MIND, BRAIN, AND EDUCATION

A Pilot Study of Classroom-Based Cognitive Skill Instruction: Effects on Cognition and Academic Performance

Allyson P. Mackey[1], Anne T. Park[1], Sydney T. Robinson[2], and John D. E. Gabrieli[3]

ABSTRACT— Cognitive skills are associated with academic performance, but little is known about how to improve these skills in the classroom. Here, we present the results of a pilot study in which teachers were trained to engage students in cognitive skill practice through playing games. Fifth-grade students at an experimental charter school were randomly assigned to receive cognitive skill instruction (CSI) or instruction in geography and typing. Students in the CSI group improved significantly more than the control group on a composite measure of cognitive skills. CSI was more effective for students with lower standardized test scores. Although there was no group effect on test scores, cognitive improvement correlated positively with test score improvement only in the CSI group. Beyond showing that cognitive skills can be improved in the classroom, this study provides lessons for the future of CSI, including changes to professional development and challenges for scalability.

Cognitive skills such as processing speed, working memory, and fluid reasoning are highly correlated with educational outcomes (Finn et al., 2014; Gathercole, Pickering, Knight, & Stegmann, 2004). These correlations suggest that improving cognitive skills could lead to better academic performance: if students could think more quickly, hold more information in mind, and solve problems more efficiently, they would be better able to learn in the classroom. Further, curricula and standardized assessments are beginning to place a stronger emphasis on cognitive skills. For example, Common Core guidelines state that students must "make sense of problems and persevere in solving them, reason abstractly and quantitatively, and evaluate the validity of the reasoning." These shifting expectations have led educators to ask the question of how to best shape and support cognitive skill development in the classroom. Here, we describe the cognitive and academic effects of a cognitive skill curriculum that resulted from an interdisciplinary collaboration between educators and neuroscientists.

To date, the programs that have been associated with the most robust cognitive skill gains have been immersive play-based curricula implemented in early grades (for review see Diamond, 2013; Diamond & Lee, 2011; Diamond & Ling, 2016). Prominent examples include Montessori (Lillard & Else-Quest, 2006) and Tools of the Mind (Blair & Raver, 2014; Diamond, Barnett, Thomas, & Munro, 2007), but new programs are continually being designed and evaluated (e.g., Hermida et al., 2015). These types of curricula are based on developmental psychology research suggesting that young children learn best through play (for review, see Bodrova & Leong, 2003; Hirsh-Pasek, Golinkoff, Berk, & Singer, 2009), and align well with neuroscience research on motivation and plasticity (for review, see Green & Bavelier, 2008). An emphasis on play is much less commonly applied to curriculum design in older grades. One reason for the shift away from play is an increased emphasis on math and reading

[1]Department of Psychology, University of Pennsylvania
[2]School of Nursing, University of Pennsylvania
[3]Brain and Cognitive Sciences, Massachusetts Institute of Technology

Address correspondence to Allyson P. Mackey, Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104; e-mail: mackeya@upenn.edu

skills. School time is limited, and it is necessary to make tradeoffs among all possible educational activities. These limitations enhance the allure of brief, modular programs to improve cognitive skills, typically referred to as "cognitive training" programs.

The majority of research on cognitive training has focused on the degree of transfer among cognitive skills (for review of this mixed literature, see Diamond & Ling, 2016; Melby-Lervåg & Hulme, 2013; Simons et al., 2016), and not on transfer to academics. Some studies of working memory training have found transfer to laboratory tests of academic skills such as reading (Loosli, Buschkuehl, Perrig, & Jaeggi, 2012) and math (Holmes, Gathercole, & Dunning, 2009), but others have found weak or null results (Dunning, Holmes, & Gathercole, 2013; Roberts et al., 2016; Rode, Robson, Purviance, Geary, & Mayr, 2014; St Clair-Thompson, Stevens, Hunt, & Bolder, 2010). Training inhibitory control in addition to working memory has led to gains in math reasoning (Blakey & Carroll, 2015). Training of a wide variety of cognitive skills with video games has led to gains in the accuracy and speed of calculation (Miller & Robertson, 2011). In the first study to assess transfer to a real-world academic measure, training inhibitory control, working memory, and planning together was found to lead to improvements in school grades, but only for students with low attendance (Goldin et al., 2014). Taken together, these results suggest that brief, modular cognitive training holds promise for boosting academic performance, but that there is room to develop more effective programs.

One possible approach to improving cognitive training programs is to more deeply involve teachers in their design and implementation. Teachers have considerable expertise in student instruction and motivation, as well as an invaluable understanding of the practical constraints of the classroom environment (e.g., what technology is available, which students can work together, how much time students need to transition between activities, etc.). Involving teachers in the creation of new programs could also enhance their ability to implement these programs independently in their classrooms. Other avenues toward improved programs include broadening the set of cognitive skills trained (Diamond, 2013; Diamond & Lee, 2011), and harnessing the motivational power of play.

In the present study, we worked with teachers to develop a game-based curriculum targeted at a broad array of cognitive skills. This curriculum is an extension of a previous study (Mackey, Hill, Stone, & Bunge, 2011), and expands this work by targeting multiple cognitive skills, having the training be implemented by teachers in the classroom as part of the regular school schedule, and including statewide standardized tests of academic achievement as an outcome measure. The cognitive skill instruction (CSI) approach in the present study relies on four hypothesized tenets of effective instruction. First, CSI involves rich social interactions between teachers and students. Second, CSI incorporates novel activities that can be tailored to a student's performance level to keep the student engaged but not frustrated. Third, the activities involve multiple cognitive domains (e.g., visuospatial, semantic, and numerical) in order to maximize breadth of transfer. Finally, students are made explicitly aware of the skills they are learning and asked to apply these skills to new contexts. In the present study, we demonstrate the feasibility and challenges of CSI, and explore its degree of transfer to academic performance.

