Supplemental material.

Far transfer to Language and Math of a short software-based gaming intervention

Andrea P. Goldin, M. Julia Hermida, Diego E. Shalom, Martín Elias Costa, Matías López-Rosenfeld, M. Soledad Segretin, Diego Fernández-Slezak, Sebastián J. Lipina, Mariano Sigman

Working Memory Game. This game is based on a non-spatial, pattern recognition working memory task, a paradigm that measures recognition memory for visual patterns, but not spatial locations (1, 2). Each trial consists of a constant number of items that appear randomly located in a grid (Fig. S1a-left). 2000 msec after selecting one of the items with a mouse click in its specific location, a new screen appears with the same list of items but in a newly randomized spatial distribution. Now the child has to select an item different from the one he chose before and the process continues until the child has chosen, without repetition, all the items of the list (correct trial) or when s/he makes a repetition (error). The game starts with a predefined low number of items (three for this study) and this number can increase or decrease depending on the child performance (i.e. after a number of consecutive correct 3-items trials, the following trials have four items to be remembered, and the game can continue up to 20 items). If, on the other hand, the child makes errors in a number of consecutive trials, in the following trial one less item will have to be remembered). Note that for small number of items children have to remember all the chosen/not yet chosen cards (to avoid repetition). When the number of items exceeds the working memory limit, kids have to adopt a chunking or ordering strategy to solve the problem.

Each item is an image (a card) defined by a list of features that can be different or not. For instance, Fig. S1a-left shows the screen of a six-item trial where all the cards have the same background and shape and an umbrella that differs in color. All cards have different number of stars. Half of the items have the girl character sit on a chair. This variable is binary and can potentially be used by the child to chunk and categorize the items.

Planning Game. This game is based on the Dog-Cat-Mouse puzzle designed by Klahr (3) and consists of three characters (a boy, a girl and a cat) that own three places ("homes"). The characters and the places are arranged on a square board, which has four paths ("bridges"), each one parallel to each side of the board, and one diagonal path between the upper left and lower right corners of the board (Fig. S1a-middle). The characters can only be moved along the paths and one at a time. The goal of a trial is to move every character to its corresponding place. To move a character, a child clicks on it and drags and drops it to the new position. Each trial can be characterized in terms of its path length (i.e. how many moves have to be done to attain the trial): the total possible problems vary from one to seven moves (3, 4).

For the present study, minimum-move trials imply two movements and after three consecutive correct trials the number of movements increases by one. For instance, Fig. S1a-middle shows the screen of one 4-moves trial. The three characters (the boy on the lower right corner, the girl on the upper right corner, and the cat on the upper left corner) have to reach their homes (boy's upper left, girl's lower right, and cat's upper right; the lower left corner is an empty place to temporarily occupy).

Training sessions were organized in the following two phases:

Free-exploration: In this first phase a trial is considered correct when the child moves all the characters to their corresponding houses regardless of the number of moves s/he makes to attain the goal. This phase finishes when the child completes all the sequence of trials (three consecutive trials in which the number of moves between the initial configuration and the goal is, respectively 2, 3, 4, 5, 6, and 7).

Restricted-movements: Consists of three consecutive stages in which a trial is considered correct only if the goal is attained in the minimal number of moves. In the first stage, the child is told how many moves are enough to solve each trial and the number of moves between the initial configuration and the goal is, respectively 2, 3, 4, 5, 6, and 7. In the second stage, the child is told how many moves are enough to solve each trial but in each trial the number of moves between the initial configuration and the goal is not consecutive but at random. After six correct three-trials blocks, this stage ends and is followed by the third stage, which is equal to the second one except that the child is not told how many moves are the minimum required to solve each trial. In this experiment, this stage had no end: it looped until the training was finished.

Inhibitory control Game. At this game, the child has two direction choices (left or right) and has to decide, as soon as s/he can, which is the correct one. In different positions of the screen a paper plane appears and the child has to choose and press one of two keyboard keys (L for right and A for left). Planes can be red or yellow and can be pointing to the right or to the left. If a plane is yellow, the task is to press the direction to which the plane is pointing at. If, on the other hand, the plane is red, the task is to press the opposite direction. This is the only of the three training games that requests to hurry up. In the center of the screen a running clock appears and if a key is not pressed before it ends, the trial is lost. As more consecutive trials are succeeded, difficulty increases by speeding up the clock.

Experimental Design. The original idea of this study was not only to test the transfer of Mate Marote's games but also to study how children learned to play them and what kind of strategies did they use to solve the different levels. Hence, we needed more children to play Mate Marote's games and first graders were randomly assigned (matching the groups with regard to gender and classroom) to trained (*n*=73, 40 males) or control (*n*=38, 21 males) groups.

The experiment ideally consisted of a total of 27 non-consecutive sessions of training on three different computer games. All children played computer games in experimental sessions which lasted about 15 minutes. 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 Fig. S1b). As the experiment was deployed inside the school, a child only played a session if s/he were at school that day. Hence, after 10 weeks the intervention finished, whether or not children completed their 27 sessions (Fig. S1c).

The order of played games for the intervention group was: 1) planning, 2) working memory, and 3) inhibitory control (Fig. S1b). Children in the control group played three cycles of three consecutive sessions of three different commercially available computer games which were selected based on their low impact on Mate Marote's trained cognitive demands but with comparable engaging and motor action (Control game 1: One character had to jump using vines, space bar was used. Control game 2: It was a rally race, not restricted by time, in which decisions between two possible paths had to be make, mouse usage was necessary. Control game 3: A ball had to jump between rocks in order not to fall into a cliff, keyboard arrows were needed).

