The Long-Term Impacts of Low-Achieving Childhood Peers: Evidence from Project STAR[☆]

Jan Bietenbeck

Lund University and IZA

September 2016

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 standardized tests at the end of kindergarten, an effect that fades out in later grades; (2) show persistent improvements in non-cognitive 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 eighteen. I argue that the positive spillovers on long-term educational attainment are driven by the differential accumulation of non-cognitive skills by repeater-exposed students during childhood. The improvements in these skills in turn are likely due to behavioral adjustments by teachers, students, or parents to the presence of low-achieving repeaters in the classroom.

[☆]I am greatly indebted to David Dorn for his guidance and support. I thank Joseph Altonji, Manuel Arellano, David Autor, Manuel Bagues, Sascha Becker, Samuel Bentolila, Susan Dynarski, Joshua Goodman, Bryan Graham, Claudio Michelacci, Pedro Mira, Magne Mogstad, Luca Repetto, Jan Stuhler, Simon Wiederhold, two 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. Author contact details: Department of Economics, Lund University, P.O. Box 7082, 220 07 Lund, Sweden; jan.bietenbeck@nek.lu.se

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, Silva, and Weinhardt, 2012). The vast majority of these papers has focused exclusively on short-term spillover effects on contemporaneous outcomes, such as the impact of kindergarten classmates on test scores at the end of kindergarten. However, in order 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 below, they are characterized by exceptionally low cognitive and non-cognitive 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 non-cognitive 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 reduces students' test scores at the end of kindergarten, a result which corrobo-

[1] I use the terms 'regular student,' 'non-repeating student,' and 'first-time kindergarten student' interchangeably throughout the paper.

rates previous findings of negative short-term spillovers from low-achieving peers. Repeater exposure however substantially *increases* students' non-cognitive skills such as effort and discipline, which are first measured at the beginning of fourth grade. While the negative spillovers from repeaters on test scores fade out rapidly after kindergarten, the gains in non-cognitive skills persist over time. Consistent with this last result, 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 potential mechanisms behind these results. Motivated by recent findings that non-cognitive skills formed early in life are a key determinant of long-term educational success (e.g. Heckman, Stixrud, and Urzua, 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 non-cognitive skills, the experimental setup of Project STAR lets me rule out a wide range of potential explanations that involve selection of students into classes or schools. This leaves behavioral responses by teachers, students, or parents as a likely channel. For instance, teachers whose classes are frequently disrupted by repeaters may focus on teaching students good classroom behavior and study skills. While this explanation has some support in the educational psychology literature, the data do not allow me to test it directly.

This paper contributes to a large existing literature on peer effects in schools, which was recently reviewed by Sacerdote (2011). This literature includes studies based on data from 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. Like almost all previous research, these studies focus exclusively on short-term impacts. Two more recent studies have expanded the research agenda on the effects of early childhood peers to include long-term outcomes: Gould, Lavy, and Paserman (2009) investigate how 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]

To the best of my knowledge, this paper provides 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 spillovers is important because kindergarten students are at an age where both cognitive and non-cognitive 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 studying such long-term impacts is further highlighted by the finding that short- and long-term effects do not necessarily go in the same direction.

While I interpret my results as reflecting the impacts of childhood exposure to low-achieving peers more generally, examining spillovers from repeaters is also interesting in its own right. Grade retention rates in kindergarten and the early elementary grades have risen steadily over the past few decades in the United States (Hauser, Frederick, and Andrew, 2007), with 6% of kindergarten students repeating the grade during the 2010-11 school year (NCES, 2013). Whereas the consequences of retention for repeaters have been widely studied, little is known about how retention policies affect the learning outcomes of non-retained students.[3] By studying spillovers from kindergarten repeaters on their classmates, this paper provides some of the first evidence on such side effects of retention.

[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 this 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 is moreover related to two studies by Bifulco, Fletcher, and Ross (2011) and Black, Devereux, and Salvanes (2013), which examine spillovers from high school peers on longer-term educational and labor market outcomes.

[3]A large literature in educational psychology has found that being retained lowers students' academic achievement and may increase their disruptive behavior (e.g. Jimerson, 2001; Pagani et al., 2001). Recently, a small number of studies by economists have documented negative short-term spillovers from repeaters on the academic achievement of their classmates (Lavy, Paserman, and Schlosser, 2012; Gottfried, 2013; Hill, 2014).

