THE LONG-TERM IMPACTS OF LOW-ACHIEVING CHILDHOOD PEERS: EVIDENCE FROM PROJECT STAR

Jan Bietenbeck
Lund University

Abstract
This paper evaluates how sharing a kindergarten classroom with low-achieving repeaters affects the long-term educational performance of regular first-time kindergarten students. Exploiting random assignment of teachers and students to classes in Project STAR, I document three sets of causal impacts: students who are exposed to repeaters (1) score lower on a standardized math test at the end of kindergarten, an effect that fades out in later grades; (2) show persistent improvements in noncognitive skills such as effort and discipline; and (3) are more likely to graduate from high school and to take a college entrance exam around the age of 18. I argue that the positive spillovers on long-term educational attainment are driven by the differential accumulation of noncognitive skills by repeater-exposed students during childhood. Results are consistent with the hypothesis that the improvements in these skills are driven by behavioral adjustments of teachers to the presence of repeaters in the classroom. (JEL: I21, I24)

1. Introduction

A large academic literature studies the effects of class composition on student performance in school. Papers in this literature have generally found positive impacts from sharing a classroom with higher-achieving and better-behaved peers (e.g., Hoxby 2000; Burke and Sass 2013; Sojourner 2013) and corresponding negative impacts from sharing a classroom with low-achieving or disruptive peers (e.g., Figlio 2007; Carrell and Hoekstra 2010; Lavy et al. 2012b). The vast majority of these papers has focused exclusively on short-term spillovers on contemporaneous outcomes, such as the effect of kindergarten classmates on test scores at the end of kindergarten. However, in order

The editor in charge of this paper was Daniele Paserman.

Acknowledgments: This paper is a revised version of a chapter of my Ph.D. dissertation. I am very grateful to David Dorn for his guidance and support. I also thank Joseph Altonji, Manuel Arellano, David Autor, Manuel Bagues, Sascha Becker, Samuel Bentolila, Susan Dynarski, Joshua Goodman, Bryan Graham, Alessandro Martinello, Claudio Michelacci, Pedro Mira, Magne Mogstad, Luca Repetto, Jan Stuhler, Petra Thiemann, Simon Wiederhold, the editor, several anonymous referees, and numerous seminar and conference audiences for helpful comments. Diane Schanzenbach generously provided a subset of the data used in this paper. Funding from the Spanish Ministry of Science and Innovation (BES-2011-050947) is gratefully acknowledged.

E-mail: mail@janbietenbeck.com (Bietenbeck)

Journal of the European Economic Association 2019 00(0):1–35 DOI: 10.1093/jeea/jvy049

to judge the overall efficacy of policies that change the student composition of classes and schools, it is important to know how such spillovers play out in the long term.

In this paper, I study how sharing a kindergarten classroom with low-achieving repeaters affects the long-term educational performance of regular, first-time kindergarten students. The empirical analysis uses data from the Tennessee Student–Teacher Achievement Ratio experiment (Project STAR), which are uniquely suited for this purpose for three reasons. First, the data allow me to identify kindergarten repeaters as a particularly low-achieving group of peers: by definition, these students have a proven track record of failure, and as I show in what follows, they are characterized by exceptionally low cognitive and noncognitive skills. Second, Project STAR randomly assigned teachers and students, including repeaters, to classes within schools. This lets me estimate causal spillover effects from repeaters that are free from selection bias. Finally, the data contain a rich set of medium- and long-term outcomes for students, including measures of noncognitive skills, high school graduation, and college-test taking.

The main empirical specifications relate regular students' exposure to repeaters in kindergarten, measured as being randomly assigned to a class containing at least one repeater, to their educational performance at different points in time.[1] Being exposed to repeaters significantly lowers students' performance on a standardized math test at the end of kindergarten, a result that corroborates previous findings of negative short-term spillovers from low-achieving peers. In contrast, repeater exposure substantially *increases* students' noncognitive skills such as effort and discipline, which are first measured at the beginning of fourth grade. Although the negative impact on math scores fades out and, if anything, turns positive over time, the gains in noncognitive skills persist. Consistent with these last results, students who share a kindergarten classroom with repeaters show improved long-term educational attainment as evidenced by higher propensities to graduate from high school and to take a college entrance exam.

In additional analyses, I explore the mechanisms behind these results. Motivated by recent findings that noncognitive skills formed early in life are a key determinant of long-term educational success (e.g., Heckman et al. 2006), I argue that the positive spillovers on high school graduation and college-test taking are driven by the differential accumulation of such skills by repeater-exposed students. As for how exactly sharing a classroom with repeaters boosts noncognitive skills, the experimental setup and detailed longitudinal data let me rule out a wide range of potential explanations that involve selection of students into classes or schools, selective attrition, or differential access to educational resources. As an alternative mechanism, I suggest that teachers, students, or parents may change their behavior in response to the presence of low-achieving repeaters. For example, teachers whose classes are frequently disrupted by repeaters may focus on teaching students good classroom behavior and study skills at the cost of kindergarten math skills. Although

1. I use the terms "regular student", "nonrepeating student", and "first-time kindergarten student" interchangeably throughout the paper.

this explanation receives some support from the educational psychology literature, the data do not allow me to test it directly.

This paper contributes to a large literature on peer effects in schools, which is reviewed in detail in Sacerdote (2011). This literature includes studies based on Project STAR, most notably by Whitmore (2005), who examines the effects of classroom gender composition, and by Graham (2008) and Sojourner (2013), who investigate spillovers from peers' academic ability. It also contains a few recent papers that document negative spillovers from repeaters on their classmates' educational achievement (Lavy et al. 2012a; Gottfried 2013; Hill 2014). All these studies, like most of the research on peer effects, focus exclusively on short-term impacts. As rare exceptions to this norm, Gould et al. (2009) show that sharing a fifth-grade classroom with immigrants affects the likelihood of natives to graduate from high school in Israel, and Cascio and Schanzenbach (2016) analyze how the average age of students' kindergarten classmates affects long-term educational outcomes in Project STAR.[2]

The main contribution of this paper is to provide the first evidence on long-term spillovers from low-achieving childhood peers, and some of the first evidence on long-term peer effects more generally. Studying these early spillovers is particularly important because kindergarten students are at an age where both cognitive and noncognitive skills are still highly malleable (Kautz et al. 2014). Moreover, the long-term impacts examined here are arguably more relevant than short-term effects for the evaluation of policies, as they may translate more directly into changes in labor market outcomes. The importance of investigating such long-term impacts is further highlighted by the finding that short- and long-term effects do not necessarily go in the same direction.

The remainder of this paper is organized as follows. In the next section, I review the experimental design of Project STAR and summarize my data construction. Section 3 details the identification strategy and presents evidence on the random assignment of repeaters. In Section 4, I analyze the effect of repeater exposure in kindergarten on regular students' short- and long-term outcomes. Section 5 discusses potential underlying mechanisms and Section 6 presents results from robustness checks. Section 7 concludes.

2. The STAR Experiment and Data

2.1. Background on Project STAR

Project STAR was a randomized experiment designed to study the effects of class size on student achievement. In the beginning of the 1985–1986 school year, 6,325

2. This paper is also related to a study by Chetty et al. (2011), which shows that kindergarten class fixed effects predict earnings at ages 25–27 of participants in Project STAR. As the authors of that study note, these "class effects" combine the impacts of peers, teachers, and any other class-level shocks and therefore cannot be interpreted as pure peer effects. This paper moreover connects to two studies by Bifulco et al. (2011) and Black et al. (2013), which examine spillovers from high school peers on longer-term educational and labor market outcomes.

kindergarten students in 79 participating Tennessee schools were randomly assigned to small classes (target size 13–17 students) or regular-sized classes (22–25 students) within their schools.[3] They were supposed to stay in their assigned class type (small vs. regular-sized) until the end of third grade, after which the experiment ended and they would return to ordinary classes. Students that joined the initial cohort in participating schools after the kindergarten year were also randomly assigned to class types, as were teachers in each grade.

This study exploits the fact that kindergarten students and teachers were randomly assigned not only to class type, but also to a particular class within each type (50 schools in the experiment had multiple classes per type). Although assignment to classes was not documented in detail in the STAR Technical Report (Word et al. 1990), several recent studies provide extensive evidence supporting randomization (Chetty et al. 2011; Sojourner 2013; Cascio and Schanzenbach 2016). Section 3 revisits some of this evidence and provides new statistical support for the claim that repeaters were randomly assigned to kindergarten classes within schools.

The eventual implementation of Project STAR deviated somewhat from the original experimental plan, with three differences being particularly important in the context of this paper. First, as the initial cohort advanced from kindergarten to third grade, there was substantial attrition due to students moving to other schools or being retained in grade. Thus, by the time the cohort reached third grade, 49% of students who had participated in the experiment in kindergarten had left the sample. In Section 5, I provide evidence that this attrition was not related to exposure to repeaters during kindergarten and therefore does not drive my long-term results.

Second, although compliance with treatment assignment was nearly perfect in kindergarten, approximately 10% of students managed to switch between small and regular-sized classes in each of the subsequent grades (Krueger 1999). Due to the focus on spillovers from repeaters in kindergarten, noncompliance with class assignment in the later grades does not affect the (reduced-form) causal interpretation of results in this paper. Nevertheless, in what follows I also present evidence that students did not systematically change to better classes after being exposed to repeaters in kindergarten.

Finally, because of complaints by some parents about their children's initial assignment, students in regular-sized classes were re-randomized at the beginning of first grade. This feature of the experiment, as well as the attrition and class switching described previously, changes the total amount of time that students spent in class with a kindergarten repeater. In turn, this affects the interpretation of the repeater-exposure treatment, a point that I discuss in more detail in the following section.[4]

3. There was also a third class type: regular-sized classes with a full-time teacher's aide. Like previous analyses of Project STAR, I do not find any differences in treatment effects between regular-sized classes with and without a full-time teacher's aide.

4. Additional details regarding the design and implementation of Project STAR can be found in Word et al. (1990), Krueger (1999), and Finn et al. (2007).

2.2. Variable Definitions

Data for students participating in Project STAR were collected by various research teams and organizations both during the experiment and in several rounds after the experiment ended. The Project STAR public use file, on which the empirical analysis is based, combines these data such that students can be followed throughout their scholastic careers until the end of high school.[5] This section gives a brief overview of the main dependent and independent variables used in the empirical analysis. Online Appendix B provides additional details on data collection procedures and the construction of variables.

Demographic Characteristics. The data contain information on students' gender, race, eligibility for free or reduced-price lunch, and exact date of birth. Children in Tennessee are supposed to enter kindergarten if they are five years or older on September 30 of a given year, and I use this rule to construct an old-for-grade indicator that takes value 1 if the student was six years or older on September 30, 1985, and 0 otherwise. Students in Project STAR may be old for grade either because they entered school late (the so-called "red-shirting") or because they were repeating kindergarten.[6]

Kindergarten Repeaters. The data include an indicator for whether each student was repeating kindergarten in the 1985–1986 school year. There are 253 repeaters in the sample, 193 of whom are old for grade. Note that *all* repeaters would be expected to be old for grade if they had entered kindergarten in accordance with Tennessee's school entry rules during one of the previous school years. Therefore, the 60 repeaters who were not old for grade must have entered school early. The empirical analysis focuses on spillover effects from the 193 old-for-grade repeaters, who first entered kindergarten at the regular entry age. Although the data do not contain information on the exact reason for their retention, these students had likely been identified by principals or teachers as having cognitive or behavioral deficiencies that would have put them at a disadvantage had they been promoted to first grade. The same is not necessarily true for the 60 other repeaters, who may have stayed in kindergarten only because they were too young to enter first grade.[7]

5. Data on some of the outcomes studied in this paper were generously provided to me by Diane Schanzenbach; see Online Appendix B for details.

