Classroom Peer Effects in Elementary School: Evidence from a Randomized Experiment[*]

Preliminary and Incomplete

Jan Bietenbeck[†]

Abstract

I use data from a randomized evaluation of the Teach for America program to study ability peer effects in elementary school. The experiment randomly assigned about 2,000 students to classes and teachers in 17 schools located in disadvantaged communities across the United States. Students were followed over the course of one school year, and standardized tests were administered to them both at the beginning and at the end of the year. I find no evidence of an effect of mean peer ability (as measured by beginning-of-year achievement) on end-of-year student test scores, grade promotion, absenteeism, or suspensions. In my preferred specification, I can rule out test score effects larger than 0.15 standard deviations for a one standard deviation increase in mean peer ability. However, this result masks considerable heterogeneity by students' own ability. In ongoing research, I further explore the existence of non-linear peer effects and I examine the role of a variety of teacher characteristics as potential moderating factors of peer effects in the classroom.

JEL Code: *I21*
Keywords: *peer effects; randomized experiment; elementary school*

---

[*]I am grateful to David Dorn for his constant guidance and support. I thank Manuel Arellano, Stéphane Bonhomme, Bryan Graham, and seminar participants at CEMFI for their valuable comments and suggestions. Special thanks go to Mathematica Policy Research, Inc. for providing me with the data and to Steven Glazerman for answering my many questions. Funding from the Spanish Ministry of Science and Innovation (ECO2010-16726) is gratefully acknowledged.

[†]CEMFI, Casado del Alisal 5, 28014 Madrid, Spain. Email: bietenbeck@cemfi.es

1 Introduction

The study of peer effects in schools has received a great deal of attention in recent years. The interest in this topic stems from the fact that the potential for peer effects is central to many recurrent education policies such as school choice, tracking by ability, and the desegregation of schools. While there is considerable agreement in the literature on the issues of gender and behavioral peer effects,[1] the jury is still out regarding the existence and potential magnitude of ability peer effects (Sacerdote 2011). Existing estimates of spillovers from peers' academic ability encompass a large range, including zero, and seem to depend a lot on the particular setting that is studied.[2] Therefore, the value of reliable evidence on ability peer effects remains high, not least because many policies explicitly sort students into groups on the basis of some measure of their academic ability.

As is well known, the identification of peer effects is complicated by a number of econometric challenges. These include (1) accounting appropriately for endogenous selection into peer groups, (2) simultaneity of outcomes (also known as the reflection problem), and (3) identification of the relevant peer group. Many recent studies have followed Hoxby (2000) and relied on quasi-random variation in peer characteristics at the grade level within schools over time in order to estimate peer effects.[3] While this identification strategy adequately addresses problems (1) and (2), it comes at the disadvantage of having to assume that all students in the same grade are the relevant peer group from which spillovers originate. However, many authors would probably agree that students in the same classroom, rather than the larger group of students in the same grade, are a much

[1]Regarding gender peer effects at school, Hoxby (2000) and Lavy and Schlosser (2011) find that an increase in the share of girls among a student's peers increases her test scores. Figlio (2007), Neidell and Waldfogel (2010), and Carrell and Hoekstra (2011) find that having peers that exhibit externalizing behavior decreases students' academic achievement.

[2]For example, recent work by Lavy, Silva, and Weinhardt (2012) finds no evidence of spillovers from mean peer ability. In contrast, Sojourner (2012) estimates that student achievement rises by 0.35 percentiles for each 1-percentile rise in mean peer ability. Sacerdote (2011) gives a more comprehensive overview of the recent literature on peer effects in education.

[3]Examples of studies that employ this identification strategy include Lavy and Schlosser (2011), Lavy, Paserman, and Schlosser (2012), and Black, Devereux, and Salvanes (2013).

more relevant peer group for studying social spillovers, and that this is especially true for students in elementary school where classrooms are typically self-contained.[4]

In this paper, I study classroom peer effects in elementary schools located in highly disadvantaged, mostly urban neighborhoods across the United States. To do so, I use data from a well-controlled evaluation of the Teach for America (TFA) program,[5] whose experimental setup exhibits gold-standard features for the study of ability peer effects. In particular, students in the experiment were randomly assigned to classes before the start of the school year, and they were supposed to stay in these classes for the following year. This allows me to study peer effects at the classroom level and renders endogenous selection into (classroom) peer groups impossible. Moreover, students were administered standardized tests in math and reading both at the beginning (fall test) and at the end of the school year (spring test). The fall test scores provide a credible measure of baseline (peer) academic ability which is not typically available in other datasets.[6] Finally, the data contains a rich set of non-academic student outcomes from administrative records, which allows me to study the effect of peer ability on outcomes such as absenteeism, suspensions, and expulsion from school.

