituto de Física de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas, C1428EGA Buenos Aires, Argentina; [b]Universidad Torcuato Di Tella, C1428BIJ Buenos Aires, Argentina; [c]Unidad de Neurobiología Aplicada, Centro de Educación Médica e Investigaciones Clínicas, Consejo Nacional de Investigaciones Científicas y Técnicas, C1431FWO Buenos Aires, Argentina; [d]Laboratorio de Inteligencia Artificial Aplicada, Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, C1428EGA Buenos Aires, Argentina; and [e]Universidad Nacional de San Martín, 1650 Buenos Aires, Argentina

Edited* by Michael I. Posner, University of Oregon, Eugene, OR, and approved March 12, 2014 (received for review October 28, 2013)

**Executive functions (EF) in children can be trained, but it remains unknown whether training-related benefits elicit far transfer to real-life situations. Here, we investigate whether a set of computerized games might yield near and far transfer on an experimental and an active control group of low-SES otherwise typically developing 6-y-olds in a 3-mo pretest–training–posttest design that was ecologically deployed (at school). The intervention elicits transfer to some (but not all) facets of executive function. These changes cascade to real-world measures of school performance. The intervention equalizes academic outcomes across children who regularly attend school and those who do not because of social and familiar circumstances.**

cognitive training intervention | school grades | Attention Network Test | school attendance | working memory

NEUROSCIENCE

PSYCHOLOGICAL AND COGNITIVE SCIENCES

The efficacy of cognitive training is controversial and constitutes a current challenge for educational neuroscience research (1–4). Although it has been well documented that directed interventions in children can change specific cognitive functions (5–8), it is unknown whether those translate to broader contexts and real-world situations of educational pertinence. Cognitive training has largely focused on executive functions (EF) (6–8), a class of processes critical for purposeful, goal-directed behavior, including working memory (WM), planning, and cognitive control (6). Research has shown that EF capabilities can be improved with practice and gaming interventions (5–7, 9). These results are particularly promising because EF are critical for educational success (10–12) and for mental and physical health (5, 13); furthermore, early self-regulation is indicative of an individual's health and social behavior as an adult (14, 15).

Because the degree of self-regulation elicited by a child can predict real-life outcomes, it is presumed that an intervention that improves EF should affect a child educational success. However, this hypothesis has never been explicitly examined based on school grades as real-world measurements of educational achievement (16). Instead, current evidence (7, 9, 17, 18) derives from laboratory measures related to school performance (for instance, the time it takes for a child to read a word). Because school performance results from an intertwined process integrating EF with temperament, socioeconomic status (SES), and cognitive skills (19–22) among other environmental factors, examining the direct outcome of an intervention on school grades is necessary to assure its practical pertinence.

Our main hypothesis is that a gaming intervention in school-age children tuned to improve aspects of EF should transfer to real-world manifestations of school performance indexed by children's grades.

In the educational system of the City of Buenos Aires, first graders devote an important amount of their school time to language and math. Grading for these subjects is largely based on objective tasks and they are examined extensively. Instead, other subjects (such as foreign language or social behavior) where one may also expect a benefit of EF training, are devoted little school time for first graders and/or graded on much more arbitrary and subjective bases (23). Thus, our hypothesis is that a gaming intervention tuned to EF improvement may result in a specific effect in language and math because: (*i*) EF individual variability correlates with educational outcomes (16, 21, 24–26); and (*ii*) compared with other subjects for which one may also predict that EF training may have an effect (e.g., natural and social sciences), language and math are the ones with the most reliable grading system. Here, we set out to examine this hypothesis by conducting an intervention based on computerized games (27) that train EF in a low-SES group of children, initially displaying broad variability in school performance.

We first demonstrate the effect of such intervention on laboratory-based measures of EF and, subsequently, that these effects propagate to an improvement in school performance. This effect is specific to grades in language and mathematics and to the group of children who, because of a low rate of school attendance, have lower than average grades before the intervention.

## Results

**Intervention.** One hundred eleven low-SES typically developing children participated in the intervention over a period of 10 wk (see *SI Appendix*, Fig. S1 and Table S1 for specific information). Children were divided into two groups. Participants in the experimental trained group played three adaptive computer games aimed at training working memory, planning, and inhibitory control skills [refs. 27 and 28; www.matemarote.com.ar (Note

### Significance

Executive functions (EF) imply processes critical for purposeful, goal-directed behavior. In children, evidence derived from laboratory measures indicates that training can improve EF. However, this hypothesis has never been explicitly examined based on real-world measures, especially of educational achievement. Here, we investigate whether a set of computerized games might yield transfer on low-socioeconomic status otherwise typically developing 6-y-olds in an intervention deployed at their own school. The games elicit transfer of some EF, which cascades to real-world measures of school performance. More importantly, the intervention equalizes academic outcomes across children who regularly attend school and those who do not because of social and familiar circumstances.

---

Author contributions: A.P.G., M.J.H., M.E.C., M.S.S., D.F.-S., S.J.L., and M.S. designed research; A.P.G., M.J.H., and M.L-R. performed research; A.P.G., D.E.S., and M.S. analyzed data; A.P.G., S.J.L., and M.S. wrote the paper; and M.E.C., M.L.-R., and D.F.-S. programmed software.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: apgoldin@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1320217111/-/DCSupplemental.

that this platform is constantly being developed and updated so current games may not be identical to the ones used in this specific experiment.)]. Participants in an active control group were trained on three equally motivating games that required similar motor responses but which were less cognitive demanding.