6. See Deming and Dynarski (2008) for an analysis of the red-shirting phenomenon in the United States.

7. Children are required to be six years old on September 30 of the year they start first grade. This rule was likely enforced more strictly than the kindergarten entry rule since kindergarten attendance was not mandatory in Tennessee at the time of Project STAR. Empirically, the 60 "young" repeaters come from more favorable demographic backgrounds and exhibit better cognitive and noncognitive outcomes than the 193 old-for-grade repeaters. If all 253 repeaters are used as treatment in the empirical analysis, the estimated spillover effects are usually smaller than the ones reported in the paper.

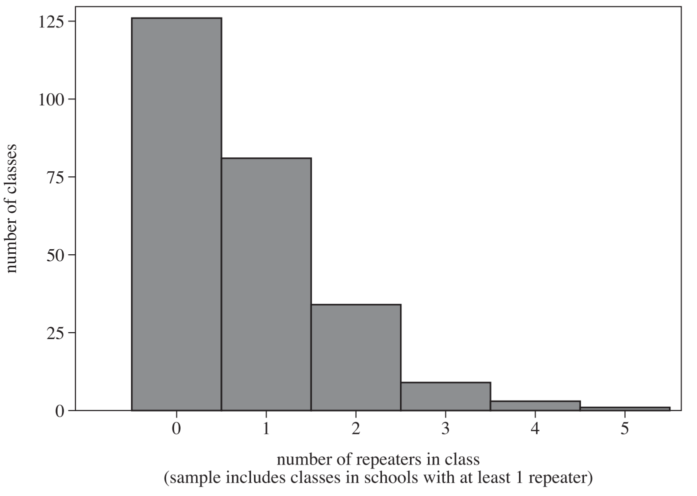

FIGURE 1. Distribution of repeaters across classes. The figure displays a histogram of the number of repeaters in class. The sample includes only the 60 (out of 79) schools with at least one repeater in kindergarten. There are 254 classes in this sample, with a mean (median) number of repeaters of 0.76 (1).

Repeater Exposure. Figure 1 shows the distribution of repeaters across classes in the 60 schools with at least one repeater.[8] 126 of the 254 classes contain no repeater, 81 contain exactly one repeater, and only 47 contain two or more repeaters. In view of this heavily skewed distribution, the main specifications of the empirical analysis will distinguish just between classes with and without repeaters. In Section 6, I also present estimates from alternative specifications in which I measure repeater exposure as the actual number of repeaters in class, or as the share of repeaters in class. Results from these regressions suggest that outcomes are similar for students who are exposed to one or to several repeaters, which implies that the main specifications using a dummy variable for the presence of at least one repeater in class do not unduly miss heterogeneous treatment effects.

An important question for the interpretation of results is whether the spillovers on long-term outcomes documented in this paper arise from exposure to repeaters during kindergarten or from exposure over a longer time horizon. If all children had stayed in their assigned kindergarten classes until the end of the experiment, regular students would have been exposed to repeaters either for four years or not at all until third grade. In practice, however, due to the various deviations from the original experimental design described previously, students who were exposed to repeaters in

8. The other 19 schools without repeaters do not contribute to the identification of spillover effects in this paper, which is based on between-class variation in the number of repeaters *within* schools. Compared to schools without repeaters, schools with repeaters are slightly smaller (average enrollment of 73 students vs. 83 students), are less likely to be located in the inner city (12% vs. 47% of schools), and contain lower fractions of black students (20% vs. 61%) and low-income students (41% vs. 67%).

kindergarten and who had not left the experiment by third grade ended up being in class with at least one of these repeaters for 2 years on average, whereas students not exposed to repeaters in kindergarten ended up being in class with repeaters for an average of 0.6 years.[9] The treatment studied in this paper thus consists of exposure to repeaters during kindergarten and an additional six months of differential exposure during grades 1–3.

Outcomes. At the end of each grade level from kindergarten through third grade, students were administered the grade-appropriate version of the Stanford Achievement Test. Moreover, in the spring of grades 5–8, all participants still attending public school in Tennessee took the Comprehensive Test of Basic Skills as part of a statewide student assessment program. Both tests are standardized multiple-choice assessments with components in mathematics and reading. The empirical analysis studies the effects of repeater exposure in kindergarten on student performance on these tests at each grade level.

In November 1989, when participants were in fourth grade, teachers in the STAR schools were asked to evaluate a random subset of their students on a set of behavioral measures. Teacher ratings were recorded on a scale from 1 to 5 and were consolidated into four indices. The effort index is based on such items as whether a student completes her homework and whether she is persistent when confronted with difficult problems. The initiative index captures such characteristics as whether a student actively participates in classroom discussions. The value index measures how much a student appreciates the school learning environment. Finally, the discipline index is based on such items as whether a student often acts restless and whether she interferes with her peers' work. In eighth grade, math and English teachers were asked to rate a different random subset of STAR participants on similar questions, the answers to which were consolidated into the same four indices. The total of eight fourth- and eighth-grade indices derived from teacher ratings serve as measures of noncognitive skills in the empirical analysis.

Most STAR participants graduated from high school in 1998, and transcripts including information on high school grade point average (GPA) and graduation status were collected from selected high schools in 1999 and 2000. Colleges and universities in the United States typically require applying students to report results from either the ACT or the SAT test. In 1998, Krueger and Whitmore (2001) matched all STAR students to the administrative records of the two companies responsible for these tests. The outcome of this process is an indicator that takes value 1 if a student took either of

9. These figures come from a regression of cumulative years of exposure at the end of third grade on a constant, an indicator for repeater exposure in kindergarten, and school fixed effects. Further analysis showed that cumulative years of exposure are very similar for students assigned to small and to regular-sized kindergarten classes. Note that these figures measure exposure to the 193 *original* repeaters for students who did not attrit from the experiment. A complete history of exposure to *any* repeaters cannot be determined for participants in Project STAR because class composition is no longer observed for students who leave the experiment and because repeater status was not recorded for students who entered the experiment after kindergarten.

these college entrance exams in 1998 and 0 otherwise. Together, high school GPA, high school graduation, and college-test taking are the measures of long-term educational attainment studied in this paper.

2.3. Sample Selection and Descriptive Statistics

The full sample includes 6,325 kindergarten students in 127 small and 198 regular-sized classes in 79 schools. I exclude 28 students for whom repeater status is not observed and five students with missing demographic characteristics from this sample. I further drop the 60 repeaters who are not old for grade as they had likely been in class with one of the old-for-grade repeaters during the previous (1984–1985) school year and are thus subject to a fundamentally different treatment. The final estimation sample thus consists of 6,232 students, 193 of whom are repeaters. Results in this paper are robust to relaxing the sample restrictions discussed in this paragraph.

Table 1 reports descriptive statistics for demographic characteristics, repeater exposure, and key outcome variables separately for nonrepeating and repeating kindergarten students in the estimation sample. Students in general exhibit lower socioeconomic characteristics than the student populations in Tennessee and the United States as a whole because Project STAR oversampled schools in low-income neighborhoods (Krueger and Whitmore 2001). Repeaters are predominantly male and are more likely to be eligible for free or reduced-price lunch than first-time kindergarten students. Repeaters are also older than nonrepeating students by definition. Since low-income schools with primarily black student populations have lower repeater shares on average, repeating students in the sample are less likely to be black. Finally, only three percent of nonrepeating students are old for grade, which shows that red-shirting was not common in the schools participating in Project STAR at the time of the experiment.

In order to facilitate easy comparison between the outcomes of regular students and repeaters, I standardize all test scores and noncognitive skill measures to have mean 0 and standard deviation 1 across nonrepeating students in the estimation sample. Table 1 shows that repeaters tend to perform substantially worse than regular students in school. For instance, they score half a standard deviation lower on the end-of-kindergarten reading test, and they are rated between a third and a full standard deviation lower on measures of effort, initiative, value, and discipline.[10] These gaps are comparable in size to those found in the educational psychology literature, which has attributed repeaters' poor performance mainly to their low levels of school readiness when entering kindergarten (Karweit 1999).[11] That literature has moreover sustained

10. In unreported regressions, I found that repeaters' measured cognitive and noncognitive skills are also significantly below those of male students, black students, and students eligible for free or reduced-price lunch. This suggests that by focusing on repeaters, I may be more successful in identifying truly low-achieving peers than by simply categorizing students as low achievers based on their demographic background, an argument that is also made by Lavy et al. (2012a).

11. These low levels of school readiness manifest themselves as low levels of cognitive ability, low attention spans, and high levels of emotional and social immaturity. Like in Project STAR, males, minority

TABLE 1. Descriptive statistics.

	Nonrepeaters			Repeaters		
	N	Mean	SD	*N*	Mean	SD
Demographic characteristics						
Male	6,039	0.51	0.50	193	0.70	0.46
Black	6,039	0.33	0.47	193	0.17	0.38
Free lunch	6,039	0.48	0.50	193	0.65	0.48
Age in years	6,039	5.48	0.31	193	6.39	0.31
Old for grade	6,039	0.03	0.17	193	1.00	0.00
Repeater exposure						
At least 1 repeater in class	6,039	0.39	0.49	–	–	–
Standardized test scores						
Kindergarten math score	5,614	0.00	1.00	175	−0.36	0.80
Kindergarten reading score	5,535	0.00	1.00	173	−0.47	0.69
8th-grade math score	4,353	0.00	1.00	102	−0.88	1.09
8th-grade reading score	4,364	0.00	1.00	108	−0.93	1.15
Noncognitive skills						
4th-grade effort	1,628	0.00	1.00	32	−1.13	1.24
4th-grade initiative	1,628	0.00	1.00	32	−1.01	1.01
4th-grade value	1,628	0.00	1.00	32	−0.83	1.25
4th-grade discipline	1,628	0.00	1.00	32	−0.32	1.20
8th-grade effort	1,731	0.00	1.00	37	−0.50	1.09
8th-grade initiative	1,731	0.00	1.00	37	−0.43	0.91
8th-grade value	1,731	0.00	1.00	37	−0.36	1.17
8th-grade discipline	1,731	0.00	1.00	37	−0.29	1.06
Long-term outcomes						
High school GPA	2,438	84.20	7.42	40	81.82	7.35
High school graduation	2,955	0.87	0.34	60	0.67	0.48
Took ACT/SAT	6,039	0.41	0.49	193	0.12	0.32

Notes: The table reports descriptive statistics of key variables separately for the 6,039 nonrepeating students and the 193 repeaters in the estimation sample. A student is considered old for grade if based on her age and Tennessee's kindergarten entry cutoff date of September 30 she would be expected to attend at least first grade in the 1985–1986 school year. Repeater exposure is measured as an indicator taking value 1 if the student's kindergarten class contains at least one repeater and 0 otherwise. Repeater exposure is not defined for repeaters because this paper studies spillovers from repeaters on nonrepeating students. The noncognitive skill measures are indices summarizing teacher ratings of student behavior in four areas: effort, initiative, value, and discipline. All test scores and measures of noncognitive skills are standardized to have mean 0 and standard deviation 1 across nonrepeating students. High school GPA is measured on a scale from 0 to 100. Took ACT/SAT is an indicator for whether the student took either of these tests in 1998, when most students were in their final year of high school.

that repeaters' initial disadvantages are exacerbated by retention itself, with grade repetition leading to lower academic achievement and increased disruptive behavior (e.g., Jimerson 2001; Pagani et al. 2001). Importantly for this paper, the exceptionally

students, and students from low socioeconomic backgrounds have been found to be more likely to repeat a grade in the educational psychology literature. For a detailed characterization of repeaters, see Karweit (1999).

low cognitive and noncognitive skills of repeaters might well impact the learning of their classmates, an issue that I investigate in detail in what follows.