METHODS

Participants

This research was approved by the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology. Participants were fifth grade students at a new charter school. Participating in either the CSI curriculum or the control curriculum (typing and geography) was a mandatory part of the school day. Students were pseudo-randomly assigned to either CSI or typing/geography. Two groups were randomly generated based on de-identified data until they did not differ significantly ($p > .2$) on age, gender, number of students with special needs, or standardized test scores from the previous year (fourth grade).

Students and parents were given the opportunity to decide whether or not to participate in assessments, and whether or not to make demographic and educational data available to researchers. Of the 48 students enrolled at the beginning of the school year, 46 participated in research (22 in the CSI group, 24 in the control group). Demographic information for these participants is shown in Table 1. Fourth-grade test scores were on the Massachusetts Comprehensive Assessment System and fifth grade scores were on the Partnership for Assessment of Readiness for College and Careers test. Concordance tables created by the Massachusetts Department of Education were used to compare these two measures. Tables are available at http://www.doe.mass.edu/parcc/results.html.

School Setting and Professional Development

We collaborated with a new experimental charter school as the school enrolled its first class of students. An innovative feature of this school is its focus on small-group instruction led by "tutors." The tutor program is similar in some ways to larger scale programs that bring high-achieving college students into the classroom (e.g., Teach for America). Like Teach for America, acceptance to the tutoring program is highly competitive and training is brief but intensive during the summer prior to the first year of teaching

Table 1
Participant Demographics

	CSI (n = 22)	*Control* (n = 24)
Age at time 1	11.63 (0.41)	11.67 (0.35)
Gender	12 M, 10 F	13 M, 11 F
Percent free or reduced price lunch	82	71
Race	7 African American 1 Asian 14 White	6 African American 0 Asian 18 White
Ethnicity	14 Hispanic	16 Hispanic
Special needs	4 (1 communication, 1 emotional, 1 health, 1 neurological)	6 (1 autism, 2 communication, 1 health, 2 specific learning disabilities)
Previous year's Math standardized test score	247.5 (11.0) 62nd percentile *n* = 19	249.8 (16.6) 67th percentile *n* = 23
Previous year's English/Language Arts standardized test score	247.3 (14.6) 67th percentile *n* = 19	250.3 (15.2) 73rd percentile *n* = 23

Note. Demographic information and test scores were provided by the school with parental consent.

(Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005). Unlike Teach for America, experienced master teachers supervise the tutors, and determine instructional plans.

As part of tutor training, we gave a presentation on the importance and measurement of cognitive skills, and led a discussion of how these skills could be supported in the classroom. We explained the purpose and rules of the custom games (see below), and asked the tutors to play the games and practice explaining them to each other. This professional development session lasted 2 hr. Throughout the year, an open line of communication was kept between the tutors, the master teachers, and the research team. At the end of the year, the tutors were asked to complete a survey about their experiences with CSI.

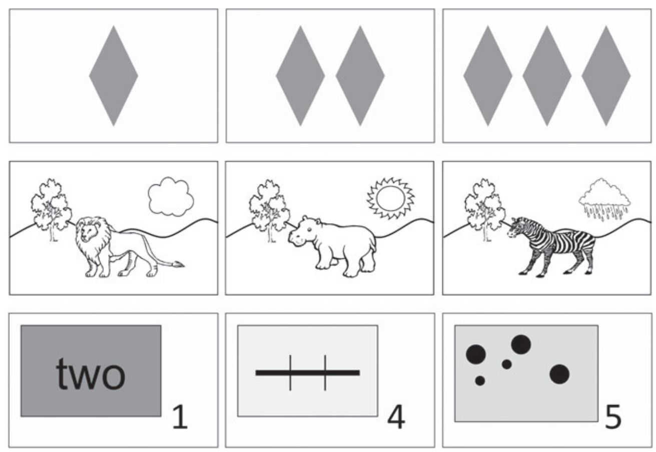

Fig. 1. Examples of "sets" for all three decks of cards. In the first set, the cards have the same symbol, color, and shading, and are different in number. In the second set, the cards have the same tree and horizon, and have different animals and weather. In the third set, the cards are different for all four properties (color, notation, number, and relationship). The colors used in the number set were blue, yellow, and gray.

Intervention Structure and Control Curriculum

CSI and the control curriculum were implemented for 20 min per day, 4 days per week. Both curricula were administered by tutors under the direct supervision of master teachers. Students who were assigned to the control curriculum learned typing and geography by playing computer games (Typing Club and Sheppard Software, respectively). Students played each game 2 days per week.

Card Games

When initially designing the CSI curriculum, the master teachers suggested we choose as few games as possible to limit the demands on the tutors. We focused on four card games, each of which could be played with three different decks (Figure 1). Two custom-made card decks were prepared based on the commercially available game, SET®. Each card deck had four different properties, and each property had three different options. In the traditional SET deck, the four properties are color, shape, number, and shading. For the CSI Safari deck, the four properties were horizon, plant, animal, and weather. For the CSI Ciphers deck, the four properties were color, notation, number inside the rectangle, and relationship between the number inside the rectangle and the number outside the rectangle (same, greater, or smaller). The students were responsible for keeping track of these various properties while playing four custom card games: Speed, Series, Set, and Sort. Each game was played on a different day of the week for 9 weeks.