Children played at school only one game in each 15-min session, and a total of no more than three sessions per week. The three games were alternated for all children throughout the intervention (*SI Appendix*, Fig. S1*B*). The mean number of sessions did not differ significantly between groups (*SI Appendix*, Fig. S1*C*; control group: 23.13 ± 0.56; trained group: 23.94 ± 0.40; $t_{109} = 1.12$; $P = 0.26$).

Before and after the intervention, we obtained school records of the children and measured their EF performance through several standardized tasks.

**Effects of the Intervention on Attention Performance.** The child Attention Network Test (ANT; ref. 29) is organized in three 32-trials blocks. In each trial, children have to rapidly catch an animal under different flanking conditions. The response times (RT) of children in correct trials showed a clear peak in the first trial of each block, which decreased rapidly and then remained stable (Fig. 1*A*). A block analysis revealed that nonstationary effects such as fatigue or learning are not a major factor (*SI Appendix*, Table S2). Hence, subsequent results are shown for the distribution of RT collapsed across all trials of the three blocks.

The intervention resulted in a sustained and large reduction of RT that was more pronounced in the trained group (Fig. 1*A*), revealed by a phase and phase–group interaction in a linear mix model (LMM) (*SI Appendix*, Table S3). Post hoc analyses showed that before the intervention, RT of both groups were similar (Pretest: $RT_{Control}$, 1,265.59 ± 9.29 ms; $RT_{Trained}$, 1,248.25 ± 7.11 ms; $t_{109} = 0.26$; $P = 0.80$), whereas after the intervention, RT were significantly shorter for the trained group (Posttest: $RT_{Control}$, 1,177.64 ± 8.42 ms; $RT_{Trained}$, 1,066.36 ± 6.59 ms; $t_{109} = 2.72$; $P < 0.008$). The fraction of correct responses showed a moderate increase for the trained group and a decrease for the control group (*SI Appendix*, Tables S4 and S5), showing that the reduction of RT in the trained group was not at the expense of an increase in the number of errors merely reflecting a change in the speed accuracy tradeoff.

A major advantage of child ANT is that it can discriminate between three different dimensions of attention: alerting, which relates to the capacity of maintaining a state of arousal; executive control, which relates to goal-directed behavior, and orienting, which refers to the capacity to shift the focus of attention (30). Following the ANT procedures (29), we measured performance in these three components of attention by comparing RT medians in different types of trials. This method was run for each individual independently in the pretest and posttest stages (Fig. 1 *B–G*). An emergent picture observed in all conditions is a decrease in RT after the intervention, which is significantly more pronounced in the trained group (Fig. 1 *B–D*). This pattern corroborates the main effects of phase and group–phase interaction shown in Fig. 1*A* and *SI Appendix*, Table S3 and indicates that this result is robust to all conditions. Above and beyond this main effect, we can investigate in which cases these changes show some specificity for certain conditions.

The alerting network scores derive from comparing no-cue and double-cue trials. The rationale for this comparison is that the double-cue provides a temporal signal that indicates the trial onset and triggers the alerting network. As expected, RT were slower for the no-cue condition (Fig. 1*B*). The effect of the intervention was very similar for both type of trials (Fig. 1 *B* and *E*) as confirmed by an LMM analysis, which showed that neither group (trained and control) nor phase (pretest and posttest) nor their interaction were significant for alerting scores (*SI Appendix*, Table S6).

The executive control scores derive from comparing noncongruent and congruent trials. As expected, RT were slower for the incongruent compared with the congruent condition (Fig. 1*C*). The effect of congruency is highly reliable and was observed in virtually every single child in our sample (Fig. 1*F*, *Lower*). RT for the control group showed a similar decrease in both conditions (Fig. 1*C*) and, hence, the effect of congruency remained similar before and after the intervention (Fig. 1*F*). In the trained group, instead, the decrease was more pronounced for the incongruent condition (Fig. 1*C*). As a consequence, the effect of conflict decreased from approximately 100 ms (in the pretest) to approximately 60 ms (in the posttest; Fig. 1*F*). The greater specificity of the intervention on the trained group for incongruent compared with congruent trials did not reach significance when calculated as an LMM analysis with group, phase, and their interaction as main regressors (*SI Appendix*, Table S6).

The orienting scores derive from comparing central- and spatial-cue trials. The rationale for this comparison is that the spatial-cue trials provide information that can trigger an exogenous allocation of attention to the position of the target. RT in the control group showed a large decrease in the spatial condition and only a modest decrease in the central condition. Instead, the decrease in RT for the trained group was comparable in both types of trials (Fig. 1*D*). This differential effect between group and type of trial was confirmed by an LMM that revealed a significant effect of phase and of the interaction between phase and group (*SI Appendix*, Table S6). Post hoc *t* tests showed that the phase effect was accounted for a difference only in posttests (control vs. trained groups: Pretest: $t_{109} = -1.50$, $P = 0.13$; Posttest: $t_{109} = 3.11$, $P < 0.003$) and suggested that the interaction effect is