3. Identification Strategy

3.1. Identification Based on Between-Class Variation in Repeater Exposure

Identification of spillovers from repeaters in this paper is based on between-class variation in repeater exposure within schools. The regression framework, which is described in detail in what follows, thus compares the outcomes of regular students who attend kindergarten in the same school but who are randomly assigned to classes with and without repeating schoolmates. This identification strategy requires that these classes do not systematically differ from each other in any other dimension. In nonexperimental data, this requirement will not be met if, for example, school principals systematically assign low-achieving repeaters to classes with high-achieving other students or more effective teachers. In contrast, random assignment in Project STAR ensures that classes with and without repeaters are balanced on characteristics of regular students and teachers.

One challenge to identification arises because repeater exposure is positively correlated with class size. In particular, repeaters are more likely to be observed in regular-sized classes because (i) larger classes are more likely to contain at least one repeater when students are randomly assigned to classes, and (ii) the sample contains more regular-sized classes than small classes.[12] Previous analyses of Project STAR have documented large negative effects of class size on student outcomes (see Schanzenbach 2006 for an overview of these findings). Therefore, a regression of student performance on repeater exposure that does not control for class size will yield an estimate that is biased. I avoid such bias by controlling for class size in all of my regressions. In what follows, I also present results from specifications that allow the effects of repeater exposure to vary with class size.

Section 4 reports estimates of the following empirical model:

$$y_{ics} = \alpha_s + \beta_1 \mathrm{EXPOSURE}_{cs} + \beta_2 \mathrm{SMALL}_{cs} + X_{ics}\gamma + \varepsilon_{ics}, \tag{1}$$

where y_{ics} is a kindergarten or long-term outcome for nonrepeating student i randomly assigned to kindergarten class c in school s, $\mathrm{EXPOSURE}_{cs}$ is an indicator for whether student i's class contains at least one repeater, SMALL_{cs} is an indicator for small class in kindergarten, and X_{ics} is a vector containing the five student demographic characteristics shown in Table 1. Because random assignment to classes took place within schools, the model also controls for a vector of school fixed effects (α_s).

12. Consider, for example, a school with the typical configuration of one small class of 15 students and two regular-sized classes of 23 students. If this school contains one repeater (the mode among schools with positive numbers of repeaters), this repeater has a 46/61 probability of being assigned to a regular-sized class and a 15/61 probability of being assigned to the small class.

3.2. Evidence on Random Assignment of Repeaters

The key identification assumption underlying the specification in equation (1) is that conditional on class size and school fixed effects, classes with and without repeaters do not differ systematically in any other dimension. Intuitively, this assumption holds here because of the random assignment of students and teachers to classes in Project STAR. This intuition is supported by evidence from previous analyses of the experimental data (e.g., Chetty et al. 2011; Cascio and Schanzenbach 2016), which show that classes are balanced on a wide range of student, parent, and teacher characteristics. Here, I complement this evidence by evaluating whether repeaters were indeed randomly assigned to classes within schools.

As a first test for random assignment, I checked whether the within-school variation in repeater exposure observed in the data is consistent with a random allocation process. To that end, I performed a Monte Carlo simulation in which students were randomly assigned to classes within schools and in which the number and size of classes and the number of repeaters in each school were based on the actual data. In the simulated data, I then regressed the number of repeaters in a student's class on school fixed effects and collected the residuals; intuitively, these residuals describe the identifying variation used in the empirical analysis. Figure 2 plots the distribution of the residuals from 1,000 replications of this exercise and from an equivalent regression using the actual data. The two distributions look very similar to each other, supporting the assumption that repeaters were randomly assigned to classes within schools.

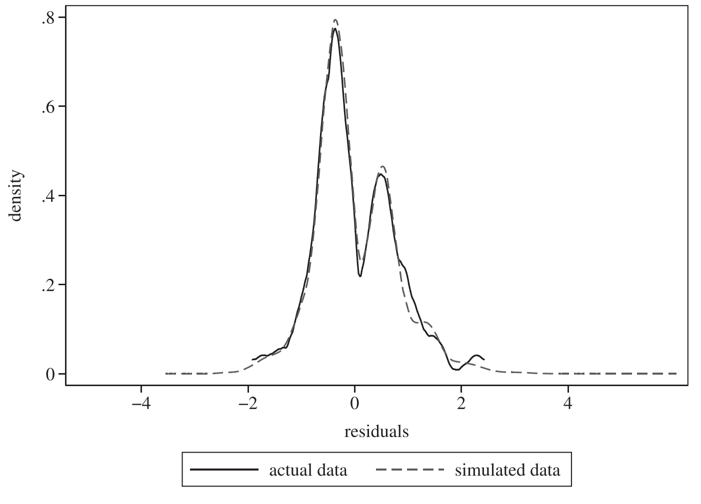

FIGURE 2. Actual and simulated variation in the number of repeaters in class. The figure shows kernel density plots of residuals from regressions of the number of repeaters in class on school fixed effects. The solid line corresponds to residuals from a single regression using the actual data, whereas the dashed line corresponds to residuals from 1,000 regressions using simulated data in which students were randomly assigned to classes within schools. In both cases, the sample is restricted to schools containing at least one repeater. Density calculations are based on an Epanechnikov kernel with the optimal bandwidth of 0.124 in the actual data.

As a second test for random assignment, I regressed an indicator taking value 1 if the student is a repeater and 0 otherwise on school and class fixed effects (omitting one class per school to avoid collinearity). Following the intuition described in Chetty et al. (2011), if assignment to classes was indeed random, then class indicators should not predict predetermined repeater status in this regression. Consistent with this idea, the p-value from an F-test for joint significance of the class fixed effects was 0.65, suggesting that repeater status is indeed balanced across classes.[13]

Finally, I tested whether being exposed to a repeater predicts regular students' and teachers' observable characteristics. To this end, I regressed each predetermined student and teacher variable on the repeater-exposure dummy and school fixed effects, with alternative specifications additionally controlling for class size as in Equation (1). As shown in Table A.1, the estimated coefficients on repeater exposure were mostly small, and they were statistically significant at the 10% level in only 2 out of 16 regressions (12.5% of cases), close to what would be expected by chance. These estimates suggest once more that repeaters were indeed randomly assigned to classes in Project STAR.

4. Main Results

4.1. Effects on End-of-Kindergarten Test Scores

I begin the empirical analysis by estimating the impact of repeater exposure on regular students' math and reading performance at the end of kindergarten. These short-term estimates serve as a benchmark for comparison with findings from the previous literature and with the estimates for long-term outcomes reported later on. Column (1) of Table 2 shows that in a regression of math scores on repeater exposure and small-class and school dummies, being exposed to repeaters reduces regular students' math scores by 9% of a standard deviation on average. Column (2) adds controls for students' demographic background to this regression. Due to the random assignment of students to classes, these controls do not change the coefficient estimate for the repeater-exposure treatment, but they slightly improve its precision. Columns (3) and (4) show the corresponding results for reading scores. The estimated impact of repeater exposure in these specifications is also negative, but it is substantially smaller than that in the math regressions and not statistically significant at conventional levels.[14]

The finding that repeater exposure decreases test scores in the short term is in line with the results from the previous literature, which has documented negative spillovers

13. In a similar vein, I ran a separate regression of repeater status on class dummies for each school and tested for joint significance. Online Appendix Figure C.1 shows the distribution of p-values from these F-tests. As would be expected under random assignment, the distribution is roughly uniform with a mean close to 0.5.

14. Larger impacts on math scores than on reading scores are a frequent finding in the economics of education literature; see Horoi and Ost (2015) for a recent example.

TABLE 2. Repeater exposure in kindergarten and end-of-kindergarten test scores.

	Math (1)	Math (2)	Reading (3)	Reading (4)
Repeater exposure	−0.090**	−0.090**	−0.014	−0.014
	(0.043)	(0.041)	(0.046)	(0.044)
Male		−0.144***		−0.175***
		(0.024)		(0.025)
Black		−0.355***		−0.249***
		(0.051)		(0.053)
Free lunch		−0.411***		−0.450***
		(0.029)		(0.029)
Age in years		0.550***		0.408***
		(0.044)		(0.048)
Old for grade		−0.411***		−0.346***
		(0.081)		(0.074)
Small class	0.169***	0.158***	0.194***	0.185***
	(0.045)	(0.043)	(0.043)	(0.042)
Observations	5,614	5,614	5,535	5,535
Mean of dep. var.	0.00	0.00	0.00	0.00

Notes: The table reports estimates from regressions of end-of-kindergarten math and reading scores on the variables listed in rows and school fixed effects. Test scores are standardized to have mean 0 and standard deviation 1 across nonrepeating students in the estimation sample. Repeater exposure is measured as an indicator taking value 1 if the student's class contains at least one repeater and 0 otherwise. Standard errors in parentheses allow for clustering at the class level. $^{**}p < 0.05$; $^{***}p < 0.01$.

from low-achieving and disruptive peers (e.g., Figlio 2007; Carrell and Hoekstra 2010; Lavy et al. 2012b). Although it is difficult to compare effect sizes across studies with different treatments, the impact of repeater exposure on math scores reported in Table 2 appears relatively large: for example, it is about half the magnitude of the class-size effect. One possible reason for this is that spillovers from repeaters are greater because they have lower cognitive skills and are more disruptive than other, more frequently studied low-achieving peers (see footnote 10). Another possibility is that by measuring repeater exposure at the class level, I better capture real-life interactions than previous studies, which often define peer groups at the grade-within-school level (see Burke and Sass 2013, for a similar argument).

