



# WWC Single Study Review

A review of the design and summary of findings for an individual study



August 2015

## WWC Review of the Report “Effects of the FITKids Randomized Controlled Trial on Executive Control and Brain Function”<sup>1</sup>

The findings from this review do not reflect the full body of research evidence on *Fitness Improves Thinking in Kids (FITKids)*.

### What is this study about?

The study measured the impact of the *Fitness Improves Thinking in Kids (FITKids)* afterschool program on the executive control (i.e., maintaining focus, performing multiple cognitive processes) and physical fitness of preadolescent students. The study was conducted between 2009 and 2013. The *FITKids* program was held at a recreational facility on the University of Illinois’ campus and included 2 hours of activities after each school day. Each 2-hour session of the program included three components: a fitness component, a rest period with an educational component, and game play.

Study authors randomly assigned 221 students ages 7–9 to either participate in *FITKids* or to be in a business-as-usual comparison group. Students in the intervention group participated in *FITKids* after school, while students in the comparison group did not participate in the program.

Executive control was measured before and after the *FITKids* program by assessing students’ response accuracy and response time on two tasks. The first task (attentional inhibition; also known as the flanker task) required students to resist distracting information. In this task, students were repeatedly shown an array of fish on a computer screen

and asked to identify whether the middle fish faced left or right. The second task (cognitive flexibility; also known as the switch task) required students to perform multiple cognitive duties at the same time. In this task, students were repeatedly shown a blue or green circle or square on a computer screen and asked to determine either the shape or the color. The easier portion of this task (homogeneous trials) included constant instruction to identify either the shape or the color. The more difficult portion of this task (heterogeneous trials) included changing instructions to identify the shape in some instances or the color in others.

Study authors also measured fitness levels of students in *FITKids* and the comparison group both before and after the intervention. Fitness levels were measured with body mass index (BMI) and maximal oxygen consumption during aerobic activity ( $VO_{2peak}$ ).

### WWC Rating

***The research described in this report meets WWC group design standards without reservations***

The study is a randomized controlled trial with low attrition.

### **What did the study find?**

The study authors found that the *FITKids* program increased accuracy in the attentional inhibition task and accuracy in the more difficult portion of the cognitive flexibility task. However, the WWC did not confirm this finding to be statistically significant after adjusting for multiple comparisons. The study also showed a statistically significant positive impact for one subscale of the attentional inhibition task, and the WWC confirmed this finding to be statistically significant. Additionally, the study authors found that *FITKids* had a statistically significant positive effect on aerobic fitness. Moreover, BMI for students in *FITKids* increased by a smaller amount than for students not in the program, and this result was statistically significant. The WWC confirmed the findings for aerobic fitness and BMI.

### **Features of *Fitness Improves Thinking in Kids (FitKids)***

The *FITKids* afterschool program is designed to increase physical fitness through a variety of activities. For this study, the program was run for 2 hours per day after school, for 150 days of the school year. The *FITKids* program included:

- a 30-minute fitness component, where students engaged in a series of 4–6 activities (e.g., cardiovascular endurance, muscular strength) at activity stations;
- a rest period, which included a healthy snack and nutrition education; and
- a 45- to 55-minute session focused on low-organizational games (e.g., tag, active rock-paper-scissors challenge)

Overall, an average of 70 minutes of moderate-to-vigorous physical activity was included in each afterschool session.

**Appendix A: Study details**

Hillman, C. H., Pontifex, M. B., Castelli, D. M., Khan, N. A., Raine, L. B., Scudder, M. R., Drollette, E. S., Moore, R. D., Wu, C-T, & Kamijo, K. (2014). Effects of the *FITKids* randomized controlled trial on executive control and brain function. *Pediatrics*, *134*(4), 1062–1071.

**Setting** The study was conducted at a recreational facility on the University of Illinois at Urbana-Champaign’s campus between 2009 and 2013.