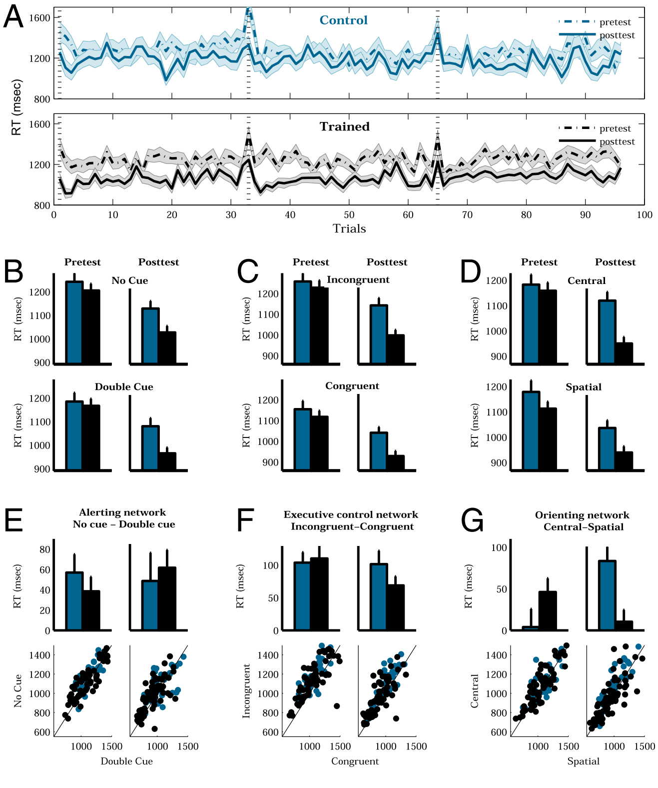

**Fig. 1.** (*A*) Response time to the child ANT task for both test phases (dotted line, pretest; full line, posttest) and experimental group (*Upper*, blue: control; *Lower*, black: trained). Lines indicate the mean response times (RT) for each trial. The dotted vertical lines indicate the first trial of each block. Shadowed areas indicate SEs. (*B–D*) Attentional network components of the child ANT task. Black, trained; blue, control. (*Left*) Pretest. (*Right*) Posttest. Bars indicate mean RT. Error bars indicate SEs. (*E–G*, *Upper*) Mean scores of each attentional network (substraction). Black, trained; blue, control. (*Left*) Pretest. (*Right*) Posttest. Bars indicate mean RT. Error bars indicate SEs. (*E–G*, *Lower*) Correlation between the two types of trial for each child (black dots, trained; blue dots, control).

built on an increase in posttest compared with pretest for the control group ($t_{74} = -2.83$; $P < 0.006$) (Fig. 1*G*).

**Effects of the Intervention on Inhibitory Control/Flexibility Performance.** The Heart-Flower Stroop (31) is a nonverbal task that measures aspects of inhibitory control and flexibility (1). In congruent trials, children have to press a button indicating the side of the screen in which a figure appears. In incongruent trials, they have to indicate the opposite side. Blocks are organized in fix (all trials of the same kind) or mix (congruent and incongruent trials are intermingled) assays.

In the fix blocks, RT significantly decreased in the posttest compared with the pretest in both groups (Fig. 2*A* and *SI Appendix*, Table S7), testified by a significant effect of RT in the phase regressor in an LMM (*SI Appendix*, Table S8). Current data cannot confirm whether this nonspecific effect is a consequence of the intervention (e.g., motor practice or other motivational nonspecific aspects of the intervention) or simply a consequence of development (29).

However, in the mix condition—which is more demanding in terms of cognitive flexibility—a significant effect was observed only in the group-by-phase interaction, both for congruent and incongruent trials (*SI Appendix*, Table S8). This interaction was accounted for by the fact that RT after the intervention were faster for children in the trained group (*SI Appendix*, Table S7). This difference was significant in the mix-congruent ($t_{105} = 1.69$; $P < 0.05$) and showed a trend in the mix-incongruent ($t_{104} = 1.59$; $P < 0.06$) conditions. Instead, the control group RT did not change between pretest and posttest. Another way of interpreting this result is that the effect of the intervention on RT was comparable for mix and fix blocks for the trained group but only observed in the fix blocks for the control group (Fig. 2).

When all of the conditions were grouped together, we did not observe any effect of group, phase, or their interaction with performance (*SI Appendix*, Table S9). When we analyzed performance for different conditions, we observed a main effect of phase for all conditions. The effect of phase was positive for the mix condition (reflecting an overall increase in performance for mix trials) and negative in the fix condition (reflecting an overall decrease in performance for both conditions) (*SI Appendix*, Tables S10 and S11). A likely explanation for this atypical finding is that after completing many playing sessions, the fix condition becomes boring, which seems particularly clear in the easiest and monotonous "fix congruent" condition where performance dropped from close to 98% (almost at ceiling) to approximately 90% (*SI Appendix*, Table S10). Also, we observed that in the mix congruent condition performance increased more for the control than for the trained group after the intervention. This unexpected effect may result from an atypical high performance of the trained group before the intervention (74% compared with 64% in the control group). After the intervention, performance of the trained group remained higher (76% compared with 74%), but because the starting level was much higher, the effect was smaller than for the control group. An LMM for this condition showed a significant effect of group and phase but not of the critical group–phase interaction.

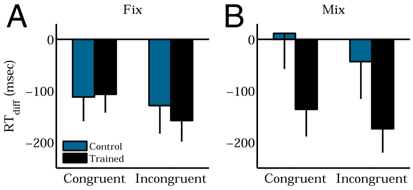

**Fig. 2.** Mean RT differences (posttest minus pretest) for fix (*A*) and mix (*B*) blocks for both experimental groups (blue, control; black, trained). Negative numbers indicate that RT in the postintervention session are faster than RT in the preintervention session. Error bars indicate SEs.