For Speed, a tutor divided a deck of cards between two students, and announced the two card properties that the students would need to keep in mind. The game began with each student flipping a card face up onto the table, and drawing a hand of three cards. On the count of three, each student then proceeded to draw and put down cards that matched the two current face-up cards on either of the two predetermined properties. The goal of the game was to be the first player to run out of cards. Speed targeted processing speed, and was based on traditional card games (Mackey et al., 2011).

Series was an *n-back* card game in which a tutor continuously flipped over cards from a deck, and students indicated the appearance of a target card by tapping the table. A target card was defined as a card that matched a card played *n* turns ago on a predetermined property. Series targeted working memory, and was based on computerized n-back tasks used in prior cognitive training research (e.g., Jaeggi, Buschkuehl, Jonides, & Perrig, 2008).

For Set, students were shown an array of 12 cards and asked to pick out "sets" of cards, defined as a group of three cards that were either all the same or all different for each of the card properties (Figure 1). Tutors could make the game easier by walking students through a single set (i.e., picking two cards and asking students to describe the card that would complete the set), or by playing a noncompetitive version of the game (i.e., students could help each other find sets or ask the tutor for hints). Set taxed fluid reasoning and was based on a commercially available game played in Mackey et al., 2011.

For Sort, students were divided into teams of two players, a *sorter* and a *checker*. The sorter was responsible for sorting a deck of cards into two piles based on two predetermined properties (e.g., sort cards into one pile if the cards contain a zebra or a cactus, and sort all other cards into the other pile). The checker was responsible for making sure the sorter made no mistakes (to encourage accurate sorting). When the students were about halfway through their card decks, a tutor briefly stopped the game and changed the sorting rule to two different options (e.g., zebra to lion, cactus to pine). Students competed to be the first player to finish sorting their deck. Sort was based on the task-switching games played in Karbach and Kray (2009).

Expanded CSI Curriculum

After 9 weeks of playing the custom games, the tutors reported that students were no longer engaged by the limited selection of card games so we increased the number and variety of games. We chose commercially available games that taxed fluid reasoning, working memory, and processing speed across a variety of domains (visuospatial, semantic, numeric). Many of these games were also used in Mackey et al., 2011. We additionally included a new variant of a working memory game using the card decks described above. Descriptions of the games are provided in Table 2.

Another change was to shift the structure of training to include two groups of six stations in the classroom. One group of stations was implemented on Mondays and Tuesdays, and the other group was implemented on Wednesdays and Thursdays. An additional six tutors joined the program, bringing the total to 12. This change allowed each tutor to become an expert in one game rather than learning a wide variety of games. Students rotated through the stations for roughly 30 weeks. Researchers did not provide direct instruction on these games. Tutors had the freedom to adjust game difficulty and to change the rules to suit their students.

Assessments

Paper-and-pencil assessments were administered in a classroom environment by testers who were blind to group assignment (details in Table 3). Assessments were given in two sessions, the first lasting 80 min, and the second lasting 40 min. Assessments were given at the beginning and end of the school year. Assessments were chosen to measure fluid reasoning and processing speed across visuospatial, semantic, and numeric domains. Due to challenges described below, only one working memory measure was included.

The Test of Nonverbal Intelligence 4 (TONI-4; Brown, Sherbenou, & Johnsen, 2010) was chosen as a visuospatial reasoning measure. It is a matrix reasoning test, which involves finding the missing piece of a pattern or identifying a set of shapes that follows the same pattern as a target item. We modified this test for classroom administration by printing all items for each student, and giving them 10 min to answer as many items as possible. Animal Sorting from the Developmental NEuroPSYchological Assessment (NEPSY) (Korkman, Kirk, & Kemp, 1998) was selected as the semantic reasoning measure. In this task, students sorted eight pictures of animals into two groups of four using dimensions such as the direction the animals are facing, the presence or absence of water, or the presence or absence of a thick border. Three numerical reasoning tests were selected. In Number Series from the Woodcock-Johnson Test of Cognitive Ability III (Woodcock, McGrew, & Mather, 2001), students determined the pattern that governed a series of numbers and filled in the missing number. Participants were given 8 min to complete as many items as possible. In a custom measure we called *Number Patterns*, each question contained two lines: one containing a pattern that demonstrated a rule, and the other containing a pattern with a missing piece. The two lines differed in their representation of number. Representations included geometric shapes (number of sides), circles, tally marks, and coins. In a custom measure we called *Candy Calculation*, algebra problems were represented with

Table 2
Commercial Game Details

Game	*Skill*	*Brief description*
Chocolate Fix®[a]	FR	Challenge cards provide rules (shape, color, position) for placing chocolates in a 3 × 3 array.
Mastermind™[a]	FR	The codemaker chooses a sequence of four colors. The codebreaker guesses the pattern, getting feedback on color and sequence position after each guess.
SET®[a]	FR	Details described in Methods section. All three decks were used.
Swish™	FR	Transparent cards contain balls and hoops in four colors. Players find two or more cards that can be overlaid so that all balls fit into matching hoops.
Tangrams[a]	FR	Players combine geometric shapes to match a target pattern.
Tip Over®	FR	Players tip over crates to create a path for the "tipper man" to reach a target crate.
Qbitz™	FR	Similar to Wechsler Block Design. Players quickly arrange blocks to match a target pattern.
Visual Brainstorm	FR	Cards present spatial and verbal logic problems.
Space Mines Patrol, CogMed™	WM	Computerized spatial span task with asteroids. Trial version increases in difficulty within a session but does not allow progress to be saved.
Memory (Custom)	WM	Three to four cards from the custom decks are shown to a player. The cards are flipped over, and after a delay the tutor asks the player to point to the card that matches a given property. Presentation time and delay are varied to match students' ability levels.
Spot It ™	PS	Cards show eight pictures in various sizes and orientations. Players search for a picture match between their cards and a center card.
Blink ®[a]	PS	Players place cards on a center deck as quickly as possible, matching color, number, or shape.