Training and Testing Procedures. In all the experimental and training sessions, every child played accompanied by a research assistant (RA) that was there to explain the rules (the first time) or remind them (whenever necessary) and to support the child if needed. For instance, some children needed a RA to tell them that they were playing well and that it was part of the game if they lose. In these cases, support was given for every trial, independently to the children request. All RA gave the same instructions every time they explained the rules. All children understood all rules after less than three trials in all games. Importantly, none of the Research Assistants (RA) nor the teachers were informed about the train-control structure of the experiment. Moreover, none of the RA or the teachers knew that we would analyze the grades during the course of the experiment. Both teachers and RA knew that children would be playing games, they were informed about the length of the experiment, and the fact that we were investigating whether games could serve to enhance some cognitive aspects. We were extremely careful not to give any hint that might indicate that we had an expectation that one set of games might be more efficient than the other. Specifically, to assure that all adults remained comparable regardless of groups, RA were trained by the researchers for two weeks prior to the pretest phase. The training consisted in teaching RA: a) how to properly administer the three tests; b) standardized instructions they should provide children to play each game. We were careful to assure that the length of the instructions were comparable in both groups; c) how to (and when) assist children while playing if they showed difficulties or if they needed encouragement; d) how to manage logistical and administrative issues; and e) how to behave at school and with the children (including the previous talking with a child, how to talk to teachers, how gestures are important and when to consciously control them, etc.).

As emphasized, RA were completely blind to our hypothesis and the existence of a Control and a Trained group. During the RA training we emphasized that we expected that all games could promote some aspects of cognition without providing any further detail. Also importantly, each RA assisted children in each session from different experimental groups in a random assignment. In other words, there were no RA assigned to each group or to specific group of children. RA were 10 in total (6 in "Number 1 school" and 4 in the smaller "Number 2 school").

During the training and the testing sessions, children were always invited to play and never knew they were been evaluated. RA had to try that children enjoy the experience the most. Children were told that if they did not want to stay in the experimental room they could go back to the classroom as soon as they needed. Only once one child left the experimental session (because he felt ill) and all children had fun while playing and were eager to play every day they saw the experimenters at the school.

All the training and testing procedures were assessed by the RA inside the school, in appropriate rooms for these purposes. To maximize children's concentration, all instances were performed with headphones and in individual computers.

A training or testing session went as follow: Each RA first prepared the computer, mouse, chair, headset, etc. One RA accompanied one child from the classroom to the computer room. S/he checked that the child was comfortable and started the session by logging in the child's ID. This automatically presented the specific software for that children (game, level, session number...) according to the progression of the training. If the child played a game for the first time, the RA explained the instructions in detail until s/he thought that the child had understood them in full detail. In the remaining training sessions they verified (asking questions and monitoring performance) whether the child remembered how to play the game. Testing sessions proceeded as guided and indicated by each specific test (see references (5–7).

The RA remained next to the child during play time. Games ended in an automatic manner after a fixed and specified time interval. Testing sessions also finished automatically after a completion criterion (i.e. number of trials, number of failures...) The RA then accompanied the child back to the classroom.

Measures. To evaluate the impact and transference of the trained skills, two-to-one weeks before the beginning of the training and one-to-two weeks after the last playing session, all children were administered a battery of standard tests. Tests included the children version of attentional test *ANT* (Attention Network Test (6), a cognitive flexibility Stroop task (5), and the Tower of London planning test (*TOL*, (7)). There were no significant differences between groups in any transfer measures at the pretest stage.

The Attention Network Test for children (Child ANT) developed by Rueda et al. (6) assesses three dimensions of attention, namely: alerting, orienting, and executive attention. The task is presented as a computer game and requires the children to: a) determine if an animal in the center of the screen points to the left or to the right, and b) act as soon as possible by pressing one of two buttons indicating the chosen direction.

All procedures and design was implemented as in Rueda's work (6). Equal numbers of flanker animals appeared at random on either side of the target animal on all trials oriented either in the same direction (congruent), or in the opposite direction (incongruent) as the target animal. Children were instructed to focus and respond only to the orientation of the central, target animal. Cues in the form of white laid eggs appeared for 150 msec at random before the target on 75% of trials either at the central fixation point ("central"), simultaneously above and below the central fixation point ("double"), or only above or below ("spatial"). Cues were not mentioned in the instructions given to children. Trials were counterbalanced by conflict situation (congruent/incongruent number of trials) and by the different types of cues and correct response side.

The task used at this experiment was composed of 104 trials divided into four blocks. The first block of 8 trials was a practice block. The subsequent three blocks consisted of 32 trials each, for a total of 96 test trials, and were only played if children had successfully completed at least 4 trials of the practice block. All children completed all 104 trials. The three 32-trials blocks were characterized by different animal and background color and children were allowed to rest between blocks. The interstimulus interval (ISI) was 1000 msec. Less than 0.1% of responses were given before 250 msec from the start of the trial and were dismissed in the present analyses, due to its impulsive condition (less than 0.078% of all trials during pretest and 0.095% during posttest). The whole test lasted about 20 minutes.