**Effects of the Intervention on Planning Performance.** We measured performance in the Tower of London (TOL), which is a widely used task for testing planning and aspects of problem solving (32). Both groups increased their performance between pretest and posttest (highest level achieved: control pretest 4.02 ± 0.18, trained pretest 4.14 ± 0.15, control posttest 4.82 ± 0.20, trained posttest 4.68 ± 0.15). An LMM analysis revealed that only phase (pretest and posttest) had a significant effect on TOL performance; neither group (trained or control) nor group-by-phase interaction (*SI Appendix*, Table S12) had a significant effect. TOL is the only task for which we did not observe a specific effect of the intervention.

**Effects of the Intervention on School: Math and Literacy Performance.** Our main objective was to investigate real-world effects of the intervention. To this aim, we analyzed class grades as a direct indicator of school performance. We divided the first-grade curriculum into three different groups. The first includes language and math (*LM*), two subjects that receive much teaching time in first grade in the City of Buenos Aires and the assessment of which is based on concrete and objective goals (and, hence, with better precision in grade assignment) (23). We thus hypothesized that grades of the children in these subjects could increase after the intervention. The second group (control, *C*) includes control subjects such as physical education, music, arts, and collaborative work, for which we did not expect any effect of the intervention. The third group (informal, *I*) includes subjects afforded little school time and which are graded informally (without clear curricular goals) in first graders such as natural and social sciences. Finding or not an effect of the intervention on the *I* group seemed equally possible to us.

Grading was assessed by the children's teachers, who were blind to all aspects of the experiment. Grades were given every 2 mo throughout the school year (at the end of each of the four bimesters). The intervention begun before the grading of the third bimester and finished before the grading of the fourth. Hence, preintervention versus postintervention effects are based on comparisons between the second and fourth bimesters.

Grades of children showed a strong relation to school attendance and a general increase from the second to the fourth bimester (Fig. 3 and *SI Appendix*, Table S13). This result was confirmed by an LMM with attendance and bimester as main factors, which showed a significant effect of bimester and the bimester–attendance interaction (*SI Appendix*, Table S14). For example, in language in the fourth bimester, low-attendance grades were 7.50 ± 0.25, whereas the high-attendance group had 8.41 ± 0.20. Given that attendance was a major factor in grade variability, we analyzed the effect of the intervention by dividing children along the median split of school attendance of each experimental group.

We first compared how performance varied with bimester (second and fourth) and with group (trained and control). This analysis was performed with an independent linear model for the two attendance groups (high and low) and for the three subject groups (*LM*, *I*, and *C*) (Fig. 4 *A* and *B*). The sole model showing a significant effect in the group–bimester interaction was the *LM* for the low-attendance group (*SI Appendix*, Table S15). This interaction was accounted for by the fact that *LM* grades in the fourth bimester were better for the trained compared with the control group (control: 7.03 ± 0.33; trained: 7.89 ± 0.23; $t_{85} = -2.21$; $P < 0.03$). Instead, before the intervention, there was no difference in LM grades between trained and control groups (control: 6.58 ± 0.27; trained: 6.74 ± 0.20; $t_{96} = -0.47$; $P = 0.64$). After the intervention, children in the trained group achieved similar *LM* grades regardless of their degree of attendance (Fig. 4 *A* and *B*; fourth bimester trained group school grades in mathematics: high-attendance: 7.80 ± 0.25, low-attendance: 7.87 ± 0.28, $t_{64} = 0.16$; $P = 0.87$; language: high-attendance: 8.38 ± 0.24, low-attendance: 7.92 ± 0.31, $t_{64} = -1.11$; $P = 0.27$).

We followed up the prepost analysis with a distribution analysis to identify the specificity of the effect to *LM* grades. To this

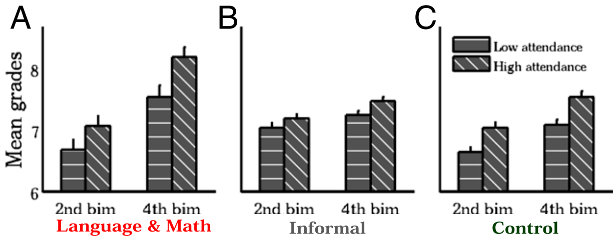

**Fig. 3.** Average class grades for preintervention (second bimester) and postintervention stages (fourth bimester) of the three groups of subjects divided by children school attendance. (*A*) Language and mathematics. (*B*) Informal subjects (e.g., history-social science). (*C*) Control subjects (e.g., technology arts). Error bars indicate SEs.

aim, we performed an analysis based in a total of 84 conditions, corresponding to 3 bimesters, 14 subjects, and 2 levels of attendance (high or low). The third bimester was excluded of this analysis because its grading partially overlapped with the intervention and, therefore, could not be clearly assigned to a pre phase or postphase of the intervention.

For each condition, we compared grades for the control and the trained groups and generated a $t$ values probability density curve for grades where each entry of the distribution is a $t$ test of one of the 84 comparisons (Fig. 4*C*). Comparisons corresponding to the first and second bimesters organized in a Gaussian distribution with close to zero mean (mean: −0.136, SD: 0.654), which is expected because it merely reflects the random assignation of each value revealing no differences before the intervention.

The hypothesis that the intervention results in a specific improvement to *LM* grades for the low-attendance children results in the following prediction: the two comparisons between trained and control group corresponding to language and math for the low-attendance groups should rank first among all 84 comparisons. Note that in this rank analysis, we are not performing multiple comparisons among all these 84 comparisons, but instead we are testing the hypothesis that the ones for which we had a hypothesis rank first among all.

