

SPARK Evaluation

Fil Babalievsky and Atishay Sehgal

8/2/2018

Background

- ▶ Dr. John Ratey: SPARK book, compiled evidence that aerobic exercise improves cognition
- ▶ Implemented in Naperville High for academically at-risk children, promising results but not randomized
- ▶ Positive results from an RCT with obese young (7-11 year old) children in Georgia, bigger effects linked to longer exercise
- ▶ Borough Hall: Closer to a true randomized design than Naperville, very different population than Georgia, not pre-selected based on how likely they were to benefit

Implementation

- ▶ One High School split a group of <100 ninth graders into 3 gym classes
- ▶ Arbitrarily split by Gym class, not based on systematic differences
- ▶ One quarter where all had same curriculum
- ▶ Aerobic exercise (heart rate near max) along with sports
- ▶ Implementation limited by state curriculum
- ▶ 3 days per week, 7 minutes in first quarter and 20 thereafter

Empirical design

Differences in Differences: Do SPARK kids' scores grow more from Marking Period 1 to Marking Periods 2, 3, 4? Is this different in different classes?

- ▶ First:

$$\Delta_{i,s,m} = \alpha_m + \beta_m \cdot SPARK_i$$

- ▶ Second: Split by classes

$$\Delta_{i,s,m} = \sum_j \alpha_{j,m} + \sum_j \beta_{j,m} \cdot SPARK_{i,j}$$

Results: Simple Design

How to read this table: a score of 2.3 in the first column and row means that SPARK kids' scores improved by 2.3 points relative to non-SPARK kids from MP1 to MP2. A score of 0.5 in the second column means that SPARK scores went up by 0.5 points from MP1 to MP3 (it is not cumulative.)

Table 1:

	<i>Dependent variable:</i>		
	MP1 to MP2 (1)	MP1 to MP3 (2)	MP1 to MP4 (3)
Spark	2.391** (0.932)	0.547 (1.248)	1.442 (0.983)
Constant	-6.385*** (0.733)	-8.835*** (0.973)	-6.029*** (0.767)
Observations	273	263	263
R ²	0.024	0.001	0.008
Adjusted R ²	0.020	-0.003	0.004
Residual Std. Error	7.478 (df = 271)	9.878 (df = 261)	7.781 (df = 261)
F Statistic	6.579** (df = 1; 271)	0.192 (df = 1; 261)	2.151 (df = 1; 261)

Note:

*p<0.1; **p<0.05; ***p<0.01

Results: Class effects

Table 2:

	<i>Dependent variable:</i>		
	MP1 to MP2	MP1 to MP3	MP1 to MP4
	(1)	(2)	(3)
Spark_math	3.232** (1.612)	0.278 (2.025)	2.857* (1.703)
Spark_sci	2.907* (1.603)	0.239 (1.996)	0.089 (1.679)
Spark_eng	0.986 (1.603)	1.396 (1.996)	1.456 (1.679)
math	2.380 (1.792)	-8.962*** (2.224)	-3.965** (1.870)
eng	2.743 (1.779)	-7.629*** (2.191)	-3.914** (1.843)
Constant	-8.086*** (1.258)	-3.371** (1.549)	-3.429*** (1.303)
Observations	273	263	263
R ²	0.048	0.153	0.041
Adjusted R ²	0.030	0.136	0.023
Residual Std. Error	7.441 (df = 267)	9.165 (df = 257)	7.708 (df = 257)
F Statistic	2.675** (df = 5; 267)	9.282*** (df = 5; 257)	2.225* (df = 5; 257)

Note:

* p<0.1; ** p<0.05; *** p<0.01

Interpretation

- ▶ Coefficients are points out of 100, so the coefficient of 2.8 on the 4th quarter math results means that SPARK kids had >0.28 grade levels of growth compared to non-SPARK
- ▶ Not huge and not as consistent as we would like, but still interesting
- ▶ Generally larger for math, in line with prior studies
- ▶ Hints that Georgia results could scale and that Naperville results could hold

Going forward

Ideally:

- ▶ Large scale RCT (maybe randomly assign all SI High School gym classes to SPARK or non-SPARK)
- ▶ Get as much leeway as possible from State Education Dept
- ▶ Test for heterogeneous effects: maybe measure baseline academic performance and fitness from kids, test if effects are bigger for less fit or lower scoring kids
- ▶ Pre register the study for greater credibility and visibility

Possible Interest

- ▶ The prospect of running an RCT with thousands of students ought to make academics salivate
- ▶ We have offered some suggestions for how to run a larger study but it would be good to bring in full time researchers who specialize in this