The Heart-Flower Stroop task was developed by Davidson et al. (5) to evaluate inhibitory control and cognitive flexibility since childhood. Briefly, the task consists of pressing one of two buttons depending on a figure that appears on the screen and its position. The stimuli figure can be a red heart or a red flower, and its position can be on the right or on the left of the screen. If the figure is the heart, the button to be pressed is the one on the same side of the figure and this trial is called “congruent” (i.e. if the heart appears on the right, the right button is the correct one to be pressed). If the figure is the flower, the button to be pressed should be the one on the opposite side and the trial is called “incongruent” (i.e. if the flower appears on the right, the left button is the correct choice).

The complete task used at this experiment consisted of three phases through which conditions changed progressively augmenting the difficulty on inhibitory control demands. The last phase additionally included a cognitive flexibility component. A first 12-trials congruent-only phase requiring short term memory (STM) was followed by a 12-trials incongruent-only phase requiring STM plus inhibition of the tendency to respond on the side where the stimulus appeared. In the third and last phase, 24 incongruent and congruent trials were intermixed, requiring STM, inhibiting attention, and flexibly switching attentional focus (congruent and incongruent trials were counterbalanced, 12 of each kind, and randomly presented). The whole task was comprised of counterbalanced left-right trials. The ISI was 500 msec. Responses given before 250 ms from the start of the trial were dismissed in the present analyses due to its impulsive condition (Table S16).

The Tower of London (TOL) task is widely used for measuring planning and aspects of problem solving (7). This test presents the participant with three colored balls that need to be moved between three different-size pegs in order to reproduce a target pattern in a set number of moves. A move involves moving a ball from one peg to another (or return it to the same) peg. Participants know the minimum number of moves they should make in order to succeed; however, a trial ends only when: a) the subject considers to have succeeded, or b) the subject asks to end it. A correct trial is the one whose final configuration is accomplished in the minimum number of moves needed. The participant must create a strategy for every trial (e.g., the particular sequence of moves) in order to succeed. Task difficulty is manipulated by increasing the number of moves required to reach the solution, which ranges from 1 up to 8 moves. For the task used at this experiment, the number of moves was maintained for five consecutive trials, after which the number of minimal moves increased by one. The task ended when the participant performed three consecutive incorrect trials.

Teachers give grades on a qualitative scale (in words), hence we converted them to grade points for analyses. As grades differ among countries and educational systems, we based our conversion on a mean calculated from Wikipedia (8). Conversion was calculated as follows: Sobresaliente (Excellent) = 10; Muy bueno (Very good) = 8.5; Bueno (Good) = 6.5; Regular (Sufficient) = 4.5; and Insuficiente (Insufficient) = 2.

Sociodemographic variables. Parents were asked to participate in a short interview about their children home environment. The majority of parents (specially mothers) agreed to participate. An individual appointment was made, and the conversation took place in a proper room at school. A socioeconomic scale (NES) (9) was administered to each parent to identify indicators of unsatisfied basic needs (UBN, poverty criteria) and other typical indicators of socioeconomic status (10). Among different questions, they were asked about their own education and occupation and about different aspects of their home. Different scores were built on this data (see next sections). Parents were also asked if their children receive state subsidy (0 = no, 1 = yes), and if their neighborhood is considered a slum (0 = no, 1 = yes).

Scoring of dwelling conditions. Scores to different items were summed (SS) (house type = 0-2; floor = 0-3; water availability = 0-3; bathroom = 0-3; ceiling = 0-3; external walls = 0-3; and home property = 0-3). A final score (FS) was assigned to this construct (i.e., dwelling) as per the following criteria: for SS between 0 and 5 the FS was 3; for SS between 6 and 11 the FS was 6; for SS between 12 and 17 the FS was 9; for SS between 18 and 20 the FS was 12. UBN households may tend to obtain the lowest scores on this indicator.

Scoring of overcrowding conditions. Overcrowding rates (R: persons per room excluding the kitchen) were calculated, thereafter a score (S) was assigned according to the following criteria: for R between 1 and 2 the S was 9; for R between 2.01 and 4 the S was 6; for R between 4.01 and 6 the S was 3; and for R higher than 6.1, the S was 0. UBN households may tend to obtain the lowest scores on this indicator.

Scoring of parental education and occupation backgrounds. The applied scoring criteria correspond to that used by the National Institute for Statistics of Argentina (INDEC). A higher score suggests better preparation. <u>Parental education</u>: without studies = 0; incomplete primary school = 1; complete primary school = 3; incomplete high school = 6; complete high school = 9; incomplete college = 10; complete college and more = 12. <u>Parental occupation</u>: unoccupied = 0; unstable worker = 1; unskilled laborer = 2; skilled laborer = 4; small independent producer = 6; administrative employee/sellers = 7; technician/nurse = 8; small business owner = 10; professional = 11; company director = 12.

Statistical analysis. We used linear mixed model (LMM) to test statistics for group comparisons. We used the lmer program of the lme4 package (version 1.0.5) (11) for estimating fixed and random coefficients. This package is supplied in the R system for statistical computing (version 3.0.2; R Core Team, 2013). We report regression coefficients (Ms), standard errors (SEs) and t- values (t = M/SE). There is no clear definition of ‘‘degree of freedom’’ for LMMs, and, therefore, precise p-values cannot be estimated. In general, however, given the large number of trials and subjects and the comparatively small number of fixed and random effects estimated, the t distribution is equivalent to the normal distribution for all practical purposes (i.e., the contribution of the degrees of freedom to the test statistic is negligible). Our criterion for referring to an effect as significant is $t = M/SE > 2.0$. The significance on fixed effects was checked with the *confint()* function implemented in the lme4 package. This function estimates confidence intervals by computing a likelihood profile and finding the appropriate cutoffs based on the likelihood ratio test. In all tests both statistics led to the same decision (n = 93). In all cases, the factors subject and classroom were considered as random.