**Study sample** In order to be eligible for the study, students ages 7–9 who lived in East Central Illinois must have (a) not been receiving special educational services related to cognitive or attentional disorders, neurologic diseases, or physical disabilities and (b) consented to be in the study (parent consent and student assent were both required). Of the 475 students screened, 221 were randomly assigned to either receive *FITKids* or to be in a business-as-usual comparison group. For the randomization procedure, pairs of students were first matched on age, gender, race, socioeconomic status, and maximal oxygen consumption ( $VO_{2max}$ ). A coin was flipped to assign one student in each pair to the intervention group and one to the comparison group. Siblings from 10 families also participated in the *FITKids* program, and analyses were re-run using one randomly selected sibling from each of the 10 families.

At pretest, students averaged 8.8 years of age in both the intervention and comparison groups. Over half of participating students (51% intervention, 59% comparison) were White, about a quarter of the study sample (25% intervention, 28% comparison) were African American, and more than one in ten (17% intervention, 12% comparison) were Asian. Slightly less than half of the students in both the intervention group (43%) and comparison group (49%) were categorized as having low socioeconomic status, which was based on an index of three factors: participation in free or reduced-price meals at school, highest educational level attained by the student’s mother and father, and the number of parents who worked full time.

The follow-up sample included 105 students in the intervention group and 101 students in the comparison group who participated in assessments at the end of the spring semester. Missing values were imputed in the analysis using multiple imputation methods, which the WWC considers to be acceptable for the reporting of study findings, since this randomized controlled trial had low attrition.

**Intervention group** Students in the intervention group participated in the *FITKids* afterschool program. The *FITKids* program was offered for 2 hours after each school day for 150 days of the 170-day school year. Each daily lesson included 30 minutes of 4–6 “instant activities” at activity stations, which focused on aerobic activities, muscular strength and endurance, and movements. Instant activities were followed by a snack and a brief educational component centered on a weekly theme involving nutrition, fitness, or benefits of physical activity. The final portion of each lesson involved 45–55 minutes of low-organizational games (e.g., tag, active rock-paper-scissors challenge) centered on a skill theme. The average attendance rate for sessions of the *FITKids* program was 80.6%. Students averaged 70 minutes of moderate-to-vigorous physical activity during each session. During each session, students took an average of 4,246 steps and had a mean heart rate of 137 beats per minute.

**Comparison group**

Students in the comparison group were randomly assigned to not receive *FITKids*; these students were placed on a wait list.

**Outcomes and measurement**

Core outcomes for the study included executive control and fitness. Executive control was measured with flanker and switch tasks. The flanker task was designed to measure inhibitory control (resisting distractions or habits to maintain focus), and the switch task was designed to assess working memory and cognitive flexibility. Fitness was assessed by measuring each student's BMI and maximal oxygen consumption ( $VO_{2\text{peak}}$ ) during strenuous exercise on a treadmill. The pretest for each assessment was administered prior to the start of *FITKids* during a 2-day testing period, and the posttest was administered following completion of the *FITKids* program for a given school year. Posttest assessments were identical to the pretest assessments. For a more detailed description of these outcome measures, see Appendix B.

**Support for implementation**

The randomization procedure was conducted by a staff member who was not involved in the data collection. No information on staff training for this intervention was provided.

**Reason for review**

This study was identified for review by receiving media attention.