4.2. Effects on Post-Kindergarten Test Scores

Previous analyses of peer effects in schools have focused almost exclusively on contemporaneous impacts like the ones reported in Table 2. In this paper, I move beyond this short-term perspective by tracking the effects of repeater exposure in kindergarten throughout students' entire school careers. I begin by estimating the impact of repeater exposure on regular students' math and reading scores at each grade level from kindergarten through eighth grade. Panel A of Figure 3 reveals a rapid fade-out of the negative spillover effect from repeaters on math scores: already one year

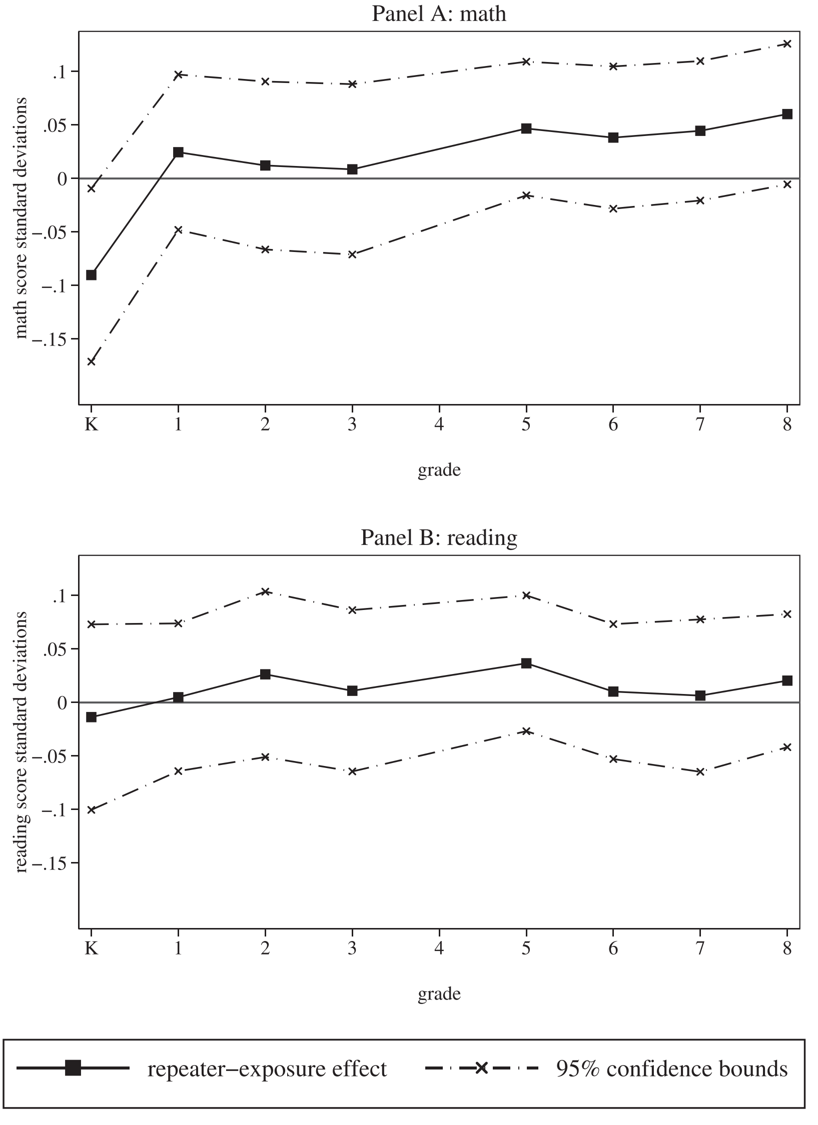

FIGURE 3. Repeater exposure in kindergarten and post-kindergarten test scores. The figure plots point estimates and 95% confidence bounds from 16 separate regressions of test scores on repeater exposure in kindergarten. The dependent variables are the math scores (panel A) and reading scores (panel B) in the year in which students were supposed to be in the grade indicated on the horizontal axis. Each regression is run on the sample of nonrepeating students observed with the corresponding test score and includes further controls as in columns (2) and (4) of Table 2. No results are reported for fourth grade because test scores are available for only a small fraction of students in that grade; see Online Appendix B for details.

after kindergarten, the estimated impact turns slightly positive, and it never falls below zero again afterward. Indeed, the magnitude of the repeater-exposure effect seems to rise over time, culminating in an estimate of a 6.0% of a standard deviation increase in eighth-grade math scores, which is marginally statistically significant. Panel B shows point estimates for reading scores that are qualitatively similar, though generally smaller in size. Overall, these results point to an interesting pattern of negative impacts of repeater exposure on test scores in the short term, which then fade out rapidly and appear to turn positive in the longer term.[15]

4.3. Effects on Noncognitive Skills

A growing literature in economics documents the importance of noncognitive skills for success in life and argues that such skills are partly formed in school (e.g., Heckman et al. 2006; Chetty et al. 2011; Heckman et al. 2012). I analyze the impacts of repeater exposure in kindergarten on noncognitive skills in Table 3. Panels A and B present results from regressions without and with demographic controls, respectively. In stark contrast to the negative short-term effects on test scores discussed previously, columns (1)–(4) show large positive spillovers from repeaters on regular students' behaviors in fourth grade. The impacts are particularly pronounced for the effort index, which measures traits such as persistence and resolution, and the discipline index, which measures good classroom behavior. Columns (5)–(8) reveal that these effects persist into eighth grade, the second and last point of measurement of these outcomes.

Column (9) shows the estimated effect of repeater exposure on a summary index of noncognitive skills. The aggregation of outcomes in this manner improves statistical power to detect effects that go in the same direction, and it allows me to present later additional results in a concise way. Following Kling et al. (2007), I construct the summary index by first averaging the standardized fourth- and eighth-grade effort, initiative, value, and discipline indices for each student (if noncognitive skills are observed in only one of the grades, I use the average of the available indices). In a second step, I then normalize this average to have mean 0 and standard deviation 1 across nonrepeating students. As can be seen in panel B of Table 3, repeater exposure raises the resulting summary index by a highly significant 11.7% of a standard deviation. As a point of comparison, the gender gap in this index is about half a standard deviation, and the gaps between black and white students and between students with and without free-lunch eligibility are about 30% of a standard deviation.[16] Thus, sharing a kindergarten classroom with repeaters has a large positive impact on regular students' noncognitive skills.

15. Figure 3 presents estimates from regressions that control for nonrepeating students' demographic characteristics. Regressions that do not include these controls yield very similar results.

16. The gender gap in the summary index of noncognitive skills is notably very similar to the gender gaps in externalizing behavior and self-control reported in Bertrand and Pan (2013) for a representative sample of fifth-grade students in the United States.

TABLE 3. Repeater exposure in kindergarten and noncognitive skills.

	4th grade				8th grade				Summary index
	Effort (1)	Initiative (2)	Value (3)	Discipline (4)	Effort (5)	Initiative (6)	Value (7)	Discipline (8)	(9)
Panel A: No demographic controls									
Repeater exposure	0.121**	0.037	0.138**	0.156***	0.160***	0.101*	0.149***	0.182***	0.127***
	(0.055)	(0.056)	(0.054)	(0.055)	(0.056)	(0.056)	(0.052)	(0.054)	(0.043)
Panel B: With demographic controls									
Repeater exposure	0.104*	0.025	0.124**	0.142***	0.169***	0.105*	0.160***	0.194***	0.117***
	(0.054)	(0.056)	(0.053)	(0.054)	(0.054)	(0.056)	(0.051)	(0.052)	(0.041)
Male	−0.430***	−0.265***	−0.422***	−0.495***	−0.528***	−0.286***	−0.460***	−0.544***	−0.527***
	(0.044)	(0.048)	(0.046)	(0.053)	(0.048)	(0.048)	(0.049)	(0.046)	(0.037)
Black	0.048	0.120	0.010	−0.106	−0.281**	−0.060	−0.371***	−0.583***	−0.243**
	(0.142)	(0.152)	(0.129)	(0.151)	(0.121)	(0.120)	(0.142)	(0.160)	(0.109)
Free lunch	−0.315***	−0.364***	−0.174***	−0.191***	−0.215***	−0.204***	−0.183***	−0.119*	−0.276***
	(0.061)	(0.063)	(0.056)	(0.058)	(0.060)	(0.061)	(0.056)	(0.062)	(0.046)
Age in years	0.202**	0.139	0.042	0.181**	0.141	0.005	0.063	0.203**	0.132*
	(0.088)	(0.090)	(0.091)	(0.086)	(0.090)	(0.088)	(0.083)	(0.081)	(0.069)
Old for grade	−0.191	−0.184	−0.049	0.096	−0.177	−0.223	−0.032	−0.040	−0.150
	(0.143)	(0.152)	(0.142)	(0.144)	(0.145)	(0.165)	(0.132)	(0.137)	(0.111)
Small class	0.072	0.071	0.047	0.088*	−0.035	−0.037	−0.067	0.019	0.043
	(0.047)	(0.048)	(0.045)	(0.048)	(0.049)	(0.049)	(0.044)	(0.046)	(0.040)
Obs. (both panels)	1,628	1,628	1,628	1,628	1,731	1,731	1,731	1,731	2,589
Mean of dep. var.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Effect on KG math	−0.075	−0.075	−0.075	−0.075	−0.046	−0.046	−0.046	−0.046	−0.055
score in subsample	(0.061)	(0.061)	(0.061)	(0.061)	(0.057)	(0.057)	(0.057)	(0.057)	(0.051)

Notes: The table reports estimates from regressions of students' noncognitive skills, measured in fourth and eighth grade, on their exposure to repeaters in kindergarten. The outcome variables in columns (1)–(8) are indices summarizing teacher ratings of student behavior in four areas: effort, initiative, value, and discipline. The outcome variable in column (9) is a summary index of noncognitive skills that combines the available information from fourth and eighth grade for each student; see text for further details. Repeater exposure is measured as an indicator taking value 1 if the student's kindergarten class contains at least one repeater and 0 otherwise. All specifications control for an indicator for small class in kindergarten and kindergarten school fixed effects, with specifications in panel B additionally controlling for nonrepeating students' demographic characteristics. The bottom row reports estimates from regressions of the end-of-kindergarten math score on repeater exposure like in Table 2, column (2), with the difference that the sample is restricted to students observed with the outcome in the column head. Standard errors in parentheses allow for clustering at the kindergarten class level. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

TABLE 4. Repeater exposure in kindergarten and long-term educational attainment.

	High school GPA (1)	High school graduation (2)	Took ACT/SAT (3)	Summary index (4)
Panel A: No demographic controls				
Repeater exposure	0.698**	0.023*	0.034**	0.075**
	(0.334)	(0.013)	(0.016)	(0.031)
Panel B: With demographic controls				
Repeater exposure	0.552*	0.021*	0.033**	0.074***
	(0.308)	(0.013)	(0.015)	(0.028)
Male	−3.243***	−0.072***	−0.160***	−0.338***
	(0.298)	(0.013)	(0.012)	(0.024)
Black	−2.371***	0.011	0.022	0.025
	(0.641)	(0.029)	(0.027)	(0.055)
Free lunch	−3.406***	−0.141***	−0.284***	−0.590***
	(0.400)	(0.019)	(0.015)	(0.032)
Age in years	0.677	−0.031	0.079***	0.077*
	(0.494)	(0.021)	(0.019)	(0.040)
Old for grade	0.277	−0.054	−0.126***	−0.201***
	(0.812)	(0.042)	(0.037)	(0.077)
Small class	−0.063	0.007	0.010	0.043*
	(0.285)	(0.012)	(0.012)	(0.024)
Obs. (both panels)	2,438	2,955	6,039	6,039
Mean of dep. var.	84.20	0.87	0.41	0.00
Effect on KG math	−0.067	−0.131***	−0.090**	−0.090**
score in subsample	(0.054)	(0.050)	(0.041)	(0.041)

Notes: The table reports estimates from regressions that relate students' educational attainment, measured at the end of high school, to their exposure to repeaters in kindergarten. See the notes to Table 1 for descriptions of the outcome variables in columns (1)–(3). See text for details on the construction of the summary index of long-term educational attainment used as outcome in column (4). Repeater exposure is measured as an indicator taking value 1 if the student's kindergarten class contains at least one repeater and 0 otherwise. All specifications control for an indicator for small class in kindergarten and kindergarten school fixed effects, with specifications in panel B additionally controlling for nonrepeating students' demographic characteristics. The bottom row reports estimates from regressions of the end-of-kindergarten math score on repeater exposure like in Table 2, column (2), with the difference that the sample is restricted to students observed with the outcome in the column head. Standard errors in parentheses allow for clustering at the kindergarten class level. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

4.4. *Effects on Long-Term Educational Attainment*

The scholastic outcomes of participants in Project STAR were last tracked at the end of high school through collection of data on high school GPA, high school graduation, and college-test taking. Table 4 reports estimates from regressions that relate these measures of long-term educational attainment to students' exposure to repeaters in kindergarten. Panels A and B again present results from regressions without and with demographic controls, respectively. Focusing on the results in panel B, sharing a classroom with repeaters raises regular students' high school GPA by 0.6 on a scale from 0 to 100 (column (1)) and increases their propensity to graduate from high school by 2.1 percentage points (column (2)). Strikingly, repeater exposure also raises students' likelihood of taking a college entrance exam by 3.3 percentage points (column (3)), which corresponds to a sizeable 8% increase over the base rate of 41%.