SOM – References

1. Luciana M, Nelson CA (1998) The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia* 36:273–93.
2. Petrides M, Milner B (1982) Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia* 20:249–262.
3. Klahr D (1985) Solving problems with ambiguous subgoal ordering: preschoolers’ performance. *Child Dev* 56:940–952.
4. Goldin AP et al. (2013) Training Planning And Working Memory In Third Graders. *Mind, Brain, Educ* 7:136–146.
5. 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* 44:2037–2078.
6. Rueda MR et al. (2004) Development of attentional networks in childhood. *Neuropsychologia* 42:1029–1040.
7. Shallice T (1982) Specific impairments of planning. *Philos Trans R Soc London* 298:199–209.
8. http://en.wikipedia.org/wiki/Grading_(28education)29.
9. 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* 39:49–60.
10. Boltvinik J (1998) in *United Nations Development Programme, Social Development & Poverty Elimination Division (SEPED) Series on Poverty Reduction No. 3* (UNDP, New York).
11. Bates DM (2010) *lme4: Mixed-effects modeling with R* (Springer, New York, NY).

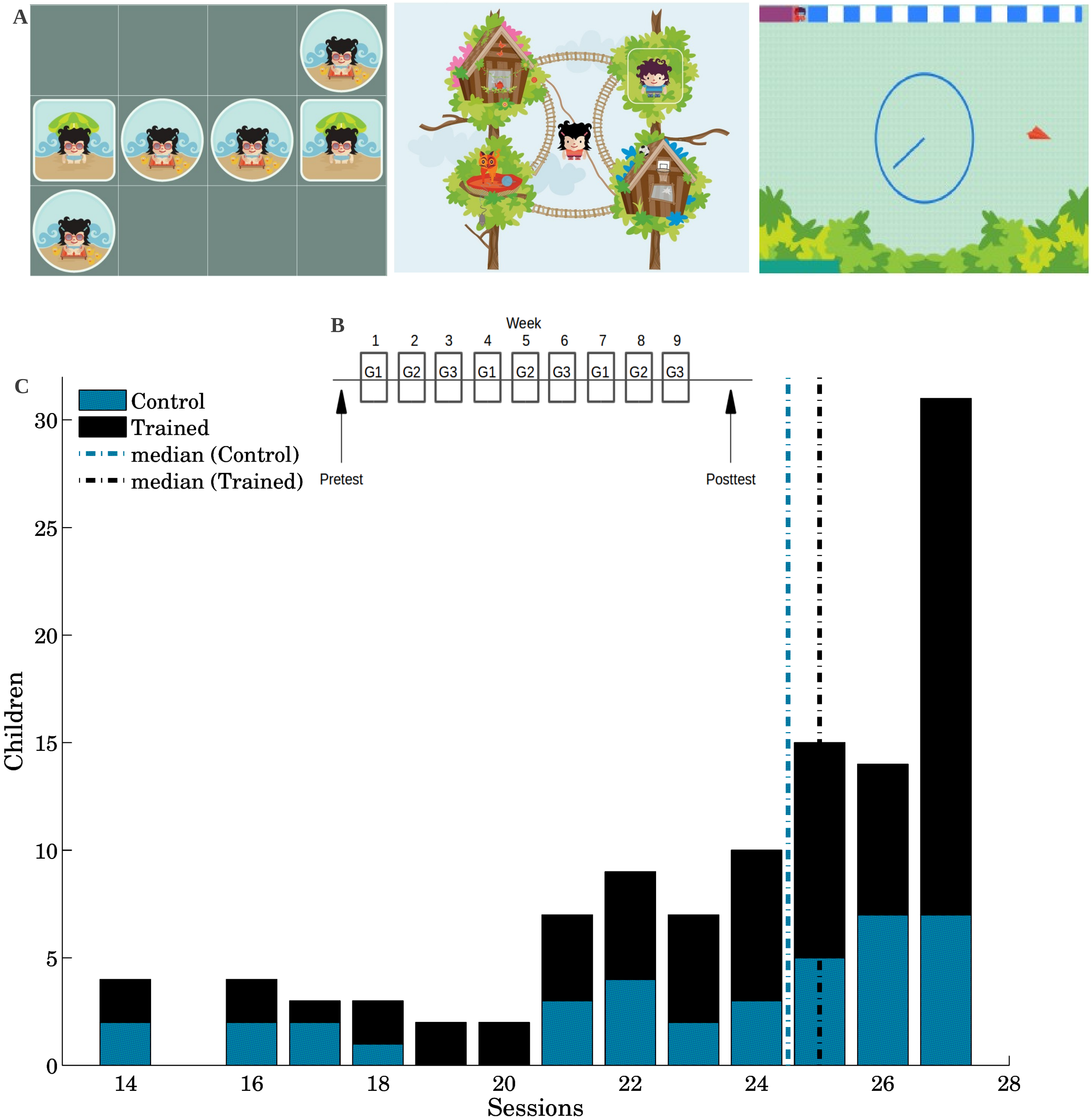

Fig. S1. (S1a) Mate Marote games screens. From left to right: Working memory, planning, and inhibitory control.