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 1985-86 school year, 6,325 kindergarten students in 79 participating Tennessee schools were randomly assigned to small (target size 13-17 students) or regular-sized (22-25 students) classes within their schools.[4] Students were supposed to stay in their assigned class type (small versus 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, including repeaters, 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). Early analyses of Project STAR were reluctant to conclude that this was indeed the case, mainly because the STAR Technical Report (Word et al., 1990) does not describe the exact procedure by which students were allocated to specific classes. However, several more recent studies (Chetty et al., 2011; Sojourner, 2013; Cascio and Schanzenbach, 2016) also rely on random assignment of students and teachers to classes in Project STAR and provide new evidence in support of this assumption. Section 3 revisits some of this evidence and provides statistical support for the claim that repeaters were randomly assigned to kindergarten classes within schools.

The eventual implementation of Project STAR differed somewhat from the original experimental design. Three of these differences are particularly important in the context of this paper. First, as the initial cohort of students 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

[4]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.

who had participated in the experiment in kindergarten had left the sample. Second, 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. Third, while 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. In contrast, selective attrition from the sample could potentially confound some of the estimates on long-term outcomes. I provide extensive evidence that this is not the case in Section 5 below. Finally, the three aspects of the implementation mentioned in the previous paragraph change the total amount of time that students spent in class with a kindergarten repeater. This affects the interpretation of the repeater-exposure treatment, a point that I discuss in more detail in the following subsection.[5]

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 below is based, combines these data such that students can be followed throughout their scholastic careers until the end of high school.[6] This subsection gives a brief overview of the dependent and independent variables used in the empirical analysis. Online appendix A provides additional details on data collection procedures and on the construction of outcome variables.

Demographic characteristics. The data contain information on students' gender, race, eligibility for free or reduced-price lunch, and exact date of

[5] 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).

[6] Data on some of the outcomes studied in this paper were generously provided to me by Diane Schanzenbach; see Online appendix A for details.

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 which 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.[7]

Kindergarten repeaters. The data include an indicator for whether each student was repeating kindergarten in the 1985-86 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 below focuses on spillover effects from the 193 old-for-grade repeaters, who first entered kindergarten at the regular entry age. While 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.[8]

Repeater exposure. Figure 1 shows the distribution of repeaters across classes in the 60 schools with at least one repeater.[9] 126 of the 254 classes

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

[8]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 non-cognitive 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.

[9]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 versus 83 students), are less likely

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 above, students who were exposed to repeaters in 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.[10] 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

to be located in the inner city (12% versus 47% of schools), and contain lower fractions of black students (20% versus 61%) and low-income students (41% versus 67%).

[10]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.

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 below 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 1-5 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 non-cognitive skills in the empirical analysis below.

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 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 (1984-85) school year and are thus subject to a fundamentally different treatment. Finally, while schools without repeaters do not contribute to the identification of spillover effects in this paper, they are kept in the sample in order to increase the precision of the estimated impacts of other covariates included in the regressions. 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 non-repeating 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 non-repeating students. Repeaters are also older than non-repeating 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 non-repeating 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 non-cognitive skill measures to have mean 0 and standard deviation 1 across non-repeating 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.[11] 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).[12] That literature has moreover sustained 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 low cognitive and non-cognitive skills of repeaters might well impact the learning of their classmates, an issue that I investigate empirically below.

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 below, 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 non-experimental data, this requirement will not be met if, for example, school principals assign low-achieving repeaters to classes with high-achieving other students or teachers. In contrast, random assignment in Project STAR ensures that classes with and without repeaters are balanced on characteristics of regular students and teachers.

[11] Repeaters' measured cognitive and non-cognitive skills are also significantly below those of male students, black students, and students eligible for free or reduced-price lunch (results are available upon request). 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, Paserman, and Schlosser (2012).

[12] 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 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).

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.[13] 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 negatively biased. I avoid such bias by controlling for class size in all of my regressions. Below, 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:

$$y_{ics} = \alpha_s + \beta_1 \text{EXPOSURE}_{cs} + \beta_2 \text{SMALL}_{cs} + X_{ics}\gamma + \varepsilon_{ics}, \qquad (1)$$

where y_{ics} is a kindergarten or long-term outcome for non-repeating student i randomly assigned to kindergarten class c in school s, 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).

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 studies of the experiment (e.g. Chetty

[13]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.

et al., 2011; Cascio and Schanzenbach, 2016), which show that classes are balanced on a wide range of student demographics 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. I then computed the within-school standard deviation in repeater exposure, which is a summary measure of the identifying variation used in this paper, in the re-randomized data. Across 1,000 replications, the median standard deviation was 0.381 with a narrow 90% empirical confidence interval of [0.369, 0.391]. This confidence interval comfortably contains the within-school standard deviation of 0.383 observed 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 the joint significance of the class fixed effects was 0.65, suggesting that repeater status is indeed balanced across classes.