Column (4) shows results from regressions of a summary index of these three long-term outcomes. To construct this index, I first normalize each outcome by subtracting its mean and dividing by its standard deviation. I then take the simple average of all available outcomes for each student, and normalize it again to have mean 0 and standard deviation 1. The resulting index should be interpreted as a broad indicator of educational attainment, as measured around the typical age of high school completion. As can be seen in Table 4, repeater exposure raises the index by 7.4% of a standard deviation, an estimate that is highly statistically significant. Thus, being exposed to repeaters in kindergarten has important benefits for students' long-term educational attainment.

4.5. *Heterogeneity Analysis*

An interesting question is whether the spillovers from repeaters documented previously affect all students equally. I begin exploring the potential heterogeneity of effects in Table 5. Panel A reports results from regressions of four key outcomes in which the repeater-exposure dummy is interacted with regular students' demographic characteristics.[17] An interesting pattern emerges: students who tend to do worse in school—males, black students, and students eligible for reduced-price lunch—appear to suffer larger initial declines in test scores and experience smaller gains in educational attainment if exposed to repeaters. However, none of these interactions is statistically significant at conventional levels.[18]

To increase statistical power, I combine the five demographic variables into an index of predicted ability as follows. For each student, I run a leave-me-out regression of the average of the end-of-kindergarten math and reading scores on these variables. I then compute the fitted values and standardize them to have mean 0 and standard deviation 1. Panel B of Table 5 shows results from regressions in which the repeater-exposure treatment is interacted with the thus constructed index of predicted ability. As expected, the coefficients on the interaction term are always positive, and they are more precisely estimated than those in panel A. The results show, for example,

17. For the sake of brevity, Table 5 and subsequent tables present estimates from regressions in which noncognitive skills and long-term educational attainment are measured by the respective summary indices. Results for individual long-term outcomes are presented in Online Appendix C and are discussed in Section 6.

18. Results are qualitatively and quantitatively similar if each demographic characteristic is interacted with repeater exposure in a separate regression, rather than in the same regression as done in Table 5. I also tested whether the impacts of repeaters differ by students' relative age, defined as the difference between own age and classmates' average age, but found little evidence of such heterogeneity.

TABLE 5. Heterogeneity by regular students' characteristics and by class size.

	Kindergarten math score (1)	8th-grade math score (2)	Noncog. index (3)	Long-term index (4)
Panel A: Heterogeneity by demographic characteristics				
Repeater exposure	−0.048	0.087*	0.147**	0.142***
	(0.054)	(0.051)	(0.059)	(0.044)
× male	−0.035	−0.025	−0.093	−0.029
	(0.047)	(0.056)	(0.075)	(0.049)
× black	−0.018	−0.034	−0.123	−0.112
	(0.079)	(0.083)	(0.121)	(0.070)
× free lunch	−0.056	0.000	0.104	−0.072
	(0.054)	(0.070)	(0.089)	(0.062)
× age in years	0.144	0.065	0.013	−0.061
	(0.092)	(0.109)	(0.138)	(0.082)
× old for grade	0.054	−0.261	0.159	0.134
	(0.160)	(0.225)	(0.243)	(0.150)
Panel B: Heterogeneity by predicted academic ability				
Repeater exposure	−0.098**	0.052	0.120**	0.064**
	(0.041)	(0.034)	(0.046)	(0.029)
× predicted ability	0.054*	0.033	0.005	0.059**
	(0.029)	(0.032)	(0.044)	(0.028)
Panel C: Heterogeneity by class size in kindergarten				
Repeater exposure	−0.078	0.089**	0.119**	0.065**
	(0.048)	(0.039)	(0.050)	(0.032)
× small class	−0.043	−0.099	−0.007	0.031
	(0.095)	(0.064)	(0.084)	(0.056)
Obs. (all panels)	5,614	4,353	2,589	6,039

Notes: The table reports estimates from regressions that probe for heterogeneous effects of repeater exposure by regular students' demographic characteristics (panel A), their predicted academic ability (panel B), and class size (panel C). Each column in each panel reports the results from a single regression in which the repeater-exposure dummy is interacted with the variables indicated in the leftmost column. All regressions control for students' demographic background, an indicator for small class in kindergarten, and kindergarten school fixed effects. See text for definitions of the outcome variables and predicted academic ability. Standard errors in parentheses allow for clustering at the kindergarten class level. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

that although repeater exposure raises long-term attainment by 6.4% of a standard deviation for students at the mean of predicted ability, this impact is close to zero (but still positive) for students with predicted ability one standard deviation below average. Thus, sharing a kindergarten classroom with repeaters may widen the attainment gap between students from different demographic backgrounds.

I next investigate whether the effects of repeater exposure differ along the distributions of four continuous outcomes: end-of-kindergarten and eighth-grade math scores, the noncognitive skills index, and high school GPA. Figure 4 shows results from quantile regressions that reveal that throughout these distributions, the impacts

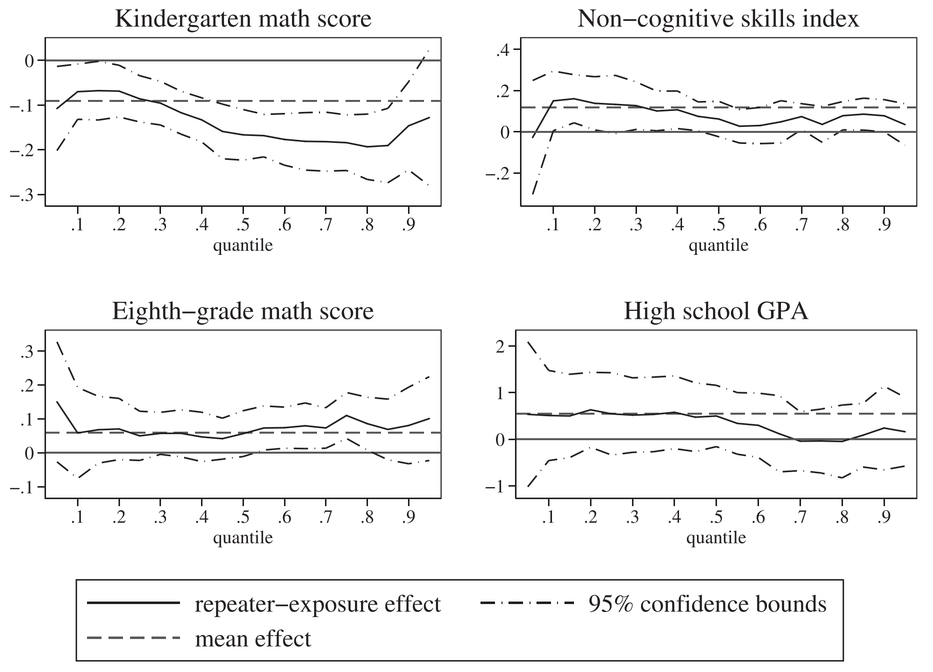

FIGURE 4. Quantile treatment effects. The figure shows coefficient estimates and 95% confidence bounds from quantile regressions of the four outcomes indicated previously each panel on repeater exposure in kindergarten. Estimation is based on the two-step estimator described in Canay (2011). Standard errors are obtained via boostrapping with 100 replications.

are qualitatively similar to the mean effects presented previously.[19] Thus, repeaters affect not only a small subset of their classmates; rather, the spillovers reflect general impacts for the whole class. At the same time, the estimates also suggest that students at lower quantiles experience smaller decreases in end-of-kindergarten math scores and somewhat larger gains in noncognitive skills and high school GPA. One potential explanation for this pattern is that the presence of repeaters leads teachers to shift their focus toward teaching skills that are particularly important for lower-achieving students, an interpretation that I discuss in more detail in the next section.[20]

19. The results are based on the fixed-effects quantile regression approach described in Canay (2011). This approach lets me account for differences in outcome distributions between schools, which is important because randomization into classes took place within schools in Project STAR. The additional assumption made compared to classical quantile regression is that school fixed effects affect the outcomes of all students within the same school in the same way, regardless of their location in the outcome distribution.

20. The finding that the negative effect on end-of-kindergarten math scores is both less pronounced for students at lower quantiles and more pronounced for students with low predicted ability might seem counter-intuitive. These findings are reconciled by the fact that the distributional impacts are similar for students with low and high predicted ability, but that the decrease in scores for the latter group is smaller throughout the achievement distribution. A similar argument holds for the results for high school GPA.

Finally, panel C of Table 5 reports estimates from specifications in which the effect of repeater exposure is allowed to vary with class size. To the extent that smaller classes allow teachers to better respond to the individual needs of each student, one might expect spillovers from repeaters to be attenuated in these classes, a conjecture that would notably be consistent with the well-known theoretical model by Lazear (2001). The empirical results do not lend support to this intuition: the estimated coefficients on the interaction terms are usually small relative to the main repeater-exposure effect, have different signs across different outcomes, and are always imprecisely estimated. In unreported regressions, I also confirmed that estimates are qualitatively similar, though less precise, when the entire empirical analysis is conducted separately for small and for regular-sized classes. Thus, there is little evidence that spillovers from repeaters differ by class size.

5. Discussion and Mechanisms

5.1. Noncognitive Skills as a Channel for Long-Term Impacts

Section 4 documents important spillovers from repeaters on their kindergarten classmates. Repeater-exposed students initially score lower on standardized tests, but this impact fades out rapidly after kindergarten and, if anything, turns positive in later grades. In contrast, there are lasting positive effects on noncognitive skills and on long-term educational attainment. Similar patterns of fading impacts on test scores but persistent effects on noncognitive skills and adult outcomes have recently been documented for other early childhood interventions, for example by Chetty et al. (2011) and Heckman et al. (2012).[21] Motivated by findings that noncognitive skills formed early in life are a key determinant of long-term educational success (e.g., Heckman et al. 2006), these authors argue that the differential accumulation of noncognitive skills by treated children is a key channel through which these interventions affected long-term outcomes. In this section, I provide suggestive evidence that exposure to repeaters in kindergarten similarly raises regular students' long-term educational attainment via improving their noncognitive skills. The following section then discusses in detail the possible ways in which sharing a classroom with repeaters may enhance these skills.

Table 6 reports results from a test of the hypothesis that improvements in noncognitive skills are driving the results for long-term outcomes. This test is based on the intuition that if the hypothesis holds, then controlling for intermediate noncognitive skills in the regression of long-term outcomes should substantially attenuate the estimated coefficient on repeater exposure. More formally, this augmented regression will yield the Average Controlled Direct Effect, that is, the effect of repeater exposure when noncognitive skills are fixed at the same value for all students

21. A notable difference to my analysis is that in these studies, the impacts on test scores and noncognitive skills go in the same direction.