(S1b) Time line of the experiment. A diagram of the whole experimental procedure is shown. Children played only one game in each session and performed at least three sessions per week (boxes). Children in the trained group played Mate Marote's games. Trained group game 1 (*G1*): Planning game; trained group game 2 (*G2*): Working Memory game; trained group game 3 (*G3*): inhibitory control. Children in the control group played three different games (*G1*, *G2*, and *G3*) which were selected to minimize cognitive demands with comparable engaging and motor action than Mate Marote's games. Evaluation for transfer was assessed one-to-two weeks before (pretest) and after (posttest) the intervention sessions. Evaluations are shown by arrows.

(S1c) Final number of sessions played. Blue: Control; black: Trained. Dot lines show the median number of sessions played for each experimental group (Blue: Control, 24.5; black: Trained, 25.0).

Table S1

Variable	*Control* *n*	*M (SD)*	*95% CI*	*Trained* *n*	*M (SD)*	*95% CI*	*F*	*p*
Dwelling Score	33	10.18 (1.83)	9.53 – 10.83	62	10.06 (1.81)	9.60 – 10.52	0.16	0.851
Overcrowding Score	33	5.76 (2.50)	4.87 – 6.64	62	6.65 (2.23)	6.08 – 7.21	0.38	0.823
Live in a slum	35	0.77 (0.43)	0.63 – 0.92	68	0.85 (0.36)	0.77 – 0.94	0.15	0.702
Parental Education Score	33	5.55 (3.04)	4.47 – 6.62	62	5.58 (3.17)	4.78 – 6.38	0.98	0.436
Parental Occupation Score	36	1.17 (1.50)	0.66 – 1.67	68	1.12 (1.86)	0.67 – 1.57	0.33	0.801
Receiving subsidies	33	0.52 (0.51)	0.34 – 0.70	62	0.61 (0.49)	0.49 – 0.74	0.02	0.886

Table S1: SES descriptive statistics by experimental group. One-way ANOVA was performed. No significant differences were found.

Table S2

RT (msec)		*Pretest*	*Posttest*
Control	*1st block*	1245 ± 47	1116 ± 34
	2nd block	1186 ± 42	1119 ± 34
	3rd block	1186 ± 41	1061 ± 38
Trained	*1st block*	1142 ± 30	957 ± 27
	2nd block	1151 ± 30	942 ± 23
	3rd block	1186 ± 32	1013 ± 29

Table S2: Mean ± SEM RT in msec of the Child ANT task three blocks

Table S3

Fixed effects	*M*	*SE*	*t-value*
Experimental group	-29.85	36.11	-0.83
Block	4.31	3.82	1.13
Test phase	-87.29	10.62	**-8.22**
Group x phase	-88.45	13.09	**-6.76**

Random effects	*Variance*	*SD*
Subject (n = 111)	30415	174
Classroom (n = 5)	1117	33
Residual (n = 20393)	184352	429

Table S3: Summary of the RT LMM-based statistics for the Child ANT task. Significant differences are stated in bold.

Table S4

	Pretest		*Posttest*	
Correct responses (%)	*Control*	*Trained*	*Control*	*Trained*
Central Cue	91.34 ± 1.72	89.90 ± 1.55	88.71 ± 1.59	91.67 ± 1.24
Spatial Cue	91.78 ± 1.71	88.93 ± 1.51	90.68 ± 1.66	91.10 ± 1.36
No Cue	91.56 ± 1.64	89.33 ± 1.55	87.72 ± 2.10	90.58 ± 1.33
Double Cue	90.24 ± 2.03	90.07 ± 1.45	89.58 ± 1.95	91.44 ± 1.28
Congruent Trial	93.15 ± 1.42	94.52 ± 0.79	90.68 ± 1.48	93.49 ± 0.87
Incongruent Trial	89.31 ± 2.13	84.59 ± 2.35	87.66 ± 1.95	88.90 ± 1.75

Table S4: Percentage of correct responses for every type of cue and trial of the Child ANT task.

Table S5

Fixed effects	*M*	*SE*	*t-value*
Experimental group	-0.02	0.02	-0.79
Test phase	-0.02	0.02	-1.24
Group x phase	0.04	0.02	1.81

Random effects	*Variance*	*SD*
Subject (n = 111)	0.01	0.08
Classroom (n = 5)	<0.01	0.02
Residual (n = 222)	0.01	0.07

Table S5: Summary of the LMM-based statistics for the Child ANT task performance differences. No significant differences were found.

Table S6

Attentional Network Component	*Fixed effects*	*M*	*SE*	*t-value*	*Random effects*	*Variance*	*SD*
	Experimental group	-18.55	27.91	-0.66	Subject (n = 111)	2065	45.44
Alerting	Test phase	-8.24	30.26	-0.27	Classroom (n = 5)	102	10.09
	Group x phase	31.43	37.31	0.84	Residual (n = 222)	17397	131.90
	Experimental group	6.37	29.81	0.21	Subject (n = 111)	4093	63.98
Executive	Test phase	-2.34	30.88	-0.08	Classroom (n = 5)	<0.01	<0.01
	Group x phase	-38.92	38.07	-1.02	Residual (n = 222)	18114	134.59
	Experimental group	42.19	25.78	1.64	Subject (n = 111)	<0.01	<0.01
Orienting	Test phase	79.45	29.56	**2.69**	Classroom (n = 5)	137	11.70
	Group x phase	-115.11	36.45	**-3.16**	Residual (n = 222)	16602	128.90

Table S6: Summary of the LMM-based statistics for each Attentional Network component of the Child ANT task. Significant differences are stated in bold.