I begin my empirical analysis by estimating the effect of mean peer ability (as measured by beginning-of-year test scores) on student achievement in the spring test. I find little evidence of such an effect; in fact, the point estimate on mean peer ability is slightly negative though statistically not significant in all of my regressions. The 95% confidence interval in these estimates allow me to rule out effects larger than 0.15 standard deviations

[4]A recent paper by Burke and Sass (2013) confirms this intuition. Using a student and teacher fixed-effects approach, the authors find that spillovers at the classroom level are one order of magnitude larger than spillovers at the grade level.

[5]Teach for America is a non-profit organization which recruits high-achieving recent college graduates to teach in schools in low-income communities across the United States for a minimum of two years. See http://www.teachforamerica.org for more detailed information about the organization.

[6]Many other papers rely instead on proxy measures for peer ability, such as parental education or income (e.g., Ammermueller and Pischke 2009, Black, Devereux, and Salvanes 2013). Two recent studies at the school level that also use lagged test scores to measure peer ability are Lavy, Silva, and Weinhardt (2012) and Sojourner (2012).

for a one standard deviation increase in mean peer ability. I also find no evidence of an effect of mean peer ability on other student outcomes in the data. However, these results mask considerable heterogeneity by students' own ability, with high-ability students being harmed by being placed in a class with better-achieving peers. In ongoing research, I further explore the existence and magnitudes of heterogeneous and non-linear peer effects, and I examine the role of a variety of teacher characteristics as potential moderating factors of peer effects in schools.

2 Background on NETFA and Data

2.1 Study Design and Implementation

Between 2001 and 2003, Mathematica Policy Research, Inc. conducted the National Evaluation of Teach for America (NETFA). The evaluation sought to address concerns about the effectiveness of the TFA program, which had started to expand heavily around that time. In particular, given the challenging environment in which TFA teachers worked, some education experts had started to doubt whether individuals with high academic ability but little teacher training could really improve student outcomes in these schools. The basic idea underlying NETFA was to conduct a randomized controlled trial in which students would be randomly assigned to classes and to TFA and non-TFA teachers, and to compare the test score gains over the course of one year of students with a TFA teacher to those of students with a non-TFA teacher. The study found a modest positive impact of having a TFA teacher on math scores gains, but no significant impact on reading scores.

At the time of the experiment, TFA operated in 15 regions (defined as school districts or clusters of school districts) across the United States. Out of these, six regions were randomly selected to participate in the study after stratifying on urbanicity and on the dominant race and ethnicity of the students served. The six regions were: Baltimore, Chicago, Los Angeles, Houston, New Orleans, and Mississippi Delta. Within each of these regions, a random sample of primary schools that fulfilled the following two criteria was drawn: (1) TFA and non-TFA teachers were employed in the same grade; (2) students were not grouped into classrooms according to their academic ability and were taught by the same teacher in all subjects. The resulting sample of 17 schools is broadly representative of the schools in which TFA operated at the time of the study (Decker, Mayer, and Glazerman 2004).

The experiment was conducted in two phases. First, a pilot study was run in Baltimore in the 2001-2002 school year. This was followed by a full-scale evaluation in the other five regions in the 2002-2003 school year. Before the start of the school year, all students entering a particular grade in one of the participating schools were randomly assigned to classes and teachers. That is, random assignment took place at the school-by-grade level (a so-called block-randomized design). Students that enrolled in participating schools after the start of the school year were also randomly assigned to classes. Throughout the school year, the authors of the experiment conducted roster checks in all of the participating schools in order to enforce the random assignment. At the beginning of the school year, students took an abbreviated form of the Iowa Test of Basic Skills (ITBS) in math and reading (fall test). Students were tested again at the end of the school year. Importantly, students that had left the school but who had stayed in the same district were followed and were also administered the test.[7]

2.2 Data and Sample Selection

This paper makes use of the NETFA Public-Use Data File, which is freely available from Mathematica Policy Research, Inc. The data contains information on 1,969 students in 100 classrooms in 37 randomization blocks (ie school-grade cells) in 17 schools. Next to

[7]The attrition rate from the study, not counting within-district movers, was 11%.

test scores, the data contains student demographic characteristics from school records including student gender, age, race, ethnicity, and free-lunch status. Further information from school records includes indicators for being assigned to summer school, being promoted to the next grade at the end of the school year, the number of days a student was absent, the number of suspensions and the total number of days suspended, and school expulsion. Finally, the data contains highly detailed information on teachers' background from end-of-year teacher questionnaires, as well as information on teachers' subjective impression of the classroom environment along several dimensions. Table 1 presents summary statistics for the most important variables for this paper.