In very tight agreement with this prediction, the language and mathematics values of the fourth bimester (after the intervention) for low-attending children were the two higher ranked comparisons (i.e., showing a greater difference between trained and control group) of all 84 comparisons including all subjects, bimesters, and groups (Fig. 4*C*). The probability of the null hypohesis that this order resulted by chance can be obtained from permutation analysis of the empirical distribution of scores for the first two bimesters and corresponds to $P < 0.0022$ in language and $P < 0.0160$ in mathematics. The joint probability of obtaining by chance a specific effect in language and math in the fourth bimester is $P < 0.00003$.

## Discussion

Before and after a relatively short software intervention (approximately 7 h grouping together the time of all sessions), we evaluated transfer to different facets of executive function, including networks of attention (29), inhibitory control (31), and planning (32), and to school performance (grades).

In the child ANT task, the trained group showed faster RT compared with the control group after the intervention, but not before. After the intervention, the RT decrease was 181 ms faster for the trained group and 88 ms faster for the control group. This 100-ms difference is large, comparable to the spontaneous developmental decrease achieved between 4 and 7 y of age, based on the average data from two independent studies with middle- and low-income samples (29, 33).

The ANT task can examine the effect of the intervention on the three networks of attention (executive control, orienting, and alerting) by comparing RT in different types of trials. We observed a differential effect of the intervention between the trained and control group in the orienting network. Children in the trained group showed equivalent (large) RT gains on all trials regardless of orienting cues. Instead, the gains of children in the control group were specific to trials with an exogenous orienting cue indexing the position of the target (spatial cue). In trials in which the exogenous cue did not indicate the location of the target (central cue), the gain in the control group after the intervention was modest. Such a result might shed light on the difference between control video games that encourage exogenous orienting, and the training activities that encourage endogenous orienting in the context of cognitive control.

Our hypothesis was that the trained group should have shown a benefit in EF. The ANT executive control scores showed only a nonsignificant trend revealing a specific effect of the intervention. The control group showed a comparable decrease in RT in congruent and incongruent trials and, hence, the effect of congruency remained similar before and after the intervention (approximately 100 ms). In the trained group, instead, the decrease was more pronounced for the challenging incongruent condition. As a consequence, the effect of conflict decreased from approximately 100 ms (in the pretest) to approximately 60 ms (in the posttest). Similarly, we observed a moderate increase in performance (*SI Appendix*, Table S4) in incongruent trials specific for the trained group.

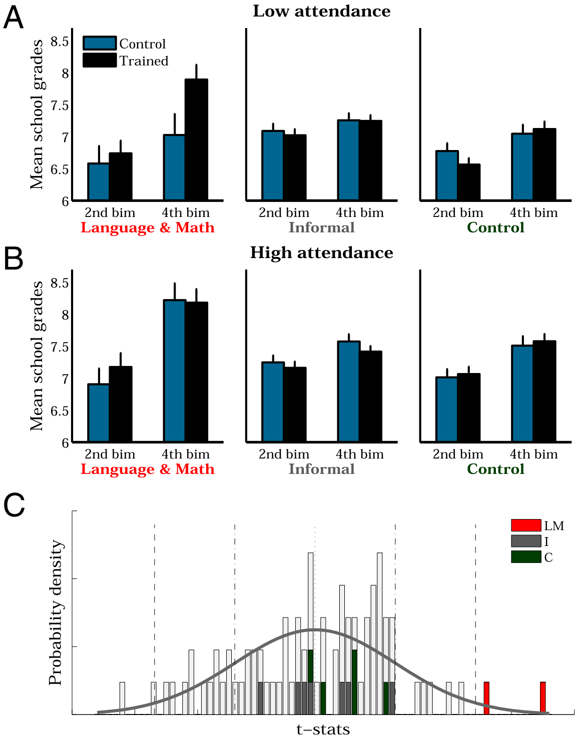

**Fig. 4.** (*A* and *B*) Average class grades divided by children school attendance: low (*A*) or high attendance (*B*). Grades are shown for the second (before the intervention) and the fourth bimester (after the intervention). Error bars indicate SEs. (*C*) Probability density for all 84 $t$ values trained-to-control comparisons (14 subjects, 2 groups for high and low school attendance, and 3 bimesters for first, second, and fourth). The dotted line indicates the mean and the dashed lines, the first and second SD of the distribution. $t$ values corresponding to comparisons of low-attending children after the intervention (fourth bimester) are colored according to the group of subjects (dark gray, informal subjects; green, control subjects; red, language and math). $t$ values for all subjects for first and second bimester and for the fourth bimester of the high attendance group are light gray. Note that language and math after the intervention (red) correspond to the higher ranked (to the foremost right) $t$ values of all comparisons.

The Stroop battery showed a reliable and large effect size on RT. However, we did not observe clear and reliable effects on performance. This result may seem intriguing because some researchers (31, 34) have shown that in similar tasks (although more focused to task switching), accuracy is more sensitive than RT in identifying EF. However, whether RT or performance is the most sensitive measure may vary widely with specific task contexts. In our experiment, performance varied in a narrow range (~10% variations across groups and phases for each condition; *SI Appendix*, Table S10) and RT showed relatively low variability (SEs ~50 ms in RT ~1,000 ms, *SI Appendix*, Table S7), which may explain why RT are more informative about EF. In addition, our children were older than in previous experiments where performance seemed to be the best indicator of EF (1).