Finally, I tested whether being exposed to a repeater predicts non-repeating students' observable characteristics. To this end, I regressed the five demographic characteristics available in the data on the repeater-exposure dummy and school fixed effects, with alternative specifications additionally controlling for class size as in equation 1. As shown in Online Appendix Table B.1, the estimated coefficients on repeater exposure were always close to zero and never statistically significant at conventional levels in these regressions. To summarize, the evidence presented here strongly suggests that repeaters were indeed randomly assigned to classes within schools 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 being exposed to repeaters reduces regular students' math scores by 9.0% 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 run is in line with the results from the previous literature, which has documented negative spillovers from low-achieving and disruptive peers (e.g. Figlio, 2007; Carrell and Hoekstra, 2010; Lavy, Silva, and Weinhardt, 2012). While 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. 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 11). 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).

[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.

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 2 reveals a rapid fade-out of the negative spillover effect from repeaters on math scores: already one year after kindergarten, the estimated impact turns slightly positive, and it never falls below zero again afterwards. 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 substantial negative impacts (at least in math) of repeater exposure on test scores in the short term and a rapid fade-out of these effects later on.

4.3. Effects on non-cognitive skills

A growing literature in economics documents the importance of non-cognitive skills for success in life and argues that such skills are partly formed in school (e.g. Heckman, Stixrud, and Urzua, 2006; Chetty et al., 2011; Heckman, Pinto, and Savelyev, 2012). I analyze the impacts of repeater exposure in kindergarten on non-cognitive skills in Table 3. In stark contrast to the negative short-term effects on test scores discussed above, panel A shows large positive spillovers from repeaters on regular students' effort, initiative, value, and discipline in fourth grade, when these skills are first measured. Panel B reveals that these effects persist into eighth grade, the second and last point of measurement of these outcomes.

Panel C shows the estimated effect of repeater exposure on a summary index of non-cognitive skills. Following Kling, Liebman, and Katz (2007), this index is constructed by averaging the eight individual fourth- and eighth-grade indices for each student and normalizing the resulting composite to have mean 0 and standard deviation 1 (if only fourth-grade or only eighth-grade non-cognitive skills are observed for a student, the average of the available skill variables is used). Repeater exposure raises non-cognitive skills, as measured by the summary index, by a highly significant 11.7% of a standard deviation. Finally, Online Appendix Table B.2 reports results from regressions of the 41 individual behaviors which make up the non-cognitive skill indices used in Table 3. Those estimates similarly reveal positive impacts of repeater exposure on the vast majority of student behaviors in both fourth and eighth grade. Taken together, these results point to lasting positive impacts of sharing a kindergarten classroom with repeaters on regular students' non-cognitive skills.

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. Sharing a classroom with repeaters raises regular students' high school GPA by 0.6 points on a scale of 100 points (column 1) and raises their propensity to graduate from high school by 2.1 percentage points (column 2). Strikingly, repeater exposure also increases 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%. Finally, column 4 shows a highly significant positive impact of repeater exposure on a summary index of these three long-term outcomes (which is constructed in the same way as the summary index of non-cognitive skills).

4.5. Heterogeneity analysis

An interesting question is whether the spillovers from repeaters documented above affect all students equally. I explore 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.[15] Although all of these interactions

[15]For the sake of brevity, Table 5 and subsequent tables present estimates from regressions in which non-cognitive skills and long-term educational attainment are measured

are imprecisely estimated, 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. Thus, while sharing a kindergarten classroom with repeaters does not harm the vast majority of students in the long term, it may widen the attainment gap between low- and high-achieving students.[16]

Panel B 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. Non-cognitive 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. In contrast, there are lasting positive effects on non-cognitive skills and on long-term educational attainment. Similar patterns of only temporary impacts on test scores but persistent effects on non-cognitive skills and adult

by the respective summary indices. Results are qualitatively similar if individual non-cognitive skill measures or long-term outcomes are used instead.