TABLE 6. Repeater exposure, noncognitive skills, and long-term educational attainment.

	Summary index of long-term attainment			Difference
	(1)	(2)	(3)	[(2)–(3)]
Repeater exposure		0.060	0.012	0.048***
		(0.042)	(0.038)	[$p = 0.004$]
Noncog. skills (index)	0.408***		0.408***	
	(0.019)		(0.019)	
Observations	2,589	2,589	2,589	

Notes: The table reports estimates from regressions that relate students' educational attainment to their exposure to repeaters in kindergarten and to their noncognitive skills measured in fourth and eighth grade (columns (1)–(3)). Repeater exposure is measured as an indicator taking value 1 if the student's kindergarten class contains at least one repeater and 0 otherwise. See text for descriptions of the summary indices of long-term educational attainment and noncognitive skills. All specifications control for students' demographic background, an indicator for small class in kindergarten, and kindergarten school fixed effects. Regressions include all nonrepeating students for whom noncognitive skills are observed. Standard errors in parentheses allow for clustering at the kindergarten class level. The rightmost column reports results from a test of the null hypothesis that the coefficients on repeater exposure in columns (2) and (3) are equal. The p-value in brackets is based on a Wald test conducted after re-estimating the specifications in columns (2) and (3) using seemingly unrelated regression. $^{***}p < 0.01$.

(see Acharya et al. 2016). The additional assumption made compared to the main regressions is that there are no intermediate confounders, or no other variables that are affected by the treatment and that themselves affect noncognitive skills and long-term educational attainment.

Corroborating similar findings from previous studies, column (1) of Table 6 shows that noncognitive skills measured in fourth and eighth grade are highly predictive of long-term educational attainment. Column (2) replicates the estimated effect of repeater exposure on the summary index of long-term outcomes for the subsample of students observed with noncognitive skills. The coefficient is of similar size as in the main analysis given previously, but naturally less precisely estimated. This coefficient is reduced by 80% when the summary index of noncognitive skills is added to the regression as a control (column (3)), and the difference between this Average Controlled Direct Effect and the uncontrolled estimate is highly statistically significant (rightmost column). The evidence thus supports the hypothesis that the differential accumulation of noncognitive skills by exposed students during their childhood is an important channel through which repeater exposure affects long-term educational attainment.

To further evaluate this mechanism, I use the estimated impact of repeater exposure on noncognitive skills to predict its impact on long-term educational attainment and then compare this predicted impact to the actual estimate. From column (1) of Table 6, a one standard deviation increase in noncognitive skills is associated with a 40.8% of a standard deviation increase in the summary index of long-term outcomes. Repeater exposure raises noncognitive skills by 11.7% of a standard deviation (see Table 4, panel B, column (9)), which would thus predict a rise in long-term educational attainment of 4.8%. This figure is quite close to the actual estimated impact of 7.4%

reported in Table 4, lending further support to the hypothesis that noncognitive skills are the main mechanism for the effect of repeater exposure on long-term educational attainment.

5.2. Mechanisms for Impacts on Noncognitive Skills

How exactly does exposure to repeaters in kindergarten enhance regular students' noncognitive skills? One possibility is that the presence of such low-achieving and likely disruptive students in class leads teachers to emphasize the teaching of study skills and good classroom behavior. Such a shift in the focus of teaching could notably also explain the decrease in test scores at the end of kindergarten. A salient alternative explanation is that the observed pattern of results is due to selection effects. In particular, students may sort into classes or schools, or may leave the sample, in ways that are related to their exposure to repeaters. Finally, another possible mechanism is that repeater-exposed students benefit from additional resources in school, such as special tutoring. In what follows, I first present evidence that suggests that selection effects and additional resources do not cause the improvements in noncognitive skills of repeater-exposed students. I then discuss in more detail how the presence of a repeater can lead to learning of noncognitive skills by other students through behavioral adjustments by teachers, parents, or students.

Selection into Classes or Schools. Perhaps the most obvious explanation for the results reported previously is that they are due to a systematic pairing of repeaters with particular teachers or students in kindergarten. For example, school principals may assign low-achieving and disruptive repeaters to teachers who are relatively better at teaching noncognitive skills. Alternatively, students with low levels of persistence and discipline may select out of classes containing repeaters. However, the random assignment of teachers and students to kindergarten classes in Project STAR means that this mechanism cannot drive the results in this paper.

The fact that noncognitive skills are not observed in the data until the beginning of fourth grade opens up the possibility that the improvements in these skills do not happen until after kindergarten. This in turn means that selection in the later years of the experiment might be driving the results, for example if students who were exposed to repeaters in kindergarten systematically attend classes with better peers during grades 1–3. Again, the random assignment of teachers and students to classes throughout the duration of Project STAR severely limits the possibility of such sorting. The main way for students to select into particular classes is thus to change to another school. I tested whether repeater-exposed students are more likely to switch schools, but did not find any evidence of such behavior: in a regression of an indicator for leaving the experiment at any point after kindergarten on repeater exposure, the coefficient was 0.004 (standard error 0.014). It might still be the case, however, that repeater-exposed students tend to change to *better* schools than their nonexposed peers, and that this explains their improved noncognitive skills. To dispel such concerns, I re-estimated the main specifications for a sample of students who stayed in the same school throughout the experiment. Online Appendix Table C.1 shows that the results from these regressions were qualitatively and quantitatively similar to the ones in Section 4.[22]

As reported in Section 2, a few students also managed to change classes *within* Project STAR schools after kindergarten. Unfortunately, the data only report the actual class attended and not the randomly assigned class for each student, preventing me from testing directly whether such switching is related to repeater exposure. Instead, I checked whether exposure to repeaters in kindergarten predicts predetermined characteristics of students' classmates and teachers during grades 1–3. The results, which are presented in Online Appendix Table C.2, showed no systematic differences in class composition by repeater exposure, suggesting that between-class switching is not driving the effects reported in Section 4.

Selection out of the Sample. For reasons detailed in Online Appendix B, many of the outcomes studied in Section 4 are only observed for a subset of the students who attended kindergarten in Project STAR. Another possible explanation for the results is therefore that students select out of the sample based on their exposure to repeaters in kindergarten. In particular, the estimates might be picking up a so-called "healthy survivor effect:" if students who were negatively affected by repeaters in the short term are less likely to be observed with later outcomes, this could explain the positive effects on noncognitive skills and some of the long-term outcomes found previously.[23] A first piece of evidence against this explanation is provided in the lower panels of Tables 3 and 4, where I replicate the impact of repeater exposure on end-of-kindergarten math scores for the various subsamples of students observed with noncognitive skills and long-term outcomes. The coefficient estimates shown there are generally similar to the main estimate of –0.090 reported in Table 2, suggesting that the divergence of short- and long-term effects is not due to selective attrition.[24]

I investigate this matter more thoroughly in Table 7. Panel A reports estimates from regressions in which the dependent variables are indicators for being observed with six key outcomes. Across all regressions, the estimated coefficients on repeater exposure are close to zero and not statistically significant, suggesting that repeater-exposed students are not more likely to leave the sample. It might still be the case, however, that the composition of exposed versus nonexposed students changes across different outcomes. I test for such compositional changes by adding interactions between the

22. There is still the possibility that students change schools between third and fourth grade, that is, just before noncognitive skills are first measured. School identifiers in fourth grade are only observed for the subsample of students who took the Comprehensive Test of Basic Skills or who were selected for measurement of noncognitive skills in that year. This means that I can only identify 36% of regular students as staying in the same school throughout grades 1–4 with certainty. I nevertheless confirmed that results hold also in this reduced sample, although the effects are naturally less precisely estimated.

23. Notably, selection out of the sample cannot explain the impact of repeater exposure on college test-taking, which is observed for all students by construction.

24. The impact on end-of-kindergarten math scores is also similar for the subsamples of students observed with post-kindergarten test scores, see Online Appendix Table C.3.

TABLE 7. Exposure to repeaters in kindergarten and sample attrition.

	Kindergarten math score (1)	8th-grade math score (2)	4th-grade discipline (3)	8th-grade discipline (4)	High school GPA (5)	High school graduation (6)	Took ACT/SAT (7)
Panel A: Outcomes are indicators for being observed with the variable in the column head							
Repeater exposure	−0.011	0.009	−0.012	−0.020	0.011	0.007	
	(0.008)	(0.012)	(0.015)	(0.013)	(0.013)	(0.014)	
Panel B: Outcomes are indicators for being observed with the variable in the column head							
Repeater exposure	−0.016	0.033	0.016	−0.016	0.040*	0.023	
	(0.013)	(0.020)	(0.023)	(0.024)	(0.023)	(0.024)	
× male	0.005	−0.006	−0.013	−0.002	−0.017	0.014	
	(0.014)	(0.025)	(0.021)	(0.025)	(0.025)	(0.027)	
× black	−0.021	−0.004	−0.017	0.034	0.026	−0.036	
	(0.018)	(0.032)	(0.028)	(0.028)	(0.033)	(0.034)	
× free lunch	0.016	−0.050*	−0.036	−0.024	−0.068**	−0.040	
	(0.016)	(0.028)	(0.027)	(0.030)	(0.029)	(0.030)	
× age in years	0.001	−0.025	0.014	0.007	−0.042	−0.035	
	(0.026)	(0.042)	(0.039)	(0.042)	(0.041)	(0.043)	
× old for grade	0.049	0.012	−0.068	−0.048	0.023	0.080	
	(0.046)	(0.079)	(0.080)	(0.077)	(0.074)	(0.082)	
p-value (joint sign.)	0.47	0.52	0.40	0.60	0.26	0.48	
Panel C: Outcomes are the variables in the column heads, sample is restricted to nonattritors							
Repeater exposure	−0.081	0.072*	0.160**	0.228***	0.427	0.010	0.038*
	(0.057)	(0.041)	(0.072)	(0.055)	(0.332)	(0.012)	(0.022)

Notes: The table reports results from regressions that test for selective attrition by exposure to repeaters in kindergarten. In panels A and B, the dependent variables are indicators taking value 1 if the outcome in the column head is observed for a given student and 0 otherwise ($N = 6{,}039$ nonrepeating students). No results are presented for ACT/SAT test-taking because by construction, this outcome is observed for all students. The last row in panel B reports p-values from F-tests for joint significance of the repeater-exposure dummy and the five interaction terms. In panel C, the dependent variables are the variables listed in the column heads. In this panel, the sample is restricted to the 2,100 nonrepeating students who are observed with the outcomes in columns (1), (2), (5), (6), and (7) (sample sizes for the regressions in columns (3) and (4) are 861 and 1,015, respectively). Regressions in all three panels control for students' demographic background, an indicator for small class in kindergarten, and kindergarten school fixed effects. Standard errors in parentheses allow for clustering at the kindergarten class level. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

repeater-exposure dummy and the five demographic characteristics to the specifications from panel A. The corresponding estimates in panel B reveal mostly small coefficients on both the main effect and the interaction terms, which are always jointly insignificant. This suggests that there are no systematic differences between exposed and nonexposed students observed with different outcomes. Finally, to dispel any remaining concerns about selective attrition driving my results, I re-estimate the main specifications for a consistent sample of nonattritors. As panel C shows, the results are qualitatively and quantitatively similar to those found for the unrestricted sample in Section 4.[25]

Access to Additional Resources. Yet another explanation for the enhanced noncognitive skills of repeater-exposed students is that these students differentially benefited from additional resources in school. The experimental setup of Project STAR severely limits this possibility: for example, schools would not have been allowed to place additional teaching aides into classes containing low-achieving repeaters. One way in which students could nevertheless have profited from additional resources is via special education and special instruction programs, which were not controlled by the experiment. Students in such programs might for example get individualized study plans or have access to different learning materials, which may foster noncognitive skills. Thus, to the extent that repeater-exposed students are more likely to enter special education or special instruction programs, this might explain their differential accumulation of such skills. Similarly, if repeaters are more likely than other students to participate in such programs, there might be positive externalities on their classmates.