**Appendix B: Outcome measures for each domain**

<b>Executive control</b>	
<i>Flanker task</i>	The flanker task was designed to measure attentional inhibition. Students were asked to identify the direction in which a centrally targeted object (a goldfish) pointed among a group of flanking objects (i.e., goldfish pointing in either congruous or incongruous directions). Students were given a block of 40 practice trials before the assessment began, which included two blocks of 75 trials each. For a given school year, the pretest was administered prior to the start of <i>FITKids</i> during a 2-day testing period (on Day 2), and the posttest was administered following completion of the <i>FITKids</i> program.
<i>Switch task</i>	The switch task was designed to measure working memory and cognitive flexibility. Students were asked to press a response pad with their left thumb when a character on the screen was blue or a circle, and to press the response pad with their right thumb when the object was green or a square. This task also included heterogeneous trials where students performed the color and shape tasks in combination with a specific task indicated by the character's arms (i.e., arms up required a response based on the shape of the character, and arms down required a response on the color of the character). Students were given a block of 40 practice trials before the assessment began, which included two blocks of 60 homogeneous trials and three blocks of 50 heterogeneous trials. For a given school year, the pretest was administered prior to the start of <i>FITKids</i> during a 2-day testing period (on Day 2), and the posttest was administered following completion of the <i>FITKids</i> program.
<b>Fitness</b>	
<i>Body mass index (BMI)</i>	Each student's height and weight was measured at the beginning and end of the intervention and converted into a BMI. For a given school year, the pretest was administered prior to the start of <i>FITKids</i> during a 2-day testing period (on Day 1), and the posttest was administered following completion of the <i>FITKids</i> program.
<i>Aerobic fitness—maximal oxygen consumption (<math>VO_{2peak}</math>)</i>	Aerobic fitness was measured using a computerized indirect calorimetry system while students ran or walked on a treadmill. The treadmill was kept at a constant speed, and the incline was increased by 2.5% every 2 minutes until the student was no longer able to maintain a consistent level of intensity. Maximal oxygen consumption ( $VO_{2peak}$ ) was established when a student (1) had a heart rate of more than 185 beats per minute and the heart rate plateaued; (2) the respiratory exchange rate (RER) was above 1.0; (3) the student scored above 7 on the OMNI ratings of perceived exertion scale; or (4) the student plateaued in oxygen consumption corresponding to an increase of less than 2 mL/kg per minute. For a given school year, the pretest was administered prior to the start of <i>FITKids</i> during a 2-day testing period (on Day 1), and the posttest was administered following completion of the <i>FITKids</i> program.

**WWC Single Study Review**

**Appendix C: Study findings for each domain**

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Executive control</b>								
<i>Flanker task, response accuracy</i>	Full sample	221 students	84.9% (4.7)	82.7% (10.8)	+3.2%	0.41	+16	.05
<i>Flanker task, response time (ms)</i>	Full sample	221 students	503.3 (86.8)	492.5 (96.1)	+10.8	-0.12	-5	> .05
<i>Switch task, response accuracy for homogeneous trials</i>	Full sample	221 students	90.0% (6.0)	90.2% (5.6)	-0.2%	-0.04	-1	> .05
<i>Switch task, response accuracy for heterogeneous trials</i>	Full sample	221 students	79.7% (12.0)	75.0% (12.3)	+4.7%	0.40	+16	.01
<i>Switch task, response time for homogeneous trials (ms)</i>	Full sample	221 students	779.4 (176.4)	759.5 (148.4)	+19.9	-0.13	-5	> .05
<i>Switch task, response time for heterogeneous trials (ms)</i>	Full sample	221 students	1,497.1 (234.0)	1435.1 (237.4)	+62.0	-0.27	-11	> .05
<b>Domain average for executive control</b>						<b>0.04</b>	<b>+2</b>	<b>Not statistically significant</b>
<b>Fitness</b>								
<i>Body mass index (BMI)</i>	Full sample	221 students	19.1 (4.7)	19.8 (4.6)	-0.7	0.16	+6	.00
<i>Aerobic fitness (VO<sub>2peak</sub>, mL/kg/min)</i>	Full sample	221 students	41.2 (6.8)	39.9 (6.9)	+1.3	0.20	+8	.01
<b>Domain average for fitness</b>						<b>0.18</b>	<b>+7</b>	<b>Statistically significant</b>