[16]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.

outcomes have recently been documented for other early childhood interventions, for example by Chetty et al. (2011) and Heckman, Pinto, and Savelyev (2012).[17] Motivated by recent findings that non-cognitive skills formed early in life are a key determinant of long-term educational success (e.g. Heckman, Stixrud, and Urzua, 2006), these authors argue that the differential accumulation of non-cognitive skills by treated children is a key channel through which these interventions affected long-term outcomes. In this subsection, I provide evidence that exposure to repeaters in kindergarten similarly raises regular students' long-term educational attainment via improving their non-cognitive skills. The following subsection then discusses in detail the possible ways in which sharing a classroom with repeaters may enhance these skills.

Table 6 reports evidence from an informal test of the hypothesis that improvements in non-cognitive skills are driving the results for long-term outcomes. This test is based on the idea that if the hypothesis holds, then controlling for intermediate non-cognitive skills in the regression of long-term outcomes should substantially attenuate the estimated coefficient on repeater exposure. Corroborating similar findings from previous studies, column 1 shows that non-cognitive skills measured in fourth and eighth grade are highly predictive of long-term educational attainment. Column 2 reports the estimated effect of repeater exposure on the summary index of long-term outcomes for the subsample of students observed with non-cognitive skills. In line with the idea outlined above, this estimate is reduced by 80% when the summary index of non-cognitive skills is added to the regression as a control (column 3), and this reduction is highly statistically significant (column 4). The evidence in Table 6 thus seems to support the hypothesis that the impact of repeater exposure on long-term educational attainment works mainly through the differential accumulation of non-cognitive skills by exposed students during childhood.[18]

To further evaluate this mechanism, I use the estimated impact of re-

[17]A notable difference to my analysis is that in these studies, the impacts on test scores and non-cognitive skills go in the same direction.

[18]Note that the results do not let me rule out that repeater exposure improves long-term outcomes via raising different unobserved skills that are correlated with non-cognitive skills. The evidence in Table 6 should therefore be interpreted as suggestive.

peater exposure on non-cognitive 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 non-cognitive skills is associated with a 40.8% of a standard deviation increase in the summary index of long-term outcomes. Repeater exposure raises non-cognitive skills by 11.7% of a standard deviation (see panel C of Table 4), 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 column 4 of Table 4, lending further support to the hypothesis that non-cognitive skills are the main mechanism for the effect of repeater exposure on long-term educational attainment.

5.2. Mechanisms for impacts on non-cognitive skills

How exactly does exposure to repeaters in kindergarten enhance regular students' non-cognitive 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 the following, I first present evidence which suggests that selection effects and additional resources do not cause the improvements in non-cognitive skills of repeater-exposed students. I then discuss in more detail how the presence of a repeater can lead to learning of non-cognitive 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 above 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 non-cognitive skills. Alternatively, students with low levels of persistence and discipline may select out of classes containing repeaters. However, as argued in Section 3, 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 non-cognitive 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 during grades 1–3, but did not find any evidence of such behavior.[19] It might still be the case, however, that repeater-exposed students tend to change to *better* schools than their non-exposed peers, and that this explains their improved non-cognitive 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 B.3 shows that the results from these regressions were qualitatively and quantitatively similar to the ones in Section 4.[20]

As reported in Section 2, a few students also managed to change classes *within* Project STAR schools after kindergarten. Unfortunately, the data only allow me to observe the actual class attended and not the randomly assigned class for each student, which prevents me from testing whether such switching is related to repeater exposure. However, two pieces of

[19] 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).

[20] There is still the possibility that students change schools between third and fourth grade, i.e. just before non-cognitive 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 non-cognitive 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.

evidence suggest that between-class switching is not the main mechanism behind the rise in non-cognitive skills. First, such switching was relatively rare compared to the between-school switching discussed in the previous paragraph, and changing from small classes in particular was very limited (see Krueger, 1999). The fact that I find no evidence for differences in spillovers by class size thus makes between-class switching an unlikely explanation for my results. Second, I confirmed that exposure to repeaters in kindergarten does not predict the observable demographic characteristics of students' classmates in grades 1–3, suggesting that post-kindergarten class composition is as good as random (results are available upon request).

Selection out of the sample. For reasons detailed in Online Appendix A, 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 non-cognitive skills and some of the long-term outcomes found above. In Table 7, I provide evidence that such selective attrition is unlikely to drive my results.

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 non-exposed students changes across different outcomes. I tested for such compositional changes by adding interactions between the 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 non-exposed students observed with different outcomes. Finally, to dispel any remaining concerns about selective attrition driving my results, I re-estimated the main specifications for a consistent sample of non-attritors. As panel C shows, the results were qualitatively and quantitatively similar to those found for the unrestricted sample in Section 4.[21]

Additional resources. Yet another explanation for the enhanced non-cognitive 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 non-cognitive 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 B.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. While repeaters are more likely than regular students to participate in such programs in kindergarten (23% versus 6% participation rate for both programs combined), Online Appendix Table B.5 shows that results are qualitatively

[21]For the purpose of these regressions, I defined non-attritors 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 non-cognitive 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.