Participation in special education and special instruction programs was recorded for students in Project STAR in kindergarten and in first grade, which allows me to test the potential channels outlined in the previous paragraph. Online Appendix Table C.4 shows that regular students who were exposed to repeaters in kindergarten are not more likely to participate in special education or special instruction programs. Although repeaters are more likely than regular students to participate in such programs in kindergarten (23% vs. 6% participation rate for both programs combined), Online Appendix Table C.5 shows that results are qualitatively and quantitatively similar if classes with participating repeaters are excluded from the sample. Thus, there is no evidence that differential access to additional resources is behind the improvement in noncognitive skills of repeater-exposed students.[26]

25. In panel C, I define nonattritors as students who are observed with the following outcomes: kindergarten math score, eighth-grade math score, high school GPA, high school graduation, and college test-taking. I did not include fourth- and eighth-grade noncognitive skills in this list because these skills were measured only for two different random subsamples of students, such that only 493 regular students are observed with all outcomes in Table 7. In Online Appendix Figure C.2, I show that test score impacts are qualitatively similar for the subsample of students observed with noncognitive skills (the effect on educational attainment for this subsample is reported in Table 6).

26. Another way in which repeater-exposed students might profit from additional resources is if they are more likely to repeat themselves. In a regression of an indicator for being below grade in 1994, when students were supposed to be in eighth grade, on repeater exposure, the coefficient was 0.013 (standard error 0.016), suggesting that this is not the case.

Behavioral Responses by Teachers, Students, or Parents. A final explanation is that teachers, students, or parents react to the presence of repeaters in the classroom in a way that promotes the accumulation of noncognitive skills. A particularly salient possibility is that teachers emphasize the learning of basic behavioral and study skills in response to class disruption by repeaters. Such disruption could take the form of misbehavior that directly distracts other students, or could be due to repeaters' exceptionally low cognitive ability that diverts teacher resources and slows down the pace of instruction. Teachers may react to such disruption, for example, by changing their teaching practices or by explicitly focusing their attention on the subgroup of low-achieving students. Importantly, such adjustments could account for both the drop in end-of-kindergarten test scores and the improvements in noncognitive skills seen in the data.

To test this teacher-adaptation mechanism, one would ideally observe teaching practices and lesson content, perhaps from teacher time use surveys and teacher logs. Unfortunately, such data are not available for teachers in Project STAR. However, the explanation outlined in the previous paragraph is notably consistent with the distributional effects of repeater exposure discussed in Section 4. Recall that those results suggested that students in the lower quantiles of the respective distribution experience smaller decreases in the end-of-kindergarten math score and larger gains in noncognitive skills and high school GPA when exposed to repeaters. This pattern could well be due to teachers emphasizing basic behavioral skills, as high-achieving students likely already possess those skills and would have profited more from learning "actual math and reading." The quantile regression results are also more generally consistent with the idea that the presence of repeaters leads teachers to focus on students in the lower part of the achievement distribution.

Additional support for the teacher-adaptation mechanism comes from the educational psychology literature, which confirms that teachers adjust their teaching practices to students' cognitive ability and behavior (e.g., Corno 2008; Nurmi et al. 2013). More specifically, teachers tend to pay more attention and give additional instructional support to low-achieving students (Babad 1990; Kiuru et al. 2015). They are also more likely to establish explicit rules for behavior and stable routines in the classroom, and to use teacher-directed practices, if the average level of student achievement is low (Pakarinen et al. 2011; Kikas et al. 2018).[27] This tendency could notably explain the higher discipline of students in classrooms with low-achieving repeaters found in Project STAR.

Although changes in teachers' behavior are a particularly salient explanation for the results in this paper, behavioral adjustments by students or parents might also be at work. For example, Hill (2014) suggests that students may see repeaters as examples of failure and may therefore exert more effort in order not to fail themselves.

27. Teachers using teacher-directed practices demand that students follow strict rules in the classroom and tend to focus on the practice of basic skills. In contrast, child-centered practices give students more freedom to choose their own way of learning. For further details about this dichotomy, see, for example, Kikas et al. (2018).

Alternatively, parents of repeater-exposed students may compensate for a worse classroom environment by helping their children more at home, for example, by paying for private tuition. Both of these reactions may lead to the observed improvements in noncognitive skills, and the lack of data on student views and parental inputs does not let me distinguish between such different behavioral reactions.

6. Robustness of Results

6.1. Alternative Measures of Repeater Exposure

The main analysis in Sections 4 and 5 distinguishes between classes with and without repeaters, but does not further differentiate classes according to the actual number of repeaters. In Table A.2, I explore whether the results are sensitive to this particular definition of treatment. Panel A shows estimates from regressions of four main outcomes on separate indicators for being in class with one, two, and three to five repeaters. Across all specifications, the estimated impacts of exposure to one and exposure to two repeaters are qualitatively and quantitatively similar to the main effects reported in Section 4. Although the coefficients on exposure to three to five repeaters are smaller in absolute value, they are very imprecisely estimated and not statistically different from these effects either.

Panel B shows that using the class share of repeaters as treatment again yields results that are qualitatively similar to the main results. Panel C reports results from specifications that include both the repeater-exposure dummy and the share of repeaters as regressors. Although the estimated coefficients on the dummy are roughly similar to the ones reported in Section 4, the coefficients on the share are no longer significant in these regressions. This suggests that within the small range of the number of repeaters per class observed in this sample, the extensive margin (being in class with one vs. no repeaters) is more important than the intensive margin (being in class with one vs. several repeaters). This finding lends further support to the use of the repeater-exposure dummy as the main treatment variable.

6.2. Impacts of Repeaters versus Classroom Demographics

One potential concern with the treatment is that it might simply be picking up the impact of changes in the average demographic composition of peers, a measure that has been widely used to identify peer effects in the previous literature.[28] Online Appendix Table C.6 reports results from regressions of four main outcomes that additionally control for average demographic characteristics of each students' classmates, including repeaters. To the extent that the impacts documented in Section 4 operate via changing the

28. For example, Hoxby (2000) and Whitmore (2005) study the impacts of the share of female students in the classroom. Cascio and Schanzenbach (2016) study the impacts of the average age of students' kindergarten classmates in Project STAR.

demographic composition of classmates, one would expect the estimated coefficient on repeater exposure to be attenuated in these regressions. This is not the case: controlling for share of male classmates, share of black classmates, share of free-lunch classmates, classmates' average age, or share of old-for-grade classmates does not significantly alter the impacts of repeater exposure, indicating that they operate over and above any of these potential channels. This suggests once again that the focus on repeaters allows me to better identify truly low-achieving peers.[29]

6.3. Relative Measurement of Noncognitive Skills

Table 3 reports positive impacts from repeater exposure in kindergarten on regular students' noncognitive skills. A potential concern with these findings is that these improvements might simply reflect higher teacher ratings of students' behavior *relative* to the behavior of repeaters in the same class. I address this concern in Online Appendix Table C.7. In panel A, I re-estimate the impacts of repeater exposure on fourth-grade noncognitive skills for the subsample of students whose fourth-grade classes did not contain any of the 193 original kindergarten repeaters. The effects of repeater exposure in these regressions are somewhat attenuated compared to those reported in Table 3 but qualitatively similar. Although the data do not allow me to observe classroom composition during eighth grade, I can restrict the sample to students who at that time attended a *school* that did not contain any of the original repeaters (most students had switched to a different middle school by eighth grade). Panel B shows that the impacts of repeater exposure on noncognitive skills in this restricted sample are very similar to the ones reported in Table 3. Thus, the evidence does not support the idea that the positive impacts of repeater exposure on noncognitive skills capture purely mechanical effects due to relative teacher ratings.

6.4. Additional Results for Disaggregated Post-Kindergarten Outcomes

The regressions probing for heterogeneity in Section 4, the analysis of mechanisms in Section 5, and the robustness checks presented in the previous sections measure noncognitive skills and long-term educational attainment using the respective summary indices. As discussed previously, these indices are appealing because they increase statistical power to detect effects that go in the same direction within a domain

29. As can be expected from the demographic characteristics of repeaters shown in Table 1, being exposed to repeaters is associated with having a higher share of male classmates, a higher share of free-lunch classmates, a higher average age of classmates, and a higher share of old-for-grade classmates. Across the regressions in Online Appendix Table B.6, the only statistically significant impact of classroom demographics is a negative effect of the share of male classmates on end-of-kindergarten math scores, which confirms previous results by Whitmore (2005). Cascio and Schanzenbach (2016) use an instrumental-variable strategy to examine the impacts of classmates' average age; in specifications that apply the same strategy, I found some significant impacts of classmates' average age, which corroborate their results. Importantly, however, the estimated coefficients on repeater exposure from these regressions were again qualitatively and quantitatively similar to the ones reported in Section 4.

and because they allow me to present results in a concise manner. For transparency, however, the Online Appendix also presents results for more disaggregated outcomes, which I summarize here.

Online Appendix Table C.8 reports estimates of the impact of repeater exposure on the 41 individual behaviors that make up the eight noncognitive skill indices used in Table 3. Similarly to the results shown there, the estimates reveal positive impacts of repeater exposure on the vast majority of student behaviors in both fourth and eighth grade. Online Appendix Table C.9 replicates most of the results presented in all previous tables for a set of eleven (relatively) disaggregated outcomes: the eight fourth- and eighth-grade noncognitive skills measures, high school GPA, high school graduation, and college test-taking. The vast majority of estimates presented there are qualitatively similar to the ones presented in the main text, although they are often less precisely estimated. This confirms that by aggregating outcomes into the two indices, I do not inordinately ignore heterogeneous impacts of repeater exposure.

6.5. Mechanical Spillover Effects

In a recent paper, Angrist (2014) documents a mechanical bias in peer-effects regressions that arises if students both provide treatment for other students and are subject to treatment from these other students themselves.[30] Intuitively, this bias is avoided here due to the clear separation of initiators and recipients of spillover effects. I confirmed this intuition in a simulation-based falsification test. In particular, I exchanged each student's classmates with a new set of peers randomly drawn from other classes in the same school. In this way, all students were assigned to a group of placebo classmates with whom they did not interact in their real-world classroom. I then re-estimated the effect of repeater exposure, measured using the placebo classmates, on kindergarten math scores. Any effect of repeater exposure in this regression reflects purely mechanical forces. In 1,000 replications of this exercise, the median coefficient on repeater exposure was 0.017 with a 90% empirical confidence interval of [−0.028, 0.063], which excludes the coefficient of −0.090 found in the actual data.

7. Conclusion

Many education policies change the grouping of students into classes and schools, but little is known about the long-term impacts of school peers. This paper provides some of the first evidence on such impacts by evaluating how sharing a kindergarten classroom with low-achieving repeaters affects regular students' test scores, their noncognitive skills, and their long-term educational attainment.