The lack of transfer to TOL task was unexpected, because planning was specifically trained by one of the games. The most parsimonious explanation is that the planning game is highly domain specific, which means that the learning that children may obtain from playing is only a small part of the general resources of planning tested by TOL task (increasing depth of plan, inhibiting unsuccessful heuristics such as greediness). An additional explanation is that TOL was the only test that was not completed on the computer. Thus, the change in the contextual setting for our tasks might have hampered transfer.

In summary, a global analysis of our results regarding the effects of the intervention on EF shows (*i*) that there is a strong decrease in RT (~100 ms) in the trained compared with the control group in the child ANT task. (*ii*) When looking specifically at the ANT executive control network, we observe only trends of small effects in the scores in the expected direction (increase in performance and decrease in RT in the trained group), which do not reach significance. This result may be partly camouflaged by a strong and specific effect of the intervention in both congruent and incongruent trials. (*iii*) A highly significant, reliable, and large-effect size ~100 ms in the more demanding cognitive-flexibility condition of the Stroop task, which is not accompanied by significant changes in performance. (*iv*) A lack of an effect in transfer to TOL planning task. Therefore, the overall results from our cognitive batteries suggest that the training may, in effect, lead to an improvement of EF, but they also indicate that this increase is not expressed in all facets of EF or task contexts. These differential and specific effects fit well with current theories (35, 36) conceiving EF as both a convergent cognitive construct shared by multiple measures, and divergent factors of inhibition, switching, and updating. Although this idea continues to be controversial (37), understanding our results in terms of specificity of training to different factors contribute to the comprehension of the organization of EF in children.

**Real-World Changes: Equalizing Opportunities.** Our most important finding shows that the improvement in EF transfers to school outcomes based on the assessments of school performance by teachers.

Because the success of cognitive training has been controversial (1–4), it is important from a practical and theoretical perspective to understand why this specific intervention might have been successful even in measures of far transfer. Because of the complexity and high number of dimensions involved (number of combinations of games, difficulty of the games, total number, and duration of sessions), it is practically intractable to factorize all of the variables of this intervention. However, based on previous knowledge, we may speculate and pinpoint relevant aspects of the design that likely helped to make the intervention effective. (*i*) The intervention trained aspects of self-regulation, a skill that underlies early success in school (22) and is indicative of health and social behavior many years later (14, 15). (*ii*) The intervention intermingled various EF skills (working memory, categorization, planning, and inhibitory control) in different games. The relevance of the mixture for intervention success was demonstrated by Bunge and coworkers, showing that a variety of reasoning games in a classroom setting increased the gains in fluent reasoning compared with training in a specific task context (9). (*iii*) The intervention was based on games, whose rewarding nature is a motivating and engaging way of learning things (38). Moreover, playing has been shown to foster academic, cognitive, and social abilities in children (reviewed in ref. 39).

Our results specifically demonstrate that the intervention equalizes the academic outcome of children that go less than the median to school, positioning them at a similar level as children that attend more frequently. The implications of these findings are more pervasive for disadvantaged children because (*i*) for this group of children, lower marks translate into higher grade repetition and school dropout rates (ref. 20 as an example), and (*ii*) early performance in language and mathematics cascades to a broad number of social and educational factors (40).

## Materials and Methods

**Participants.** A total of 111 low-SES 6–7-y-old children (61 males) participated in the study. All participants were recruited from five first-grade classrooms in two public schools in the metropolitan area of Buenos Aires. Children's caregivers gave written consent to participate in the study, which was authorized by an institutional Ethical Committee (Centro de Educación Médica e Investigaciones Clínicas, Consejo Nacional de Investigaciones Científicas y Técnicas, protocol no. 486). Only one child was not authorized and did not participate.

**Sociodemographic Variables.** A socioeconomic scale (41) was administered to each parent to identify indicators of unsatisfied basic needs (UBN, poverty criteria) and other typical indicators of socioeconomic status (42). Applied UBN criteria were based on the identification of inappropriate dwelling (housing) and overcrowding (three or more persons per room excluding kitchen). In addition, scores were assigned directly to mothers for educational and occupational backgrounds. Analyses also included whether children receive state subsidy and whether they live in a slum. (see *SI Appendix*, Table S1 for further information and scoring criteria).

**Description of the Games.** Games are part of a growing free software, available for educational and research purposes, www.matemarote.com.ar. All games progress by using an algorithm that continuously adapts the difficulty level based on participant's performance. The adaptive nature of the game is determined by a structure of predefined levels (for instance, a small number of items in the working memory game constitutes a lower level than one with a higher number of items). All of the programs have interfaces that give feedback for correct performance. The graphics were depicted by image designers to make the aesthetics of the game enjoyable by children. The graphics design process had several iterations informed by feedback from children in the age range of 4–8 y old. Brief descriptions of the games follow; see refs. 27 and 28 and *SI Appendix* for further details and information.

***Working memory game.*** It is based on a nonspatial, pattern recognition working memory task (43). Each trial consists of a constant number of cards that appear randomly located in a grid (*SI Appendix*, Fig. S1*A*, *Left*). Children have to sequentially choose all of the cards (which reappear shuffled 200 ms after each selection) without repetition.

***Planning game.*** It is based on the Dog-Cat-Mouse puzzle designed by Klahr (44) and consists of three characters (a boy, a girl, and a cat) that own three places ("homes") (*SI Appendix*, Fig. S1*A*, *Center*). The goal of a trial is to move every character to its corresponding place (27, 44).