Table S7

RT (msec)	Pretest Control	Pretest Trained	Posttest Control	Posttest Trained
Fix congruent	874 ± 42	833 ± 28	762 ± 37	726 ± 31
Fix incongruent	1103 ± 44	1074 ± 36	974 ± 45	919 ± 31
Mix congruent	1299 ± 52	1326 ± 36	1321 ± 53	1188 ± 43
Mix incongruent	1314 ± 57	1365 ± 33	1276 ± 50	1196 ± 38

Table S7: Mean ± SEM RT in msec for the four stimuli type of the Heart-Flower Stroop task

Table S8

Stimuli	*Fixed effects*	*M*	*SE*	*t-value*	*Random effects*	*Variance*	*SD*
Fix congruent	Experimental group	-34.14	43.06	-0.79	Subject (n = 111)	38216	195.49
	Test phase	-95.80	21.22	**-4.51**	Classroom (n = 5)	1099	33.15
	Group x phase	-34.11	26.04	-1.31	Residual (n = 2415)	89365	298.94
Fix incongruent	Experimental group	-36.93	47.78	-0.77	Subject (n = 111)	40556	201.38
	Test phase	-137.53	29.34	**-4.69**	Classroom (n = 5)	1674	40.91
	Group x phase	-22.57	36.31	-0.62	Residual (n = 2081)	148977	385.98
Mix congruent	Experimental group	6.42	51.24	0.12	Subject (n = 111)	40694	201.73
	Test phase	5.53	35.37	0.16	Classroom (n = 5)	1386	37.23
	Group x phase	-157.77	42.61	**-3.70**	Residual (n = 1773)	164724	405.86
Mix incongruent	Experimental group	57.69	49.12	1.17	Subject (n = 111)	37046	192.50
	Test phase	-53.43	33.90	-1.58	Classroom (n = 5)	<0.01	<0.01
	Group x phase	-111.98	41.11	**-2.72**	Residual (n = 1730)	148335	385.10

Table S8: Summary of the RT LMM-based statistics for the four stimuli type of the Heart-Flower Stroop task. Significant differences are stated in bold.

Table S9

Fixed effects	*M*	*SE*	*t-value*
Experimental group	0.02	0.02	0.89
Test phase	<0.01	0.01	0.06
Group x phase	< -0.01	0.02	-0.03

Random effects	*Variance*	*SD*
Subject (n = 111)	0.01	0.11
Classroom (n = 5)	<0.01	<0.01
Residual (n = 9963)	0.15	0.38

Table S9: Summary of the performance LMM-based statistics for the Heart-Flower Stroop task. No significant differences were found.

Table S10

	Pretest		*Posttest*	
Correct responses (%)	*Control*	*Trained*	*Control*	*Trained*
Fix congruent	97.79 ± 0.60	98.17 ± 0.56	89.49 ± 1.79	88.28 ± 1.73
Fix incongruent	86.31 ± 1.94	81.26 ± 2.49	80.90 ± 2.15	81.13 ± 1.81
Mix congruent	64.74 ± 2.77	74.53 ± 2.31	73.64 ± 1.81	76.31 ± 2.00
Mix incongruent	64.19 ± 2.90	68.13 ± 2.64	69.82 ± 2.72	76.28 ± 2.07

Table S10: Percentage of correct responses for every block of the Heart-Flower Stroop task

Table S11

Stimuli	*Fixed effects*	*M*	*SE*	*t-value*	*Random effects*	*Variance*	*SD*
Fix congruent	Experimental group	<0.01	0.02	0.25	Subject (n = 111)	0.01	0.08
	Test phase	-0.08	0.01	**-5.36**	Classroom (n = 5)	<0.01	0.02
	Group x phase	-0.01	0.02	-0.75	Residual (n = 2582)	0.05	0.23
Fix incongruent	Experimental group	-0.06	0.04	-1.61	Subject (n = 111)	0.02	0.14
	Test phase	-0.05	0.02	**-2.20**	Classroom (n = 5)	<0.01	0.05
	Group x phase	0.06	0.03	1.95	Residual (n = 2533)	0.12	0.35
Mix congruent	Experimental group	0.09	0.04	**2.23**	Subject (n = 111)	0.03	0.17
	Test phase	0.08	0.03	**2.70**	Classroom (n = 5)	<0.01	<0.01
	Group x phase	-0.06	0.03	-1.80	Residual (n = 2413)	0.17	0.41
Mix incongruent	Experimental group	0.04	0.04	0.93	Subject (n = 111)	0.03	0.18
	Test phase	0.07	0.03	**2.40**	Classroom (n = 5)	<0.01	<0.01
	Group x phase	0.02	0.03	0.46	Residual (n = 2435)	0.17	0.41

Table S11: Summary of the performance LMM-based statistics for the four stimuli type of the Heart-Flower Stroop task. Significant differences are stated in bold.

