Spark Evaluation

*Fil Babalievsky**

May 2019

Abstract

We evaluate the effect of Project SPARK, an effort to introduce aerobic exercise to Staten Island's K-12 PE classes, on academic outcomes. We find some evidence that it may have had a positive impact, and recommend further steps.

Acknowledgement

We want to thank Principal Deirdre DeAngelis, Assistant Principal Richard Rucireto, and Julie Gambale-Kubiak at New Dorp High school. This pilot would not have been possible without their dedicated collaboration.

Introduction:

The Staten Island Borough President's office has conducted a preliminary investigation into the effects of aerobic exercise on academic achievement, in consultation with Dr. John Ratey. This program was inspired by and named for his book "SPARK"[1], an investigation of the benefits of exercise. This was a small-scale pilot program intended as a proof of concept, yet it may have had a small positive effect.

Similar studies

Dr. Ratey helped Naperville High implement a pilot scale study of the effects of exercise on at-risk children and found positive results, although the study was small and not randomized.

One other promising piece of evidence comes from a small randomized trial in Augusta, Georgia conducted by Tomporowski et al (2011). It found that obese young children randomly assigned to different levels of intense exercise did better on standardized tests and assessments of executive functioning, and that their gains scaled with the duration of the exercise.

Our study is not an ideal RCT but it targets a different group of students than the Georgia study, which helps us learn more about how effects might generalize. In particular, our study was not limited to at-risk or obese children.

Program details:

New Dorp High School[2], one of Staten Island's public schools, agreed to serve as a small scale pilot for SPARK. School administrators selected a group of just under one hundred ninth graders and, starting in the

*Babalievsky worked as the Data Fellow at the Staten Island Borough President's Office in the Summers of 2017 and 2018. He has a Bachelor's in Economics from Yale and is currently a second-year Economics Ph.D at the University of Minnesota. Atishay Sehgal provided significant assistance. Sehgal has an MA in Statistics from Columbia and worked as a Data Fellow in the Summer and Fall of 2018.

[1] Ratey, John J., Hagerman, Eric. Spark: The Revolutionary New Science Of Exercise And The Brain. New York : Little, Brown, 2008. Print.

[2] https://www.newdorphs.org/apps/pages/index.jsp?uREC_ID=249165&type=d&pREC_ID=573908

second quarter of the school year, divided them into two groups. The students in two physical education classes, comprising two thirds of the group, supplemented their standard sports-based curriculum with intense aerobic exercise. At the start of the program, this meant seven minutes of intense cardio , eventually expanded to fifteen, for three days per school week. Observers reported that, even in the early phase of the program, students were visibly exhausted. Note that both treatment and control groups of students had the same physical education curriculum in the first quarter, providing a useful baseline.

The students were still subject to the New York State physical education curriculum, with an emphasis on sports-based skill development rather than sustained aerobic activity[3]. The implementation of SPARK was, therefore, not as extensive as any of the partners involved would have wanted.

In fact, a typical physical education curriculum nationally is centered around sports-based skill development—without any specific requirements on getting students' heart rate up. Schools implementing SPARK replaced the sports-based games that students typically play with intense workouts designed to keep them at a high percentage of their maximum heart rates which we replicated for the first seven to fifteen minutes of each class in our pilot at New Dorp.

The students were split based on logistical contraints. Two gym classes were arbitrarily chosen and the third was left out as a control, but the school officials who separated the students did not do so based on any systematic difference between the students.

Research Design and Data:

We were given data on student performance in math, English, and science classes for all ninth graders in New Dorp. Our main specification consisted of a differences in differences approach. The first quarter, when SPARK pilot students in all three gym classes had the same curriculum, provideed a baseline for comparison. We tested if student academic growth differed between treatment and control over the remaining marking periods.

Our first research design compared the difference in grades between the first and subsequent marking periods for SPARK and non SPARK students, without controls for type of class. We ran three regressions in this format, with the three dependent variables being the change in grades between the first marking period and the three subsequent marking periods.

The regression design is as follows:

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

Here $\Delta_{i,s,m}$ was the change in score for student i in subject s from marking period 1 to marking period m, and $SPARK_i$ is a dummy taking value 1 if and only if the student was in SPARK.

Our second research design considered the possibility that SPARK might have had different effects on different subjects. We, therefore, added controls for the type of subject and replaced the SPARK dummy variable with an interaction term between SPARK and each of the three types of classes.

The design is as follows:

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

Here I_j takes value 1 if $j = s$ and zero otherwise, and $SPARK_{i,j}$ takes value 1 only if the student is in SPARK and if $j = s$.

[3]There is a section of the Physical Education curriculum devoted to aerobic activity and cardio, but just as schools cannot have their students play volleyball for more than one marking period, they cannot shift their entire Physical Education course load towards aerobic actibity.

The levels of j and s correspond to math, English, and science.

As a handful of students were in a higher-level math course, we ran one further test where the levels of j and s corresponded to English, science, and each of the two math courses.

All outcome variables are based on student grades, which are on the usual 0 to 100 scale. A 2.5 point coefficient, therefore, means that SPARK students saw their scores increase by a quarter of a grade level relative to non-SPARK students.

In all regression designs, the observations are student-classes. We cluster at the student level, using code made by Isidore Beautrelet.

Results:

First we report the simplest output, not broken up by subject. Note that some students did drop out of the sample in the middle of the school year. We do not know why this is the case.