Students in the study were administered standardized tests in mathematics and reading. The form and level of these tests varied according to the grade that students were in. In order to make test scores comparable across grade levels, I standardize test scores by form and level to have mean zero and unit variance. I then compute a composite score for each student as the average between her standardized math and reading scores. This is meant to increase precision by reducing measurement error. My measure of mean peer ability is the average of a student's classmates' baseline composite score. Because students that entered the participating schools after the fall test was administered lack baseline test scores and sudents that left the school district during the school year lack spring test scores, I am forced to drop about 300 observations from my data.[8] My final estimation sample consists of 1,684 students in 99 classrooms in 17 schools.

2.3 Evidence on the Validity of the Experimental Design

Identification of peer effects in this paper is based on the between-classroom variation in peer ability within randomization blocks. In this subsection, I marshal two pieces of evidence that confirm that this variation came about randomly as part of the experimental

[8]In unreported results, I confirm that attrition from the sample is unrelated to peer ability.

design. As a first test of random assignment to classes, I ran a series of two-sample Kolmogorov-Smirnov tests for the equality of the distributions of baseline test scores across all possible pairs of classes within each randomization block.[9] Of the 101 p-values resulting from these tests, only 11 were smaller than 0.10, which is close to what one would expect if students were randomly assigned to classes. As a second test of random assignment, I regressed each student's own baseline test score on the mean baseline score of her classroom peers. The coefficient estimate on mean peer ability in this regression was -0.25 with a standard error of 0.40 (clustered at the class level). Thus, there is no evidence that the assignment of students to classes in the experiment was not random.[10]

3 Results

3.1 Visual Evidence

In order to get a first idea of the magnitude of classroom peer effects, I plot regression-adjusted spring test scores against peer mean ability. I regress each student's spring test score on her fall test score, an indicator for being in a class with a TFA-teacher, and a set of randomization-block fixed effects. I then take the residuals from this regression, compute means by deciles of peer mean ability, and plot these means against each decile bin. Figure 1 shows that there is no correlation between a student's end-of-year spring score and her peers' mean ability. I confirm this result in the following regression exercise.

[9] Like in any randomized controlled trial, there were some irregularities in the implementation of the experiment. Thus, some students managed to switch to another class despite the relatively strict enforcement of the random assignment. Therefore, I use the initial, truly random class assignment in all of my specifications (ie I estimate intent-to-treat effects).

[10] In future versions of this paper, I will provide more details on the nature of these randomization checks.

3.2 Results from Linear-in-Means Specifications

Table 2 shows results from the corresponding linear-in-means models, in which a student's spring test score is regressed on her peers' mean baseline ability (as measured by the fall score). All of the regressions in this and in subsequent tables control for randomization-block fixed effects. In column (1), students' own baseline ability is the only regressor besides the treatment variable. The model in column (2) additionally includes a dummy indicating whether a student was taught by a TFA teacher or not. The model in column (3) adds controls for student demographic characteristics.

The estimates in Table 2 provide no evidence of an effect of classroom peers' mean ability on test scores. Indeed, the point estimate on the treatment variable is negative in all three specifications, though not statistically significant. The inclusion of controls reduces the absolute value of the coefficient somewhat, and leads to more precise estimates. In particular, the 95% confidence interval of the estimate in column (3) excludes positive spillovers greater than 0.15 standard deviations for a 1 standard deviation increase in mean peer ability.[11] This is considerably lower than the comparable estimate by Sojourner (2012, 0.35 standard deviations) and the conceptually similar estimates in Hoxby (2000, 0.3-0.5 standard deviations). This result is however in line with more recent evidence by Lavy, Silva, and Weinhardt (2012), who find no effect of mean peer ability, and by Burke and Sass (2013), who estimate a 0.03 increase in test scores for a 1-point increase in mean peer ability.[12]

[11] In ongoing research, I implement measurement error corrections developed in Ammermueller and Pischke (2009) and Sojourner (2012) in order to further increase the precision of my estimates.