Table S12

	Fixed effects	*M*	*SE*	*t-value*	*Random effects*	*Variance*	*SD*
Highest level achieved	Experimental group	0.11	0.25	0.43	Subject (n = 111)	0.57	0.75
	Test phase	0.79	0.22	**3.52**	Classroom (n = 5)	0.08	0.28
	Group x phase	-0.24	0.28	-0.87	Residual (n = 222)	0.96	0.98
No. of errors	Experimental group	0.20	0.43	0.46	Subject (n = 111)	0.11	0.34
	Test phase	0.42	0.49	0.86	Classroom (n = 5)	<0.01	<0.01
	Group x phase	-0.31	0.61	-0.51	Residual (n = 222)	4.59	2.14

Table S12: Summary of the LMM-based statistics for the TOL task highest level achieved and number of errors differences. Significant differences are stated in bold.

Table S13

School subjects	*Exp. Group*	*Low attendance*				*High attendance*			
		1st bim	*2nd bim*	*3rd bim*	*4th bim*	*1st bim*	*2nd bim*	*3rd bim*	*4th bim*
Language	*Control*	6.29 ± 0.30	6.36 ± 0.42	6.36 ± 0.44	6.89 ± 0.49	6.87 ± 0.33	7.13 ± 0.31	8.05 ± 0.37	8.50 ± 0.36
	Trained	6.48 ± 0.23	6.74 ± 0.29	6.98 ± 0.32	7.92 ± 0.31	6.84 ± 0.15	7.26 ± 0.23	7.91 ± 0.23	8.38 ± 0.24
Mathematics	*Control*	6.39 ± 0.29	6.81 ± 0.35	6.67 ± 0.45	7.18 ± 0.42	6.76 ± 0.36	6.68 ± 0.39	7.39 ± 0.36	7.94 ± 0.37
	Trained	6.42 ± 0.22	6.74 ± 0.26	6.84 ± 0.33	7.87 ± 0.28	6.74 ± 0.15	7.06 ± 0.26	7.19 ± 0.28	7.80 ± 0.25
Social Behavior	*Control*	7.03 ± 0.26	7.22 ± 0.33	7.09 ± 0.27	7.42 ± 0.27	7.21 ± 0.30	7.39 ± 0.29	7.39 ± 0.33	7.71 ± 0.29
	Trained	6.56 ± 0.17	6.90 ± 0.25	6.90 ± 0.28	7.24 ± 0.22	6.95 ± 0.18	7.61 ± 0.24	7.58 ± 0.21	7.50 ± 0.20
Rights and Responsibilities of Citizenship	*Control*	7.17 ± 0.24	6.00 ± 0.21	6.35 ± 0.29	6.87 ± 0.34	7.33 ± 0.30	6.67 ± 0.13	6.62 ± 0.11	6.82 ± 0.32
	Trained	7.33 ± 0.23	6.00 ± 0.16	6.50 ± 0.20	6.80 ± 0.27	7.18 ± 0.18	6.50 ± 0.10	6.79 ± 0.13	6.97 ± 0.20
Collaborative Work	*Control*	6.71 ± 0.21	6.94 ± 0.25	6.97 ± 0.26	7.03 ± 0.30	7.21 ± 0.30	7.45 ± 0.28	7.79 ± 0.35	7.74 ± 0.26
	Trained	6.69 ± 0.21	6.76 ± 0.22	6.94 ± 0.25	7.08 ± 0.21	7.00 ± 0.15	7.27 ± 0.20	7.45 ± 0.21	7.47 ± 0.18
Technology Arts	*Control*	6.50 ± 0.00	6.83 ± 0.18	7.09 ± 0.22	7.28 ± 0.23	6.71 ± 0.14	6.71 ± 0.26	7.34 ± 0.28	7.66 ± 0.23
	Trained	6.62 ± 0.09	6.95 ± 0.18	7.34 ± 0.22	7.47 ± 0.22	6.80 ± 0.11	6.84 ± 0.18	7.20 ± 0.18	7.60 ± 0.16
Visual Arts	*Control*	8.13 ± 0.25	8.00 ± 0.27	7.91 ± 0.22	7.89 ± 0.34	8.03 ± 0.27	7.63 ± 0.26	7.45 ± 0.24	7.87 ± 0.33
	Trained	8.27 ± 0.17	8.11 ± 0.21	7.97 ± 0.20	7.82 ± 0.20	8.07 ± 0.19	7.93 ± 0.21	7.64 ± 0.17	8.03 ± 0.19
Music	*Control*	6.71 ± 0.14	7.06 ± 0.26	6.85 ± 0.29	7.21 ± 0.30	7.11 ± 0.25	7.13 ± 0.22	7.13 ± 0.22	7.68 ± 0.32
	Trained	7.00 ± 0.20	7.23 ± 0.27	7.21 ± 0.22	7.16 ± 0.26	6.85 ± 0.12	6.84 ± 0.15	7.28 ± 0.20	7.20 ± 0.21
Physical Education	*Control*	7.00 ± 0.24	7.22 ± 0.29	6.85 ± 0.18	7.03 ± 0.21	7.24 ± 0.37	7.55 ± 0.34	7.21 ± 0.26	7.47 ± 0.32
	Trained	6.80 ± 0.19	6.85 ± 0.23	7.08 ± 0.19	7.08 ± 0.19	7.04 ± 0.18	7.05 ± 0.20	7.36 ± 0.20	7.54 ± 0.20
Foreign Language	*Control*	6.71 ± 0.21	6.94 ± 0.20	6.74 ± 0.15	6.61 ± 0.24	7.00 ± 0.32	7.34 ± 0.28	7.24 ± 0.23	7.32 ± 0.30
	Trained	6.44 ± 0.19	7.02 ± 0.22	6.69 ± 0.21	6.61 ± 0.25	7.05 ± 0.19	7.28 ± 0.18	7.50 ± 0.17	7.19 ± 0.19
History-Social Sciences	*Control*	6.61 ± 0.19	6.72 ± 0.15	6.85 ± 0.24	7.36 ± 0.27	6.89 ± 0.22	6.82 ± 0.17	7.24 ± 0.23	7.66 ± 0.34
	Trained	6.73 ± 0.14	6.56 ± 0.06	6.69 ± 0.17	7.45 ± 0.22	6.79 ± 0.12	6.79 ± 0.13	7.24 ± 0.18	7.79 ± 0.20
Natural Sciences	*Control*	6.61 ± 0.19	6.72 ± 0.15	6.85 ± 0.24	7.36 ± 0.27	7.00 ± 0.24	6.82 ± 0.17	7.13 ± 0.22	7.66 ± 0.34
	Trained	6.73 ± 0.14	6.56 ± 0.06	6.69 ± 0.17	7.45 ± 0.22	6.79 ± 0.12	6.79 ± 0.13	7.29 ± 0.18	7.74 ± 0.21
Responsibility	*Control*	6.39 ± 0.32	6.50 ± 0.35	6.59 ± 0.33	6.79 ± 0.37	7.11 ± 0.33	7.03 ± 0.37	7.68 ± 0.38	7.63 ± 0.31
	Trained	6.56 ± 0.23	6.31 ± 0.28	6.74 ± 0.29	7.02 ± 0.26	7.10 ± 0.16	7.22 ± 0.26	7.61 ± 0.22	7.81 ± 0.19
Attend to orders	*Control*	6.61 ± 0.36	7.00 ± 0.39	7.03 ± 0.40	7.16 ± 0.38	7.00 ± 0.36	7.08 ± 0.38	7.18 ± 0.36	7.29 ± 0.39
	Trained	6.38 ± 0.22	6.39 ± 0.30	6.71 ± 0.31	7.10 ± 0.31	6.90 ± 0.21	7.28 ± 0.29	7.33 ± 0.25	7.33 ± 0.25