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 non-cognitive skills of repeater-exposed students.[22]

Behavioral responses. 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 non-cognitive skills. A particularly salient possibility is that teachers emphasize the learning of behavioral and study skills in response to class disruption by repeaters. Such disruption could take the form of misbehavior that directly distracts other students, for example, 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 by changing their teaching practices, for instance by having students work more in small groups, or by changing the actual content of their lessons. Importantly, such adjustments could account for both the drop in end-of-kindergarten test scores and the improvements in non-cognitive skills observed in the data.

To test this explanation, one would ideally observe teaching practices and lesson content, perhaps from teacher time use surveys and teacher logs. Unfortunately, such data is not available for teachers in Project STAR. However, the mechanism outlined in the previous paragraph does receive some support from the existing literature. In particular, studies in educational psychology confirm that teachers adapt their teaching practices to students' cognitive ability and behavior (e.g. Nurmi, 2012), with one recent study finding that teachers of low-achieving students are more likely to establish explicit rules for behavior and stable routines in the classroom (Pakarinen et al., 2011). Moreover, a distinct literature in economics of education has shown that different teaching practices promote different cognitive skills and social behaviors in children (e.g. Algan, Cahuc, and Shleifer, 2011; Bietenbeck, 2014). This suggests that the rise in non-cognitive skills

[22] One other way in which repeater-exposed students might profit from additional resources is if they repeat a grade themselves. In unreported regressions, I found that this is not the case (results are available upon request).

documented above may indeed be due to teachers' adjustment to the presence of repeaters in the classroom.[23]

While 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. Alternatively, parents of repeater-exposed students may compensate for a worse classroom environment by helping their children more at home. Both of these reactions may lead to the observed improvements in non-cognitive 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 of this paper distinguishes between classes with and without repeaters, but does not further differentiate classes according to the actual number of repeaters. In Online appendix Table B.7, 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. While 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. While the estimated coefficients on the dummy are roughly similar to the ones reported in Section 4,

[23]This mechanism is also in line with recent evidence by Jackson (2012) that teachers who are good at raising test scores are not necessarily good at raising non-cognitive skills and vice versa.

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 versus no repeaters) is more important than the intensive margin (being in class with one versus more 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 which has been widely used to identify peer effects in the previous literature.[24] Online appendix Table B.7 reports results from regressions of four main outcomes which additionally control for average demographic characteristics of each students' classmates (including repeating classmates). To the extent that the impacts documented in Section 4 operate via changing the 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.[25]

[24]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.

[25]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 (results are available upon request). Across the regressions in Online appendix Table B.7, 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-variables strategy to examine the impacts of classmates' average age; in specifications which 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.

6.3. Relative measurement of non-cognitive skills

Table 3 reports positive impacts from repeater exposure in kindergarten on regular students' non-cognitive 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 B.8. In panel A, I re-estimate the impacts of repeater exposure on fourth-grade non-cognitive 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.

The data do not allow me to observe classroom composition during eighth grade. However, 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 estimated impacts of repeater exposure on non-cognitive skills in this restricted sample are very similar to the ones reported in Table 3. Therefore, the evidence does not support the idea that the positive impacts of repeater exposure on non-cognitive skills capture purely mechanical effects due to relative teacher ratings.

6.4. 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. Intuitively, this bias is avoided in this paper due to the clear separation of initiators and recipients of spillover effects. I confirmed this intuition in a simulation-based falsification test similar to the one developed by Feld and Zoelitz (2014). 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. Thus, the mechanical forces described by Angrist (2014) do not bias the results in this paper.

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 non-cognitive skills, and their long-term educational attainment.

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 worse on standardized tests at the end of kindergarten. However, these students display substantially improved non-cognitive skills, such as effort and discipline, when these are first measured at the beginning of fourth grade. While the negative spillovers from repeaters on test scores fade out rapidly after kindergarten, the positive spillovers on non-cognitive 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 eighteen. My analysis suggests that these positive long-term impacts are likely due to the differential accumulation of non-cognitive skills by exposed students, which in turn appears to be a result of behavioral adjustments by teachers, students, or their parents.

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 which 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.

References

Algan, Y., P. Cahuc, and A. Shleifer. 2011. "Teaching practices and social capital." *American economic journal: Applied economics* 5:189–210.