Table 1:

	Dependent variable:		
	MP1 to MP2	MP1 to MP3	MP1 to MP4
	(1)	(2)	(3)
Spark	2.391**	0.547	1.442
	(0.939)	(1.210)	(0.974)
Constant	−6.385***	−8.835***	−6.029***
	(0.672)	(0.718)	(0.560)
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

Next, we break out results by subject.

Table 2:

	Dependent variable:		
	MP1 to MP2	MP1 to MP3	MP1 to MP4
	(1)	(2)	(3)
Spark_math	3.232**	0.278	2.857*
	(1.634)	(1.893)	(1.607)
Spark_sci	2.907*	0.239	0.089
	(1.559)	(1.504)	(1.311)
Spark_eng	0.986	1.396	1.456
	(1.541)	(2.192)	(1.813)
math	2.380	−8.962***	−3.965**
	(1.759)	(1.866)	(1.552)
eng	2.743*	−7.629***	−3.914***
	(1.625)	(1.720)	(1.479)
Constant	−8.086***	−3.371***	−3.429***
	(1.292)	(1.125)	(0.995)
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

Finally, we break math into two separate categories.

Table 3:

	Dependent variable:		
	MP1 to MP2	MP1 to MP3	MP1 to MP4
	(1)	(2)	(3)
Spark_math	3.449**	−0.675	2.735
	(1.687)	(1.996)	(1.729)
Spark_sci	2.907*	0.239	0.089
	(1.565)	(1.510)	(1.316)
Spark_eng	0.986	1.396	1.456
	(1.547)	(2.201)	(1.820)
Spark_adv_math	−9.449***	7.800**	2.140
	(3.645)	(3.716)	(2.903)
math	2.086	−8.910***	−3.915**
	(1.879)	(1.895)	(1.578)
eng	2.743*	−7.629***	−3.914***
	(1.632)	(1.727)	(1.485)
adv_math	10.000***	−1.719	−1.656
	(1.201)	(1.383)	(1.203)
Constant	−8.086***	−3.371***	−3.429***
	(1.297)	(1.130)	(0.999)
Observations	273	263	263
R^2	0.054	0.163	0.042
Adjusted R^2	0.029	0.140	0.015
Residual Std. Error	7.444 (df = 265)	9.147 (df = 255)	7.738 (df = 255)
F Statistic	2.165** (df = 7; 265)	7.092*** (df = 7; 255)	1.587 (df = 7; 255)
Note:			*p<0.1; **p<0.05; ***p<0.01

Analysis

The point estimates for math were generally higher than the point estimates for the other subjects, and were more often significant. The greater responsiveness of math scores to an intervention is in line with much of the rest of the literature (see Graff Zivin et al 2018 or the discussion in Fryer 2017).

Most of the point estimates were positive, and a handful were significant. The largest effect size was a roughly one-third grade level difference in math from quarter one to quarter two. That is, students in the treatment group saw a change in grades that was one third of a grade level higher from marking period to marking period than the control. Still, these results were not large, and not all were significant at the five percent or ten percent levels.

We include an R notebook in the replication page for this project with more robustness tests, none of which meaningfully alter our conclusions.

Recommendations

This pilot program offers some positive evidence that replacing students' sports-based skills curriculum with sustained, intense aerobic activity may improve academic performance. Given that this is just in addition to the more obvious fitness benefits of exercise, we recommend exploring this effect further. The ideal research design would be a pre-registered randomized trial, publicly outlining the empirical strategy prior to implementing it.

We suggest that gym classes across public schools on Staten Island be assigned into a treatment and control group at random, with the treatment group switching from a sports-based curriculum to an aerobic curriculum. The current study was limited to less than 100 students and 3 classes per student; thus to get more precision we would need a larger sample. A larger sample is always preferable, and it is fair to say that we will not learn much more beyond the pilot without a sample on the order of 1,000 students.

A waiver from the New York State curriculum would allow a more faithful implementation of SPARK, which as noted was constrained to fifteen minutes a day for three days a week in our pilot. We would recommend shifting as much of the program as possible to aerobic fitness instead.

We also recommend collecting baseline data on each of the students, especially academic performance in the prior year and level of fitness going into the school year, and committing in the pre-registration plan to divide students along these pre-selected variables to test for heterogeneous effects. We further recommend that the City partner with academic institutions that specialize in program evaluation and experimental designs. This will help us learn whether less fit or less academically successful students benefit more from this intervention.

Citations

Beautrelet, I. (2016, December 13). Clustered Standard Errors in R [Web log post]. Retrieved August 2, 2018, from https://economictheoryblog.com/2016/12/13/clustered-standard-errors-in-r/

Davis, C. L., Tomporowski, P. D., McDowell, J. E., Austin, B. P., Miller, P. H., Yanasak, N. E., . . . Naglieri, J. A. (2011). Exercise improves executive function and achievement and alters brain activation in overweight children: A randomized, controlled trial. Health Psychology, 30(1), 91-98.

Fryer Jr, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experimentsa. In Handbook of Economic Field Experiments (Vol. 2, pp. 95-322). North-Holland.

Graff Zivin, J., Hsiang, S. M., & Neidell, M. (2018). Temperature and human capital in the short and long run. Journal of the Association of Environmental and Resource Economists, 5(1), 77-105.

Hendricks, Paul (2015). anonymizer: Anonymize Data Containing Personally Identifiable Information. R package version 0.2.0. https://cran.r-project.org/web/packages/anonymizer/index.html

Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

Valdeo, D. (2008, January 13). Exercise Seen as Priming Pump for Students' Academic Strides. Retrieved August 14, 2018, from https://www.mdc.edu/main/images/Exercise_Seen_as_Priming_Pump_for_Students_tcm6-22258.pdf