Note. FR = fluid reasoning; WM = working memory; PS = processing speed. [a]This game was also used in Mackey et al. (2011).

pictures of sweets. Easier problems contained one tier (e.g., 5 lollipops = 35, what is the value of one lollipop?), and harder problems contained multiple tiers and mixed sweets per tier (e.g., 3 donuts and 2 cupcakes = 23, 2 donuts and 2 cupcakes = 18, 2 donuts and 4 jellybeans = 42, what is the value of one jellybean?).

Symbol Search and Coding from the Wechsler Intelligence Scale for Children IV (WISC-IV; Wechsler, 2003) were chosen as the visuospatial speed measures. In Symbol Search, students determined whether a target symbol was present among a row of symbols. They had 2 min to complete as many items as possible. In Coding, students used a cipher to write symbols that corresponded to specific digits. They had 2 min to write as many symbols as possible. Decision Speed from the Woodcock-Johnson Test of Cognitive Ability III (Woodcock et al., 2001) was chosen as the semantic speed measure. In this task, students selected two pictures that were semantically related from a row of seven pictures. They had 3 min to complete as many rows as possible. The numerical speed measure was a custom task called the "Speeded Arithmetic" task. Students completed as many arithmetic problems as possible in 1 min. Three versions were administered: subtraction, division, and a mix of subtraction and division.

Working memory is difficult to assess in a classroom environment without computers for each student. Students get easily distracted and lose the information they were attempting to hold in mind. It is also difficult to account for individual differences in ability with a group test administration, that is, lower performing students must attempt harder items than they would have been given in a test with a ceiling. This leads to more frustration. We administered a custom measure of semantic working memory called the *Animal Span* task. In this task, students were asked to remember sequences of animals, varying from loads of two to seven. The task began with two practice trials with a load of two, followed by two trials of each load from two to seven. Line drawings of animals were projected onto a screen in front of the class. Animals appeared one at a time for 2 sec each, with 1 sec breaks between animals. After each trial, six animals with associated letters were displayed for 10 s (12 s for loads of six and seven). Students were asked to write down the letters associated with the animals shown, in the correct order. In the first version of the task, students needed to remember the animals in the forward order. In the second version, students were asked to report the animals in the reverse order. Classroom management was a challenge during this task; therefore, we are not confident that scores reflect students' true working memory capacities.

Table 3
Assessments Details

Test	*Source*	*Time (min)*	*Construct measured*
Test of Nonverbal Intelligence	Test of Nonverbal Intelligence, 4th ed. (TONI-4)	10	Fluid reasoning
Animal sorting	NEPSY (A Developmental NEuroPSYchological Assessment), modified so that all eight pictures appeared on one page	6	Fluid reasoning
Coding	Wechsler Intelligence Scale for Children, 4th ed. (WISC-IV)	2	Processing speed
Symbol search	WISC IV	2	Processing speed
Decision speed	Woodcock-Johnson Test of Cognitive Ability III, Test 16	3	Processing speed
Animal span	Custom	N/A	Working memory
Number series	Woodcock-Johnson Test of Cognitive Ability III, Test 24	5	Numerical reasoning
Number patterns	Custom	8	Numerical reasoning
Candy Calculation	Custom	8	Numerical reasoning
Speeded arithmetic	Custom, based on "Mad Minute" tests given in school	3	Numerical processing speed

Table 4
Correlations Between Assessments and Academic Data at Time 1

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
1. Cognitive composite	—												
2. Test of Nonverbal Intelligence	0.64***	—											
3. Animal sorting	0.34*	0.37*	—										
4. Number series	0.60***	0.35*	0.10	—									
5. Coding	0.38**	0.01	0.12	0.04	—								
6. Symbol search	0.46**	0.11	0.14	0.03	0.60***	—							
7. Decision speed	0.52***	0.14	−0.22	0.14	0.24	0.32*	—						
8. Speeded arithmetic	0.65***	0.26	−0.03	0.57***	0.29	0.22	0.23	—					
9. Number patterns	0.60***	0.37*	0.04	0.43**	−0.06	0.09	0.38*	0.36*	—				
10. Candy Calculation	0.49***	0.45**	0.02	0.23	−0.10	−0.10	0.19	0.24	0.37*	—			
11. Animal span	0.43**	0.11	0.17	0.06	0.14	0.21	0.15	0.27	−0.05	0.09	—		
12. Math	0.60***	0.48**	0.40**	0.63***	−0.05	0.05	0.03	0.50***	0.43**	0.53***	0.02	—	
13. ELA	0.33*	0.23	0.42**	0.25	−0.09	0.03	0.06	0.36*	−0.01	0.33*	0.07	0.62***	—

Note. ELA = English/Language Arts. Standardized test scores are from the end of the previous academic year.
* $p < .05$; ** $p < .01$; *** $p < .001$.