[12] In unreported regressions, I estimate the effect of peer mean ability on other outcomes including absenteeism, schools expulsion, and grade promotion. Similarly to the results shown in Table 2, I find no significant effect of peer mean ability in these regressions.

3.3 Heterogeneous Effects by Own Ability

One question that is of particular interest to researchers is whether the effects of peers' mean ability differ by a student's own ability. In order to shed light on this question, I define low- (high-) ability students as those students scoring in the bottom (top) 25% of the fall test score distribution. I then stratify the sample according to these own-ability groups and estimate linear-in-means specifications as in Table 2 separately for each group. Table 3 shows the results from these regressions. The results show considerable heterogeneity of the effect of peer mean ability by students' own ability. In particular, while there is no significant effect on low- and middle-ability students, high-ability students appear to be hurt by being surrounded by more high-ability peers. This result is a significant departure from results in the previous literature, which has typically found that high-ability students benefit from the presence of other high-ability students in their peer group (Sacerdote 2011).

4 Ongoing Research (*in lieu* of a Conclusion)

This is an ongoing research project in which I expect to make significant progress in the following months. Rather than giving preliminary conclusions, I will outline some of the research lines that I am planning on following here. First, clearly more evidence on the nature of peer effects in this particular setting of highly-disadvantaged elementary schools is desirable. This includes further heterogeneity results (for example by gender or by race), and, more importantly, evidence on potential non-linearities. To that end, I plan on estimating models that include the share of high- and low-achieving peers in the classroom instead of mean peer ability. Further models could include the standard deviation of peers' test scores (as in Black, Devereux, and Salvanes 2013) or other moments of the ability distribution of peers.

Second, I am planning to exploit the rich set of outcome variables available in the data set. Results on the effect of peer ability on non-academic outcomes such as the number of times a student was suspended during the school year or the number of days she was absent will provide new insights into the nature of peer effects (to the best of my knowledge, the effect of peer ability on these outcomes has not yet been studied in the literature).

Finally, and most importantly, I am planning on exploiting the rich set of teacher-level variables that are available in the data in order to peek into the black box of ability peer effects. In particular, I am going to analyze the role of teacher characteristics such as experience and training as moderating factors of peer effects in the classroom (as an example, it is plausible to think that more experienced teachers are better at exploiting the potential gains from having a heterogeneous group of students in the classroom). To do this, I am going to estimate models that interact these teacher characteristics with the relevant measure of peer ability.

References

Ammermueller, Andreas, and Jörn-Steffen Pischke (2009). "Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study." *Journal of Labor Economics* 27(3):315–348.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2013). "Under Pressure? The Effect of Peers on Outcomes of Young Adults." *Journal of Labor Economics* 31(1):119–153.

Burke, Mary A., and Tim R. Sass (2013). "Classroom Peer Effects and Student Achievement." *Journal of Labor Economics* 31(1):51–82.

Decker, Paul T., Daniel P. Mayer, and Steven Glazerman (2004). "The Effects of Teach For America on Students: Findings from a National Evaluation." Institute for Research on Poverty Discussion Paper 1285-04.

Figlio, David N. (2007). “Boys Named Sue: Disruptive Children and their Peers.” *Education Finance and Policy* 2(4):376–394.

Lavy, Victor, Daniele M. Paserman, and Analía Schlosser (2012). “Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom.” *Economic Journal* 122(559):208–237.

Lavy, Victor, Olmo Silva, and Felix Weinhardt (2012). “The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools.” *Journal of Labor Economics* 30(2):367–414.

Lavy, Victor, and Analía Schlosser (2011). “Mechanisms and Impacts of Gender Peer Effects at School.” *American Economic Journal: Applied Economics* 3(2):1-33.

Neidell, Matthew, and Jane Waldfogel (2010). “Cognitive and Noncognitive Peer Effects in Early Education.” *Review of Economics and Statistics* 92(3):562–576.

Sacerdote, Bruce (2011). “Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?” In *Handbook of the Economics of Education*, Vol. 3, ed. Erik Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North-Holland.

Sojourner, Aaron (2012). “Identification of Peer Effects with Missing Peer Data: Evidence from Project STAR.” *Economic Journal* 123(569):1–32.

