SPARK Evaluation

Fil Babalievsky and Atishay Sehgal

8/2/2018

Background

- Dr. John Ratey: SPARK book, compiled evidence
- Implemented in Naperville High, promising results but not randomized
- Borough Hall: Closer to a true randomized design

Implementation

- One High School split a group of $<$100 ninth graders into 3
- Arbitrarily split by Gym class, no systematic differences
- Aerobic exercise (heart rate near max) rather than sports

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

Table 1:

	Dependent variable:		
	MP1 to MP2	MP1 to MP3	MP1 to MP4
	(1)	(2)	(3)
Spark	2.391**	0.547	1.442
	(0.932)	(1.248)	(0.983)
Constant	−6.385***	−8.835***	−6.029***
	(0.733)	(0.973)	(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:

	Dependent variable:		
	MP1 to MP2	MP1 to MP3	MP1 to MP4
	(1)	(2)	(3)
Spark_math	3.232**	0.278	2.857*
	(1.612)	(2.025)	(1.703)
Spark_sci	2.907*	0.239	0.089
	(1.603)	(1.996)	(1.679)
Spark_eng	0.986	1.396	1.456
	(1.603)	(1.996)	(1.679)
math	2.380	−8.962***	−3.965**
	(1.792)	(2.224)	(1.870)
eng	2.743	−7.629***	−3.914**
	(1.779)	(2.191)	(1.843)
Constant	−8.086***	−3.371**	−3.429***
	(1.258)	(1.549)	(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

- Not huge, but still interesting
- Generally larger for math, in line with prior studies

Going forward

Ideally:

- Large scale RCT (randomly assign all SI High School gym classes to SPARK or not)
- Heterogeneous effects (measure baseline academic performance and fitness from kids, test if effects are bigger for less fit or lower scoring kids)
- Pre register the study