Data Analysis

Assessments were standardized (*z*-scored) across groups and time points to facilitate comparisons across measures. All *z*-scored assessments were averaged to create a cognitive composite. Analyses of variance (ANOVAs) were conducted on improvement in the cognitive composite and standardized test scores to test for Group × Time interactions. Because variability at Time 1 can influence change scores, we additionally conducted analyses of covariance (ANCOVAs) to test for an effect of group, controlling for Time 1 scores.

RESULTS

The cognitive composite at Time 1 correlated positively with entering standardized test scores (scores from the previous year) in Math ($r(40) = .60$, $p < .0001$) and English/Language Arts (ELA) ($r(40) = .33$, $p = .03$). Correlations between cognitive and academic measures are shown in Table 4. The CSI group improved significantly more than the control group on the cognitive composite ($F(1,44) = 4.62$, $p = .04$, partial $\eta^2 = .09$; Table 5 ANOVA; Figure 2a). This effect was not driven by improvement in a particular skill; rather it reflected a general positive trend across measures (Table S1, Supporting Information). Controlling for Time 1 cognitive composite did not change the results (Table 5 ANCOVAs). The CSI and Control groups did not differ significantly on any measures at Time 1 or at Time 2 (Tables 6 and S2).

Students with lower entering standardized test scores showed a greater cognitive benefit from CSI: there was a significant interaction between average entering test score (Math and ELA in fourth grade) and training group ($F(1,38) = 4.46$, $p = .04$, partial $\eta^2 = .11$). To better understand this continuous interaction, students were divided into two groups based on a median split on average entering test score within training group. Students in the lower scores

Table 5
Cognitive and Academic Change Scores

	CSI				*Control*				*ANOVA: Group × Time interaction*		*ANCOVA: Effect of group on change score*	
	M	*SD*	*t*	*p*	*M*	*SD*	*t*	*p*	*F*	*p*	*F*	*p*
All students												
Cognitive composite	0.94	0.32	13.93	<.0001	0.73	0.35	10.10	<.0001	**4.62**	**.04**	**4.86**	**.03**
Math	0.79	0.51	6.72	<.0001	0.64	0.46	6.73	<.0001	0.91	.35	0.62	.43
ELA	−0.42	0.61	−2.96	.008	−0.35	0.90	−1.87	.08	0.07	.79	0.68	.42
Lower scores												
Cognitive composite	1.02	0.35	9.22	<.0001	0.57	0.33	5.86	.0001	**9.72**	**.005**	**7.38**	**.01**
Math	0.85	0.63	4.26	.002	0.87	0.44	6.84	<.0001	0.01	.93	0.00	1.00
ELA	−0.28	0.65	−1.36	.21	−0.10	0.87	0.39	.71	0.29	.59	1.03	0.32
Higher Scores												
Cognitive composite	0.91	0.32	8.69	<.0001	0.90	0.32	9.48	<.0001	0.01	.94	0.09	.77
Math	0.71	0.35	6.08	.0003	0.39	0.34	3.83	.003	4.24	.05	0.05	.83
ELA	−0.57	0.57	−3.00	.02	−0.63	0.89	2.34	.04	0.03	.87	0.03	.87

Note. CSI: All students $n = 22$ (with test score data $n = 19$); lower scores $n = 10$; higher scores $n = 9$. Control: All students $n = 24$ (with test score data $n = 23$); lower scores $n = 12$; higher scores $n = 11$. Boldface values indicate significance ($p < .05$) for ANOVA and ANCOVA results.

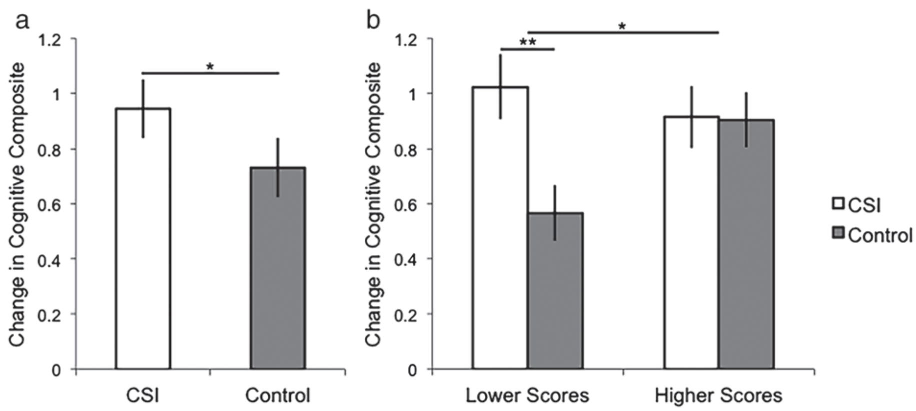

Fig. 2. (a) Cognitive composite change by curriculum group. (b) Cognitive composite change by curriculum group and test score group. $^*p < .05$; $^{**}p < .01$.

group ($n = 22$) were, on average, in the 50th percentile statewide in Math and in the 55th percentile in ELA. Students in the higher scores group ($n = 20$) were, on average, in the 85th percentile in both Math and ELA. The higher scores group also had higher cognitive composites at Time 1 ($t(40) = 2.75$, $p = .009$). Within the lower scores group, the CSI group improved significantly more than the control group on the cognitive composite ($F(1,20) = 9.72$, $p = .005$, partial $\eta^2 = .33$; Table 5, Figure 2b). No effect of CSI was found in the higher scores group ($F(1,18) = .01$, $p = .94$, partial $\eta^2 \leq .001$; Figure 2b; Table 5). Scores on all measures for the lower and higher scores groups are shown in Tables S3 and S4, respectively.