TableS13: Mean ± SEM grades for all 14 school subjects, two experimental groups, two attendance groups and four bimesters.

Table S14

Fixed effects	*M*	*SE*	*t-value*
Bimester	0.15	0.02	**8.05**
Attendance	0.08	0.15	0.57
Attendance*bimester	0.05	0.02	**2.05**

Random effects	*Variance*	*SD*
Subject (n = 111)	0.42	0.65
School subject (n = 14)	0.07	0.27
Classroom (n = 5)	0.12	0.34
Residual (n = 5920)	1.15	1.07

Table S14: Summary of the LMM-based statistics for the school grades not considering experimental groups. Significant differences are stated in bold.

Table S15

		Fixed effects	*M*	*SE*	*t-value*	*Random effects*	*Variance*	*SD*
Low attendance	*Language & Math*	Experimental group	0.14	0.35	0.40	Subject (n = 51)	0.94	0.97
		Bimester	0.55	0.24	**2.25**	Classroom (n = 5)	0.63	0.79
		Group x bimester	0.67	0.31	**2.14**	Residual (n = 189)	1.05	1.03
	Informal	Experimental group	0.04	0.21	0.19	Subject (n = 51)	0.31	0.56
		Bimester	0.44	0.14	**3.06**	Classroom (n = 5)	0.38	0.62
		Group x bimester	0.11	0.18	0.59	Residual (n = 501)	0.95	0.98
	Control	Experimental group	-0.03	0.18	-0.14	Subject (n = 51)	0.22	0.47
		Bimester	0.24	0.14	1.74	Classroom (n = 5)	0.08	0.29
		Group x bimester	0.02	0.17	0.14	Residual (n = 660)	1.15	1.07
High attendance	*Language & Math*	Experimental group	-0.02	0.34	-0.05	Subject (n = 50)	0.83	0.91
		Bimester	1.32	0.24	**5.54**	Classroom (n = 5)	0.28	0.53
		Group x bimester	0.14	0.30	0.45	Residual (n = 194)	1.02	1.01
	Informal	Experimental group	-0.11	0.25	-0.43	Subject (n = 50)	0.53	0.73
		Bimester	0.51	0.14	**3.61**	Classroom (n = 5)	0.27	0.52
		Group x bimester	0.22	0.18	1.19	Residual (n = 490)	0.95	0.97
	Control	Experimental group	-0.16	0.19	-0.82	Subject (n = 50)	0.26	0.51
		Bimester	0.33	0.13	**2.54**	Classroom (n = 5)	0.15	0.39
		Group x bimester	0.05	0.17	0.29	Residual (n = 652)	1.09	1.04

Table S15: Summary of the LMM-based statistics for each group of subjects and attendance. Significant differences are stated in bold.

Table S16

Impulsive responses (%)		*Pretest*	*Posttest*
	Congruent	0.22	2.22
Control	Incongruent	0.22	1.32
	Mix	0.33	2.74
	Congruent	0.23	1.83
Trained	Incongruent	0.57	0.80
	Mix	0.46	2.23

Table S16: Percentage of impulsive responses (i.e. RT < 250 msec) for the Heart-Flower Stroop task.