***Inhibitory control game.*** Children have to indicate as rapidly as they can the direction to which a plane points, ignoring irrelevant cues.

**Experimental Design.** First graders were balanced assigned (matching the groups with regard to sex and classroom) to trained ($n$ = 73, 40 males) or control ($n$ = 38, 21 males) groups. The pseudorandom assignment controled that in every classroom approximately one-third of girls and one-third of boys would be randomly assigned to the control group and, similarly, the remaining two-thirds to the trained group. All children played computer games in experimental sessions that lasted approximately 15 min. Children played only one game in each session and performed at least three sessions per week. Children played three cycles, each one consisting of three sessions playing one of three games and then changed the game, after having played three sessions of each of the three games the cycle restarted (see timeline in *SI Appendix*, Fig. S1*B*). Because the experiment was deployed inside the school, a child only played a session if he or she was at school that day. Hence, after 10 wk the intervention finished, whether children completed their 27 sessions or not (*SI Appendix*, Fig. S1*C*). Children in the control group played three cycles of three sessions of three different commercially available computer games (see *SI Appendix* for further details).

**Training and Testing Procedures.** Every child who played was accompanied by an adult that was there to explain the rules (the first time) or remind them (whenever necessary) and to support the child if needed. All research assistants (RA) gave the same instructions every time they explained the rules. All children understood all rules after less than three trials in all games. All RA were blind to the experimental design. All of the training and testing procedures were assessed by the RA inside the school, in appropriate rooms for these purposes. To maximize the concentration of children, all instances were performed with headphones and on individual computers.

**Grading.** First graders were evaluated in 14 different subjects by at least seven different teachers and school grades were given every 2 mo during the whole school year (four bimesters, see *SI Appendix*, Table S13 for the experiment's year school grades). All teachers were blind to the experimental design, and they did not even know there was a control group. The intervention took place during the last half of the third and more than the first half of the fourth bimester.

For school-grades analyses, we divided each experimental group by its own median of school attendance (low and high attendance) based on our records. Attendance could be calculated from school records, or from our own. As expected, both measures of attendance correlated (control group: $R = 0.46$; $P < 0.026$; trained group: $R = 0.44$; $P < 0.004$). The school record was full of missing values, because teachers often do not provide a record of attendance. We therefore chose to calculate the attendance for each child based on our records (*SI Appendix*, Fig. S1*C*).

School subjects were classified in three groups: mathematics and language (*LM*); rights and responsibilities of citizenship, music, physical education, visual arts, social behavior, technology arts, and collaborative work (*C*); and foreign language, natural sciences, history-social sciences, and two different responsibility measures (*I*).

**Statistical Analysis.** See *SI Appendix*.

**ACKNOWLEDGMENTS.** We thank Varinia Telleria for the design of characters and drawings, Milena Winograd for early work on Mate Marote, Nubis for programming, Lucía Prats and all research assistants for helping with data collection, Paola Corrales for data entry help, Escuela N.° 6 D. E. 1 "French y Beruti" and Escuela N.° 1 D. E. 1 "Juan José Castelli," Gabriel Gellon for reading and commenting the manuscript, and an anonymous reviewer for pointing us to the interpretation of the exogenous-endogenous orienting result. We specially thank Kathy Hirsh-Pasek for many helpful suggestions and critical comments. A.P.G. specially thanks Jorge Medina for all the support and confidence. This research was supported by Consejo Nacional de Investigaciones Científicas y Técnicas, Centro de Educación Médica e Investigaciones Clínicas, Ministry of Science of Argentina, Human Frontiers, and Fundación Conectar. M.S. is sponsored by the James McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition—Scholar Award.