30. One reason why this bias may arise is measurement error in peer quality. In a recent paper, Feld and Zölitz (2017) show that in settings such as Project STAR where students are randomly assigned to groups, this bias leads to an attenuation of peer effects estimates rather than to an overestimation.

The empirical analysis exploits the random assignment of teachers and students to classes in Project STAR in order to estimate causal spillover effects. Regular students who are exposed to repeaters in their kindergarten class perform significantly worse on a standardized math test at the end of kindergarten. However, these students display substantially improved noncognitive skills, such as effort and discipline, when these are first measured at the beginning of fourth grade. Although the negative spillovers from repeaters on test scores fade out rapidly after kindergarten and, if anything, turn positive, the gains on noncognitive skills persist over time. The favorable development of repeater-exposed students culminates in significantly raised propensities to graduate from high school and to take a college entrance exam around the age of 18. I argue that these positive long-term impacts are likely due to the differential accumulation of noncognitive skills by exposed students. As for how exactly repeater exposure raises these skills, the results are consistent with a mechanism of teachers adapting their teaching practices and lesson content, even though the data do not allow me to test this mechanism directly.

The striking divergence of the impacts of repeater exposure on short-term test scores and long-term educational attainment highlights the importance of studying the long-term effects of educational interventions. By themselves, the negative short-term spillovers on test scores would have suggested that policies that separate low-achieving repeaters from regular first-time students would greatly benefit the latter, who make up the vast majority of the student population in schools. However, this conclusion has to be reversed once long-term impacts are taken into account. Indeed, the overall results show that mixing students of very different abilities at an early age can be beneficial for most students in the long term.

Appendix: Additional Tables

TABLE A.1. Randomization tests.

	Student characteristics					Teacher characteristics		
	Male	Black	Free lunch	Age in years	Old for grade	Black	Experience	Master's degree
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Controlling for school fixed effects								
Repeater exposure	−0.005	−0.001	0.004	0.001	−0.003	−0.030	−0.686	−0.098*
	(0.015)	(0.007)	(0.015)	(0.009)	(0.005)	(0.038)	(0.679)	(0.059)
Panel B: Controlling for school fixed effects and class size								
Repeater exposure	−0.006	−0.002	0.001	0.004	−0.001	−0.035	−0.830	−0.113*
	(0.015)	(0.007)	(0.015)	(0.009)	(0.005)	(0.039)	(0.696)	(0.061)
Obs. (both panels)	6,039	6,039	6,039	6,039	6,039	5,998	6,018	6,018

Notes: The table reports estimates from regressions of characteristics of nonrepeating students' and their kindergarten teachers on repeater exposure in kindergarten. Repeater exposure is measured as an indicator taking value 1 if the student's kindergarten class contains at least one repeater and 0 otherwise. All specifications in panels A and B control for kindergarten school fixed effects, and specifications in panel B additionally control for an indicator for small class in kindergarten. Sample sizes are reduced in columns (6)–(8) due to missing information on teacher characteristics. Mean of dependent variable in column (6)/(7)/(8): 0.17/9.26/0.35. Standard errors in parentheses allow for clustering at the kindergarten class level. $^{*}p < 0.10$.

TABLE A.2. Alternative measures of repeater exposure.

	Kindergarten math score (1)	8th-grade math score (2)	Noncog. index (3)	Long-term index (4)
Panel A: Indicators for different numbers of repeaters				
1 repeater in class	−0.096**	0.070*	0.120***	0.072**
	(0.046)	(0.036)	(0.046)	(0.031)
2 repeaters in class	−0.092	0.039	0.134**	0.093**
	(0.070)	(0.053)	(0.065)	(0.041)
3–5 repeaters in class	−0.021	0.019	0.025	0.030
	(0.103)	(0.090)	(0.090)	(0.063)
Panel B: Linear share of repeaters in class				
Share of repeaters	−0.601	0.370	1.045**	0.781**
	(0.483)	(0.406)	(0.445)	(0.310)
Panel C: Exposure dummy and linear share of repeaters				
Repeater exposure	−0.135**	0.096*	0.123*	0.056
	(0.068)	(0.057)	(0.068)	(0.045)
Share of repeaters	0.659	−0.525	−0.080	0.256
	(0.802)	(0.677)	(0.739)	(0.484)
Obs. (all panels)	5,614	4,353	2,589	6,039

Notes: The table reports estimates from regressions that probe the robustness of results to using alternative measures of repeater exposure. In panel A, the repeater-exposure dummy is replaced by dummies for 1, 2, and 3–5 repeaters in class. Specifications in panel B include the class share of repeaters as treatment instead. Specifications in panel C include both the repeater-exposure dummy and the class share of repeaters. See text for details on the construction of the outcome variables. All regressions control for students' demographic background, an indicator for small class in kindergarten, and kindergarten school fixed effects. Standard errors in parentheses allow for clustering at the kindergarten class level. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

References

Acharya, Avidit, Matthew Blackwell, and Maya Sen (2016). "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects." *American Political Science Review*, 110, 512–529.

Angrist, Joshua D. (2014). "The Perils of Peer Effects." *Labour Economics*, 30, 98–108.

Babad, Elisha (1990). "Measuring and Changing Teachers' Differential Behavior as Perceived by Students and Teachers." *Journal of Educational Psychology*, 82, 683–690.

Bertrand, Marianne and Jessica Pan (2013). "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior." *American Economic Journal: Applied Economics*, 5, 32–64.

Bifulco, R., J. M. Fletcher, and S. L. Ross (2011). "The Effect of Classmate Characteristics on Post-Secondary Outcomes: Evidence from the Add Health." *American Economic Journal: Economic Policy*, 3, 25–53.

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, 119–153.

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

Canay, Ivan A. (2011). "A Simple Approach to Quantile Regression for Panel Data." *The Econometrics Journal*, 14, 368–386.

Carrell, Scott E. and Mark L. Hoekstra (2010). "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids." *American Economic Journal: Applied Economics*, 2, 211–228.

Cascio, Elizabeth U. and Diane Whitmore Schanzenbach (2016). “First in the Class? Age and the Education Production Function.” *Education Finance and Policy*, 11, 225–250.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan (2011). “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR.” *The Quarterly Journal of Economics*, 126, 1593–1660.

Corno, Lyn (2008). “On Teaching Adaptively.” *Educational Psychologist*, 43, 161–173.

Deming, David and Susan Dynarski (2008). “The Lengthening of Childhood.” *Journal of Economic Perspectives*, 22(3), 71–92.

Feld, Jan and Ulf Zölitz (2017). “Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects.” *Journal of Labor Economics*, 35, 387–428.

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

Finn, Jeremy D., Jayne Boyd-Zaharias, Reva M. Fish, and Susan B. Gerber (2007). “Project STAR and Beyond: Database User’s Guide.” Report, HEROS Incorporated.

Gottfried, Michael A. (2013). “The Spillover Effects of Grade-Retained Classmates: Evidence from Urban Elementary Schools.” *American Journal of Education*, 119, 1–64.

Gould, Eric D., Victor Lavy, and M. Daniele Paserman (2009). “Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence.” *The Economic Journal*, 119, 1243–1269.

Graham, B. S. (2008). “Identifying Social Interactions through Conditional Variance Restrictions.” *Econometrica*, 76, 643–660.

Heckman, James J., Rodrigo Pinto, and Peter Savelyev (2012). “Understanding the Mechanisms through which an Influential Early Childhood Program Boosted Adult Outcomes.” *American Economic Review*, 103(6), 2052–2086.

Heckman, James J., Jora Stixrud, and Sergio Urzua (2006). “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior.” *Journal of Labor Economics*, 24, 411–482.

Hill, Andrew J. (2014). “The Costs of Failure: Negative Externalities in High School Course Repetition.” *Economics of Education Review*, 43, 91–105.

Horoi, Irina and Ben Ost (2015). “Disruptive Peers and the Estimation of Teacher Value Added.” *Economics of Education Review*, 49, 180–192.

Hoxby, Caroline (2000). “Peer Effects in the Classroom: Learning from Gender and Race Variation.” NBER Working Paper 7867. National Bureau of Economic Research, Cambridge, MA.

Jimerson, Shane R. (2001). “Meta-Analysis of Grade Retention Research: Implications for Practice in the 21st Century.” *School Psychology Review*, 30, 420–437.

Karweit, Nancy L. (1999). “Grade Retention: Prevalence, Timing, and Effects.” Report, Center for Research on Students Placed at Risk, Johns Hopkins University.

Kautz, Tim, James J. Heckman, Ron Diris, Bas ter Weel, and Lex Borghans (2014). “Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success.” OECD Education Working Paper 110. Organisation for Economic Co-operation and Development, Paris, France.

Kikas, Eve, Eija Pakarinen, Piret Soodla, Kätlin Peets, and Marja-Kristiina Lerkkanen (2018). “Associations Between Reading Skills, Interest in Reading, and Teaching Practices in First Grade.” *Scandinavian Journal of Educational Research*, 62, 832–849.

Kiuru, Noona, Jari-Erik Nurmi, Esko Leskinen, Minna Torppa, Anna-Maija Poikkeus, Marja-Kristiina Lerkkanen, and Pekka Niemi (2015). “Elementary School Teachers Adapt their Instructional Support According to Students’ Academic Skills: A Variable and Person-Oriented Approach.” *International Journal of Behavioral Development*, 39, 391–401.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz (2007). “Experimental Analysis of Neighborhood Effects.” *Econometrica*, 75, 83–119.

Krueger, Alan B. (1999). “Experimental Estimates of Education Production Functions.” *The Quarterly Journal of Economics*, 114, 497–532.

Krueger, Alan B. and Diane M. Whitmore (2001). "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR." *The Economic Journal*, 111, 1–28.

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

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

Lazear, Edward P. (2001). "Educational Production." *The Quarterly Journal of Economics*, 116, 777–803.

Nurmi, Jari Erik, Noona Kiuru, Marja Kristiina Lerkkanen, Pekka Niemi, Anna Maija Poikkeus, Timo Ahonen, Esko Leskinen, and Anna Liisa Lyyra (2013). "Teachers Adapt their Instruction in Reading According to Individual Children's Literacy Skills." *Learning and Individual Differences*, 23, 72–79.

Pagani, L., R. E. Tremblay, F. Vitaro, B. Boulerice, and P. McDuff (2001). "Effects of Grade Retention on Academic Performance and Behavioral Development." *Development and Psychopathology*, 13, 297–315.

Pakarinen, Eija, Marja-Kristiina Lerkkanen, Anna-Maija Poikkeus, Martti Siekkinen, and Jari-Erik Nurmi (2011). "Kindergarten Teachers Adjust their Teaching Practices in Accordance with Children's Academic Pre-skills." *Educational Psychology*, 31, 37–53.

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, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Elsevier, Amsterdam, pp. 249–277.

Schanzenbach, Diane Whitmore (2006). "What have Researchers Learned from Project STAR?" *Brookings Papers on Education Policy*, 9, 205–228.

Sojourner, Aaron (2013). "Identification of Peer Effects with Missing Peer Data: Evidence from Project STAR." *The Economic Journal*, 123, 574–605.

Whitmore, Diane (2005). "Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment." *American Economic Review*, 95(2), 199–204.

Word, Elizabeth, John Johnston, Helen P. Bain, Dewayne B. Fulton, Charles M. Achilles, Martha N. Lintz, John Folger, and Carolyn Breda (1990). "The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Technical Report 1985–1990." Report, Tennessee State Department of Education.

Supplementary Data

Supplementary data are available at *JEEA* online.