Across all students, Math standardized test scores improved from 2014 to 2015 ($t(41) = 9.52$, $p < .0001$), but ELA scores worsened ($t(41) = -3.18$, $p = .003$). The CSI group did not improve significantly more than the control group on standardized tests of Math or ELA (Tables 5 and 6). Neither the lower scores group nor the higher scores group showed an effect of CSI on test scores (Tables 5 and 6). However, within the CSI group, cognitive improvement was significantly positively correlated with Math improvement ($r(17) = .50$, $p = .03$; Figure 3), and there was a trend toward a significant positive correlation between cognitive improvement and ELA improvement ($r(17) = .44$, $p = .06$). A relationship between cognitive change and test score change was not observed in the control group (Math: $r(21) = -.33$, $p = .13$; ELA: $r(21) = -.22$, $p = .31$). Correlations among cognitive and academic change measures were not different by test score group (Supporting Information Supplemental Table 5).

Eight tutors responded to our request for feedback at the end of the year. The majority (7 out of 8) reported that students worked hard while playing the games. Five out

Table 6
Cognitive and Academic Scores by Group and Time Point

	CSI group				*Control group*				*Time 1 group difference*		*Time 2 group difference*	
	Time 1		*Time 2*		*Time 1*		*Time 2*					
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*t*	*p*	*t*	*p*
Cognitive composite	−0.53	0.42	0.42	0.51	−0.31	0.42	0.42	0.60	1.75	.09	0.02	.98
Math	−0.44	0.77	0.34	0.84	−0.28	1.16	0.36	0.94	0.53	.60	0.08	.94
ELA	0.07	1.07	−0.35	0.90	0.29	1.11	−0.06	0.85	0.65	.52	1.05	.30

Note. Scores were standardized for each measure across groups and time points. CSI: $n = 22$ (with test score data $n = 19$). Control: $n = 24$ (with test score data $n = 23$).

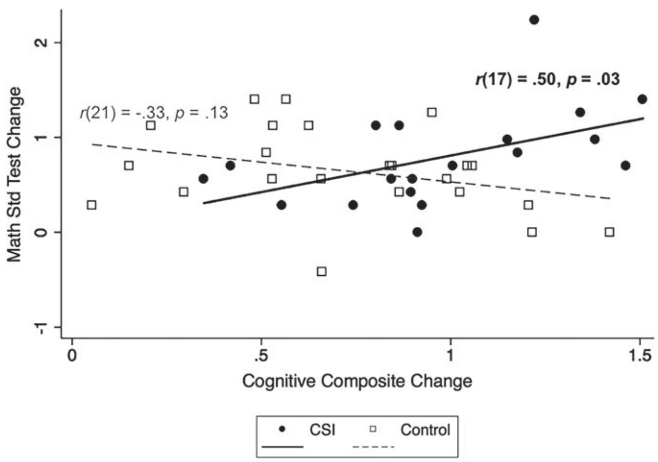

Fig. 3. Cognitive change correlates with math standardized test score change in the CSI group, but not in the control group.

of eight reported that students seemed excited to play the games. Four out of eight reported that they thought playing games were a good use of time in the classroom. Challenges cited included student frustration, boredom, and lack of attention/focus.

DISCUSSION

Students in the CSI group improved significantly more than the control group on a composite measure of cognitive skills. Both the CSI and control groups improved substantially on the composite measure over the course of the school year, likely due to a combination of practice effects, cognitive development, and other changes associated with school wide instructional strategies, which included activities such as aerobic exercise (Hillman et al., 2014) and mindfulness meditation (Schonert-Reichl et al., 2015; Tang, Hölzel, & Posner, 2015; Zenner, Herrnleben-Kurz, & Walach, 2014). Another possible contributor to the large cognitive gains across the school could have been the initial school wide professional development session in which we discussed how to support cognitive skills during academic instruction.

Students with lower academic standardized test scores from the previous year showed a greater cognitive benefit from CSI, a finding that is concordant with prior research (Mackey et al., 2011). In the prior study, in which students showed substantial cognitive benefits from training, students had test scores in the 20th percentile statewide in California. In the present study, students had test scores, on average, in the 70th percentile statewide in Massachusetts. The students with lower test scores (~50th percentile) showed benefits from CSI, but students with higher test scores (~85th percentile) did not.

The finding that cognitive training was more effective for students with lower scores on tests of academic achievement adds to the growing body of evidence that lower performing students stand to gain more from cognitive interventions (Diamond & Ling, 2016; Titz & Karbach, 2014). Studies that involve a broad range of students may not find any apparent overall academic benefit from cognitive training if students with average or above-average test scores show little or no benefit. The nature of the cognitive training may also be important. In the present study, training was deliberately broad across multiple cognitive skills. A prior study that trained only working memory failed to show academic benefits among students with low working memory (Dunning et al., 2013). Thus, a broad-spectrum cognitive training program targeted at students with low scores on tests of academic achievement is, at present, the most promising path forward.