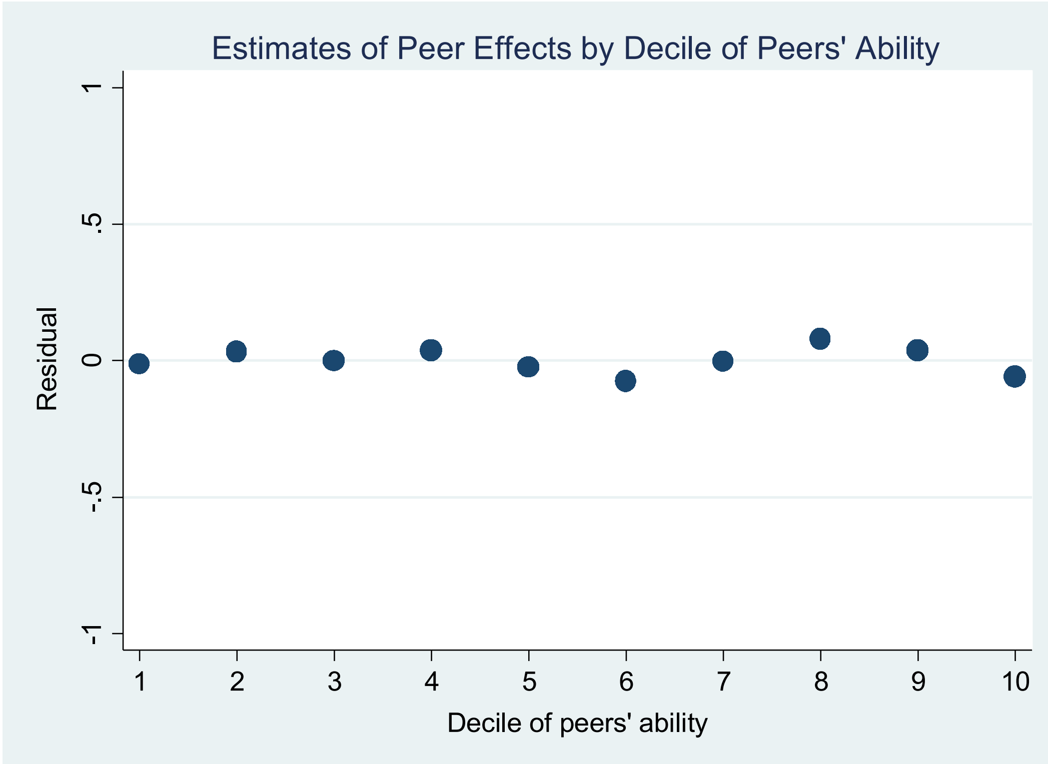

Figure 1

Table 1
Summary Statistics

	N	Mean	SD
Student characteristics			
Female	1,969	0.49	0.50
Black	1,969	0.68	0.47
Hispanic	1,969	0.26	0.44
Free lunch	1,513	0.98	0.13
Age	1,519	9.38	1.48
Teacher and class characteristics			
TFA teacher	1,969	0.44	0.50
Female teacher	1,815	0.75	0.43
Teaching experience (years)	1,799	5.66	8.64
Class size	1,969	20.91	4.92
Student outcomes from school records			
Days absent	1,676	8.24	7.78
Expelled from school	1,438	0.01	0.29
Assigned to summer school	1,635	0.34	0.47
Promoted to next grade	1,774	0.91	0.29

Table 2
Evidence on Ability Peer Effects from Linear-in-Means Models

	(1)	(2)	(3)
Peer mean ability	-0.054	-0.027	-0.009
	(0.106)	(0.092)	(0.075)
Own ability	0.743**	0.744**	0.742**
	(0.015)	(0.015)	(0.016)
Control for TFA-status	No	Yes	Yes
Additional student controls	No	No	Yes
Number of students	1,684	1,684	1,684
R-squared	0.557	0.559	0.562

Notes: All specifications control for randomization-block fixed effects. Student background variables include the student's age, gender, and dummies for black and Hispanic. Standard errors in parentheses clustered at the classroom level. ** p<0.01

Table 3
Heterogeneous Effects

	Own ability		
	Lowest 25%	Middle 50%	Top 25%
	(1)	(2)	(3)
Peer mean ability	0.043	-0.004	-0.191~
	(0.112)	(0.120)	(0.106)
Own ability	0.641**	0.769**	0.714**
	(0.078)	(0.061)	(0.059)
Number of students	419	861	404
R-squared	0.286	0.212	0.372

Notes: All specifications control for randomization-block fixed effects, TFA status, and student background variables (age, gender, and dummies for black and Hispanic). Standard errors in parentheses clustered at the classroom level. ~ p<0.1, ** p<0.01