**Table Notes:** For effect size and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding.

**Study Notes:** The p-values presented here were reported in the original study, and standard deviations were derived from 95% confidence intervals presented in the original study. A correction for multiple comparisons was needed and resulted in a WWC-computed critical p-value of .008 for response accuracy in the heterogeneous trials of the switch task and .017 for response accuracy in the flanker task; however, there are only two statistically significant measures in an outcome domain with six measures. Therefore, the WWC does not find the results for this outcome domain to be statistically significant. The WWC calculated the program group mean using a difference-in-differences approach (see WWC Handbook) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group post-test means. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information. This study is characterized as having an indeterminate effect on executive control because the mean effect reported is neither statistically significant nor substantively important. The study is characterized as having a statistically significant positive effect on fitness because at least one measure is positive and statistically significant and no effects are negative and statistically significant, accounting for multiple comparisons. For more information, please refer to the WWC Standards and Procedures Handbook (version 3.0), pp. 25–26.

**WWC Single Study Review**

**Appendix D: Supplemental findings for the executive control domain (flanker task)**

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Executive control</b>								
<i>Flanker task, congruent trials, response accuracy</i>	Full sample	221 students	89.5% (9.1)	85.6% (10.5)	+3.9%	0.41	+16	< .05
<i>Flanker task, incongruent trials, response accuracy</i>	Full sample	221 students	82.4% (10.7)	79.9% (12.0)	+2.5%	0.22	+9	> .05
<i>Flanker task, congruent trials, response time (ms)</i>	Full sample	221 students	485.4 (84.7)	478.1 (95.4)	+7.3	-0.08	-3	> .05
<i>Flanker task, incongruent trials, response time (ms)</i>	Full sample	221 students	522.8 (90.7)	508.2 (98.7)	+14.6	-0.16	-6	> .05

**Table Notes:** The supplemental findings presented in this table are additional findings that do not factor into the determination of the evidence rating. For effect size and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding.

**Study Notes:** The p-values presented here were reported in the original study, and standard deviations were derived from 95% confidence intervals presented in the original study. The WWC calculated the program group mean using a difference-in-differences approach (see WWC Handbook) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information.

## **Endnotes**

<sup>1</sup> Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the authors) to assess whether the study design meets WWC group design standards. The review reports the WWC's assessment of whether the study meets WWC group design standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the single study review protocol, version 2.0. A quick review of this study was released in November 2014, and this report is the follow-up review that replaces that initial assessment. The reported analyses in this SSR are only for those eligible outcomes that either met WWC group design standards without reservations or met WWC group design standards with reservations, and do not necessarily apply to all results presented in the study.

## **Recommended Citation**

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2015, August). WWC review of the report: *Effects of the FITKids randomized controlled trial on executive control and brain function*. Retrieved from <http://whatworks.ed.gov>



**Glossary of Terms**

- Attrition** Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.
- Clustering adjustment** If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.
- Confounding factor** A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.
- Design** The design of a study is the method by which intervention and comparison groups were assigned.
- Domain** A domain is a group of closely related outcomes.
- Effect size** The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.
- Eligibility** A study is eligible for review if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.
- Equivalence** A demonstration that the analytic sample groups are similar on observed characteristics defined in the review area protocol.
- Improvement index** Along a percentile distribution of individuals, the improvement index represents the gain or loss of the average individual due to the intervention. As the average individual starts at the 50th percentile, the measure ranges from -50 to +50.
- Multiple comparison adjustment** When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.
- Quasi-experimental design (QED)** A quasi-experimental design (QED) is a research design in which study participants are assigned to intervention and comparison groups through a process that is not random.
- Randomized controlled trial (RCT)** A randomized controlled trial (RCT) is an experiment in which eligible study participants are randomly assigned to intervention and comparison groups.
- Single-case design (SCD)** A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.
- Standard deviation** The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample are spread out over a large range of values.
- Statistical significance** Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ( $p < .05$ ).
- Substantively important** A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the [WWC Procedures and Standards Handbook \(version 3.0\)](#) for additional details.



Intervention  
Report



Practice  
Guide



Quick  
Review



Single Study  
Review

A **single study review** of an individual study includes the WWC's assessment of the quality of the research design and technical details about the study's design and findings.

This single study review was prepared for the WWC by Mathematica Policy Research under contract ED-IES-13-C-0010.