Overall, students at the charter school showed a large improvement in Math standardized test scores from the prior academic year, but no improvement in ELA. Large gains in Math and small gains in ELA are typical of Boston-area charter schools (Angrist, Cohodes, Dynarski, Pathak, & Walters, 2013). The students who participated in CSI did not improve significantly more in Math or ELA than the students who received the control curriculum. As one of the central goals of cognitive training should be to improve educational performance, the lack of a group difference in math and ELA test score changes was a disappointing result. It indicates that the CSI program in its current implementation is not sufficient to improve standardized test scores for all students within the school year of the intervention. However, within the CSI group, and not within the control group, cognitive improvement was positively correlated with math test score improvement, indicating that students whose cognitive skills improved the most from CSI also showed the greatest math gains. This result is consistent with some degree of transfer to an important education measure. Variability in cognitive and academic benefits could be related to variability in engagement with and enjoyment of the games, as has been found previously (Jaeggi, Buschkuehl, Jonides, & Shah, 2011). It is also possible that greater transfer may become evident months or years after the intervention if cognitive skills serve as a scaffold for learning. Indeed, prior studies have found delayed benefits to cognitive training (Blair & Raver, 2014; Holmes et al., 2009).

The assessments in this study were administered in a classroom setting, which may have added noise to the data due to nonstandard administration (paper-and-pencil versions, added time limits) and distractions in the testing environment. Classroom testing was especially problematic for the working memory measure, which did not show expected patterns of correlations with other measures. Advances in web-based cognitive testing, along with improvements in classroom technology, will be necessary to make better controlled testing feasible on a large scale.

A surprising finding from this study was the feedback from tutors. Only half thought the games were a good use of time, and many complained about challenges in managing student behavior (e.g., students got too excited and got out of their chairs). It was difficult to both to make the games fun and to keep a controlled classroom environment. Further, teaching cognitive skills differs from instilling crystallized knowledge in that there is no explicit feedback. It is more difficult to determine, in real time, whether or not students are learning. Tutors likely would have benefited from more professional development time dedicated to CSI. This program's reliance on intensive professional development suggests that it may not scale well, especially in traditional schools with greater student-to-teacher ratios. Computerization of CSI could address this issue, but programs would need to be structured differently from prior software-based cognitive training attempts to achieve transfer to academic performance. It is unknown whether these programs would be more effective if they included key features of CSI (social interaction with peers and instructors, individualized difficulty progression, varied tasks and domains, emphasis on metacognition). A major advantage of a software-based approach is that it could integrate frequent assessments to provide feedback to students and teachers.

These promising initial results warrant follow-up studies with more students, and longer term academic data. Additional work is needed to fine tune both the activities involved in CSI, and the classroom implementation. The long-term goal of this research is to create a program that can be independently implemented by schools that improves both cognitive skills and academics. This program will likely vary substantially by student age and developmental stage. In the present research, we have conceptualized this program as a brief stand alone module. Although our program contains some numerical reasoning, it is largely free of standard academic content. A potentially interesting alternative is to develop curricula that integrate cognitive skill practice into academics, for example, by combining working memory and reasoning training with history instruction (Ariës, Groot, & van den Brink, 2014). Another more intensive approach is to redesign the entire school day to integrate cognitive skill support with learning goals. This approach has been taken in preschool and kindergarten classrooms (Blair & Raver, 2014; Diamond et al., 2007; Hermida et al., 2015; Lillard & Else-Quest, 2006), but could be effective in older grades as well.

Whatever the format, the development and evaluation of cognition-focused instruction must be done with great care. Classroom time is precious and should not be spent on programs that produce narrow benefits on cognitive tests in the absence of demonstrable educational benefits. More research is needed to understand not only the average effects of cognitive interventions, but also individual differences in responsivity to such programs. Ultimately, it may be possible to use cognitive, academic, and neuroimaging measures to predict benefits from particular interventions and to tailor interventions to individual students (Gabrieli, Ghosh, & Whitfield-Gabrieli, 2015).

Acknowledgments—Authors are grateful for the close collaboration with Match Education that made this work possible. In particular, they would like to thank Ryan Holmes for his assistance with curriculum design, tutor training, and study implementation. Authors would also like to thank Andrew Jeong and Ray Schleck for helping with study implementation and obtaining academic data. Finally, they would like to thank Kelly Halverson for guidance in choosing and administering study assessments.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Table S1. Change Scores on Each Measure, by Group. All Measures Were *z*-Scored Across Groups and Time Points

Table S2. Assessment Scores by Group and Time Point. Scores Were Standardized for Each Measure Across Groups and Time Points

Table S3. Change Scores on Each Measure for Students With Lower Entering Standardized Test Scores

Table S4. Change Scores on Each Measure for Students With Higher Entering Standardized Test Scores

Table S5. Cognitive and Academic Change Correlations by Curriculum Group and Test Score Group. No Correlations Are Significant

REFERENCES

Angrist, J. D., Cohodes, S. R., Dynarski, S. M., Pathak, P. A., & Walters, C. D. (2013). *Charter schools and the road to college readiness: The effects on college preparation, attendance and choice*. Retrieved from http://users.nber.org/~dynarski/Charters_and_College_Readiness.pdf

Ariës, R. J., Groot, W., & van den Brink, H. M. (2014). Improving reasoning skills in secondary history education by working memory training. *British Educational Research Journal*, *41*(2), 210–228. doi:10.1002/berj.3142

Blair, C., & Raver, C. C. (2014). Closing the achievement gap through modification of neurocognitive and neuroendocrine function: Results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten. *PLoS One*, *9*(11), e112393

Blakey, E., & Carroll, D. J. (2015). A short executive function training program improves preschoolers' working memory. *Frontiers in Psychology*, *6*(Nov), 1–8. doi:10.3389/fpsyg.2015.01827

Bodrova, E., & Leong, D. (2003). The importance of being playful. *Educational Leadership*, *60*(7), 50–53.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *TONI-4, Test of Nonverbal Intelligence*. (4th ed.) Austin, TX: PRO-ED.