1. Diamond A, Barnett WS, Thomas J, Munro S (2007) Preschool program improves cognitive control. *Science* 318(5855):1387–1388.
2. Owen AM, et al. (2010) Putting brain training to the test. *Nature* 465(7299):775–778.
3. Shipstead Z, Redick TS, Engle RW (2012) Is working memory training effective? *Psychol Bull* 138(4):628–654.
4. Jolles DD, Crone EA (2012) Training the developing brain: A neurocognitive perspective. *Front Hum Neurosci* 6:76.
5. Diamond A (2012) Activities and programs that improve children's executive functions. *Curr Dir Psychol Sci* 21(5):335–341.
6. Diamond A, Lee K (2011) Interventions shown to aid executive function development in children 4 to 12 years old. *Science* 333(6045):959–964.
7. Thorell LB, Lindqvist S, Bergman Nutley S, Bohlin G, Klingberg T (2009) Training and transfer effects of executive functions in preschool children. *Dev Sci* 12(1):106–113.
8. Bryck RL, Fisher PA (2012) Training the brain: Practical applications of neural plasticity from the intersection of cognitive neuroscience, developmental psychology, and prevention science. *Am Psychol* 67(2):87–100.
9. Mackey AP, Hill SS, Stone SI, Bunge SA (2011) Differential effects of reasoning and speed training in children. *Dev Sci* 14(3):582–590.
10. Best JR, Miller PH, Jones LL (2009) Executive functions after age 5: Changes and correlates. *Dev Rev* 29(3):180–200.
11. Gathercole SE, Pickering SJ, Knight C, Stegmann ZOE (2004) Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Appl Cogn Psychol* 18(1):1–16.
12. Bull R, Espy KA, Wiebe SA (2008) Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Dev Neuropsychol* 33(3):205–228.
13. Simonds J, Kieras JE, Rueda MR, Rothbart MK (2007) Effortful control, executive attention, and emotional regulation in 7–10-year-old children. *Cogn Dev* 22(4):474–488.
14. Moffitt TE, et al. (2011) A gradient of childhood self-control predicts health, wealth, and public safety. *Proc Natl Acad Sci USA* 108(7):2693–2698.
15. Mischel W, et al. (2011) 'Willpower' over the life span: Decomposing self-regulation. *Soc Cogn Affect Neurosci* 6(2):252–256.
16. Titz C, Karbach J (2014) Working memory and executive functions: Effects of training on academic achievement. *Psychol Res*, 10.1007/s00426-013-0537-1.
17. Barnett SM, Ceci SJ (2002) When and where do we apply what we learn? A taxonomy for far transfer. *Psychol Bull* 128(4):612–637.
18. Loosli SV, Buschkuehl M, Perrig WJ, Jaeggi SM (2012) Working memory training improves reading processes in typically developing children. *Child Neuropsychol* 18(1):62–78.
19. Rueda MR, Posner MI, Rothbart MK (2005) The development of executive attention: Contributions to the emergence of self-regulation. *Dev Neuropsychol* 28(2):573–594.
20. Brooks-Gunn J, Duncan GJ (1997) The effects of poverty on children. *Future Child* 7(2): 55–71.
21. Blair C, Razza RP (2007) Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev* 78(2): 647–663.
22. Stevens C, Bavelier D (2012) The role of selective attention on academic foundations: A cognitive neuroscience perspective. *Dev Cogn Neurosci* 2(Suppl 1):S30–S48.
23. Mendoza S (dir.) (2004) *Diseño curricular para la escuela primaria: Primer ciclo de la escuela primaria-educación general básica* [Curriculum design for primary school: Primary school first stage - Basic general education] (Secretaría de Educación del Gobierno de la Ciudad de Buenos Aires. Dirección General de Planeamiento. Dirección de Currícula). Spanish.
24. Best JR, Miller PH, Naglieri JA (2011) Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learn Individ Differ* 21(4):327–336.
25. Cybele Raver C, McCoy DC, Lowenstein AE, Pess R (2013) Predicting individual differences in low-income children's executive control from early to middle childhood. *Dev Sci* 16(3):394–408.
26. Hillman CH, et al. (2012) From ERPs to academics. *Dev Cogn Neurosci* 2(Suppl 1): S90–S98.
27. Goldin AP, et al. (2013) Training planning and working memory in third graders. *Mind, Brain. Educ* 7(2):136–146.
28. Lopez-Rosenfeld M, Goldin AP, Lipina SJ, Sigman M, Fernandez Slezak D (2013) Mate Marote: A flexible automated framework for large-scale educational interventions. *Comput Educ* 68:307–313.
29. Rueda MR, et al. (2004) Development of attentional networks in childhood. *Neuropsychologia* 42(8):1029–1040.
30. Posner MI, Petersen SE (1990) The attention system of the human brain. *Annu Rev Neurosci* 13:25–42.
31. Davidson MC, Amso D, Anderson LC, Diamond A (2006) Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia* 44(11):2037–2078.
32. Shallice T (1982) Specific impairments of planning. *Philos Trans R Soc Lond B Biol Sci* 298(1089):199–209.
33. Mezzacappa E (2004) Alerting, orienting, and executive attention: Developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child Dev* 75(5):1373–1386.
34. Diamond A, Kirkham N (2005) Not quite as grown-up as we like to think: Parallels between cognition in childhood and adulthood. *Psychol Sci* 16(4):291–297.
35. Miyake A, et al. (2000) The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognit Psychol* 41(1):49–100.
36. Lehto JE, Juujarvi P, Kooistra L, Pulkkinen L (2003) Dimensions of executive functioning: Evidence from children. *Br J Dev Psychol* 21(1):59–80.
37. van der Ven SHG, Kroesbergen EH, Boom J, Leseman PPM (2013) The structure of executive functions in children: A closer examination of inhibition, shifting, and updating. *Br J Dev Psychol* 31(Pt 1):70–87.
38. Howard-Jones P, Demetriou S, Bogacz R, Yoo JH, Leonards U (2011) Toward a science of learning games. *Mind, Brain. Educ* 5(1):33–41.
39. Fisher K, Hirsh-pasek K, Golinkoff RM, Singer DG, Berk L (2011) *The Oxford Handbook of the Development of Play*, ed Pellegrini A (Oxford Univ Press, New York), pp 341–362.
40. Alexander K, Entwisle D, Horsey C (1997) From first grade forward: Early foundations of high school dropout. *Sociol Educ* 70(2):87–107.
41. Lipina SJ, Martelli MI, Vuelta B, Colombo JA (2005) Performance on the A-not-B task of Argentinean infants from unsatisfied and satisfied basic needs homes. *Rev Interam Psicol J Psychol* 39(1):49–60.
42. Boltvinik J (1998) Poverty measurement methods - an overview. Available at http://web.archive.org/web/20060521215431/www.undp.org/poverty/publications/pov_red/Poverty_Measurement_Methods.pdf. Accessed March 25, 2014.
43. Petrides M, Milner B (1982) Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia* 20(3):249–262.
44. Klahr D (1985) Solving problems with ambiguous subgoal ordering: Preschoolers' performance. *Child Dev* 56:940–952.