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, *13*(42), n42.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, *318*(5855), 1387–1388. doi:10.1126/science.1151148

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*(6045), 959–964.

Diamond, A., & Ling, D. S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience*, *18*, 34–48. doi:10.1016/j.dcn.2015.11.005

Dunning, D. L., Holmes, J., & Gathercole, S. E. (2013). Does working memory training lead to generalized improvements in children with low working memory? A randomized controlled trial. *Developmental Science*, *16*, 915–925. doi:10.1111/desc.12068

Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., … Gabrieli, J. D. E. (2014). Cognitive skills, student achievement tests, and schools. *Psychological Science*, *25*, 736–744. doi:10.1177/0956797613516008

Gabrieli, J. D. E., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, *85*, 11–26. doi:10.1016/j.neuron.2014.10.047

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1–16. doi:10.1002/acp.934

Goldin, A. P., Hermida, M. J., Shalom, D. E., Elias Costa, M., Lopez-Rosenfeld, M., Segretin, M. S., … Sigman, M. (2014). Far transfer to language and math of a short software-based gaming intervention. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(17), 6443–6448. doi:10.1073/pnas.1320217111

Green, C. S., & Bavelier, D. (2008). Exercising your brain: A review of human brain plasticity and training-induced learning. *Psychology and Aging*, *23*, 692–701.

Hermida, M. J., Segretin, M. S., Prats, L. M., Fracchia, C. S., Colombo, J. A., & Lipina, S. J. (2015). Cognitive neuroscience, developmental psychology, and education: Interdisciplinary development of an intervention for low socioeconomic status kindergarten children. *Trends in Neuroscience and Education*, *4*(1), 15–25. doi:10.1016/j.tine.2015.03.003

Hillman, C. H., Pontifex, M. B., Castelli, D. M., Khan, N. A., Raine, L. B., Scudder, M. R., … Kamijo, K. (2014). Effects of the FITKids randomized controlled trial on executive control and brain function. *Pediatrics*, *134*(4), e1063–e1071. doi:10.1542/peds.2013-3219

Hirsh-Pasek, K., Golinkoff, R. M., Berk, L. E., & Singer, D. G. (2009). *A mandate for playful learning in preschool: Presenting the evidence*. New York, NY: Oxford University Press.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9–F15.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–6833. doi:10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(25), 10081–10086. doi:10.1073/pnas.1103228108

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, *12*, 978–990.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY: A developmental neuropsychological assessment*. San Antonio, TX: Psychological Corporation.

Lillard, A., & Else-Quest, N. (2006). The early years: Evaluating Montessori education. *Science*, *313*(5795), 1893–1894. doi:10.1126/science.1132362

Loosli, S. V., Buschkuehl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology*, *18*(1), 62–78. doi:10.1080/09297049.2011.575772

Mackey, A. P., Hill, S. S., Stone, S. I., & Bunge, S. A. (2011). Differential effects of reasoning and speed training in children. *Developmental Science*, *14*, 582–590. doi:10.1111/j.1467-7687.2010.01005.x

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270–291. doi:10.1037/a0028228

Miller, D. J., & Robertson, D. P. (2011). Educational benefits of using game consoles in a primary classroom: A randomised controlled trial. *British Journal of Educational Technology*, *42*, 850–864.

Roberts, G., Quach, J., Spencer-Smith, M., Anderson, P. J., Gathercole, S., Gold, L., … Wake, M. (2016). Academic outcomes 2 years after working memory training for children with low working memory: A randomized clinical trial. *JAMA Pediatrics*, *170*(5), e154568.

Rode, C., Robson, R., Purviance, A., Geary, D. C., & Mayr, U. (2014). Is working memory training effective? A study in a school setting. *PLoS One*, *9*(8), 1–8. doi:10.1371/journal.pone.0104796

Schonert-Reichl, K. A., Oberle, E., Lawlor, M. S., Abbott, D., Thomson, K., Oberlander, T. F., & Diamond, A. (2015). Enhancing cognitive and social-emotional development through a simple-to-administer mindfulness-based school program for elementary school children: A randomized controlled trial. *Developmental Psychology*, *51*, 52–66.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, *17*(3), 103–186.

St Clair-Thompson, H. S., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology Review*, *30*(2), 203–219.

Tang, Y.-Y., Hölzel, B. K., & Posner, M. I. (2015). The neuroscience of mindfulness meditation. *Nature Reviews Neuroscience*, *16*(4), 1–13. doi:10.1038/nrn3916

Titz, C., & Karbach, J. (2014). Working memory and executive functions: Effects of training on academic achievement. *Psychological Research*, *78*, 852–868. doi:10.1007/s00426-013-0537-1

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children–Fourth edition (WISC-IV)*. San Antonio, TX: The Psychological
Corporation.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III Tests of Cognitive Abilities. In J. A. Naglieri & S. Goldstein (Eds.), *Update* (pp. 371–401). Itasca, IL: Riverside Publishing.

Zenner, C., Herrnleben-Kurz, S., & Walach, H. (2014). Mindfulness-based interventions in schools—A systematic review and meta-analysis. *Frontiers in Psychology*, *5*, 603. doi:10.3389/fpsyg.2014.00603