Angrist, J.D. 2014. "The Perils of Peer Effects." *Labour Economics* 30:98–108.

Bietenbeck, J. 2014. "Teaching practices and cognitive skills." *Labour Economics* 30:143–153.

Bifulco, R., J. Fletcher, and S. 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, S.E., P.J. Devereux, and K.G. Salvanes. 2013. "Under Pressure? The Effect of Peers on Outcomes of Young Adults." *Journal of Labor Economics* 31:119–153.

Burke, M.A., and T.R. Sass. 2013. "Classroom Peer Effects and Student Achievement." *Journal of Labor Economics* 31:51–82.

Carrell, S.E., and M.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, E.U., and D.W. Schanzenbach. 2016. "First in the class? Age and the Education Production Function." *Education Finance and Policy* 11:225–250.

Chetty, R., J.N. Friedman, N. Hilger, E. Saez, D.W. Schanzenbach, and D. Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *The Quarterly Journal of Economics* 126:1593–1660.

Deming, D., and S. Dynarski. 2008. "The Lengthening of Childhood." *Journal of Economic Perspectives* 22(3):71–92.

Feld, J., and U. Zoelitz. 2014. "Understanding Peer Effects: On the Nature, Estimation and Channels of Peer Effects." mimeo.

Figlio, D.N. 2007. "Boys Named Sue: Disruptive Children and Their Peers." *Education Finance and Policy* 2:376–394.

Finn, J.D., J. Boyd-Zaharias, R.M. Fish, and S.B. Gerber. 2007. "Project STAR and Beyond: Database User's Guide." Report, HEROS Incorporated.

Gottfried, M.A. 2013. "The Spillover Effects of Grade-Retained Classmates: Evidence from Urban Elementary Schools." *American Journal of Education* 119:1–64.

Gould, E.D., V. Lavy, and M.D. Paserman. 2009. "Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence." *The Economic Journal* 119:1243–1269.

Graham, B. 2008. "Identifying Social Interactions through Conditional Variance Restrictions." *Econometrica* 76:643–660.

Hauser, R.M., C.B. Frederick, and M. Andrew. 2007. "Grade Retention in the Age of Standards-Based Reform." In A. Gamoran, ed. *Standards-based Reform and the Poverty Gap: Lessons for No Child Left Behind*. Brookings Institution Press, pp. 120–153.

Heckman, J., R. Pinto, and P. Savelyev. 2012. "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." *American Economic Review* 103:2052–2086.

Heckman, J., J. Stixrud, and S. Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24:411–482.

Hill, A.J. 2014. "The Costs of Failure: Negative Externalities in High School Course Repetition." *Economics of Education Review* 43:91–105.

Horoi, I., and B. Ost. 2015. "Disruptive Peers and the Estimation of Teacher Value Added." *Economics of Education Review* 49:180–192.

Hoxby, C. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper No. 7867.

Jackson, C.K. 2012. "Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina." NBER Working Paper No. 18624.

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

Karweit, N.L. 1999. "Grade Retention: Prevalence, Timing, and Effects." Report, Center for Research on Students Placed at Risk, Johns Hopkins University.

Kautz, T., J.J. Heckman, R. Diris, B.t. Weel, and L. Borghans. 2014. "Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success." OECD Education Working Paper No. 110.

Kling, J.R., J.B. Liebman, and L.F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75:83–119.

Krueger, A.B. 1999. "Experimental Estimates of Education Production Functions." *The Quarterly Journal of Economics* 114:497–532.

Krueger, A.B., and D.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, V., M.D. Paserman, and A. Schlosser. 2012. "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, V., O. Silva, and F. Weinhardt. 2012. "The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools." *Journal of Labor Economics* 30:367–414.

Lazear, E.P. 2001. "Educational Production." *The Quarterly Journal of Economics* 116:777–803.

NCES. 2013. "Digest of Education Statistics 2012." Report, National Center for Education Statistics.

Nurmi, J.E. 2012. "Students' characteristics and teacher-child relationships in instruction: A meta-analysis." *Educational Research Review* 7:177–197.

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, E., M.K. Lerkkanen, A.M. Poikkeus, M. Siekkinen, and J.E. Nurmi. 2011. "Kindergarten teachers adjust their teaching practices in accordance with children's academic pre-skills." *Educational Psychology* 31:37–53.

Sacerdote, B. 2011. "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" Elsevier, vol. 3 of *Handbook of the Economics of Education*, pp. 249 – 277.

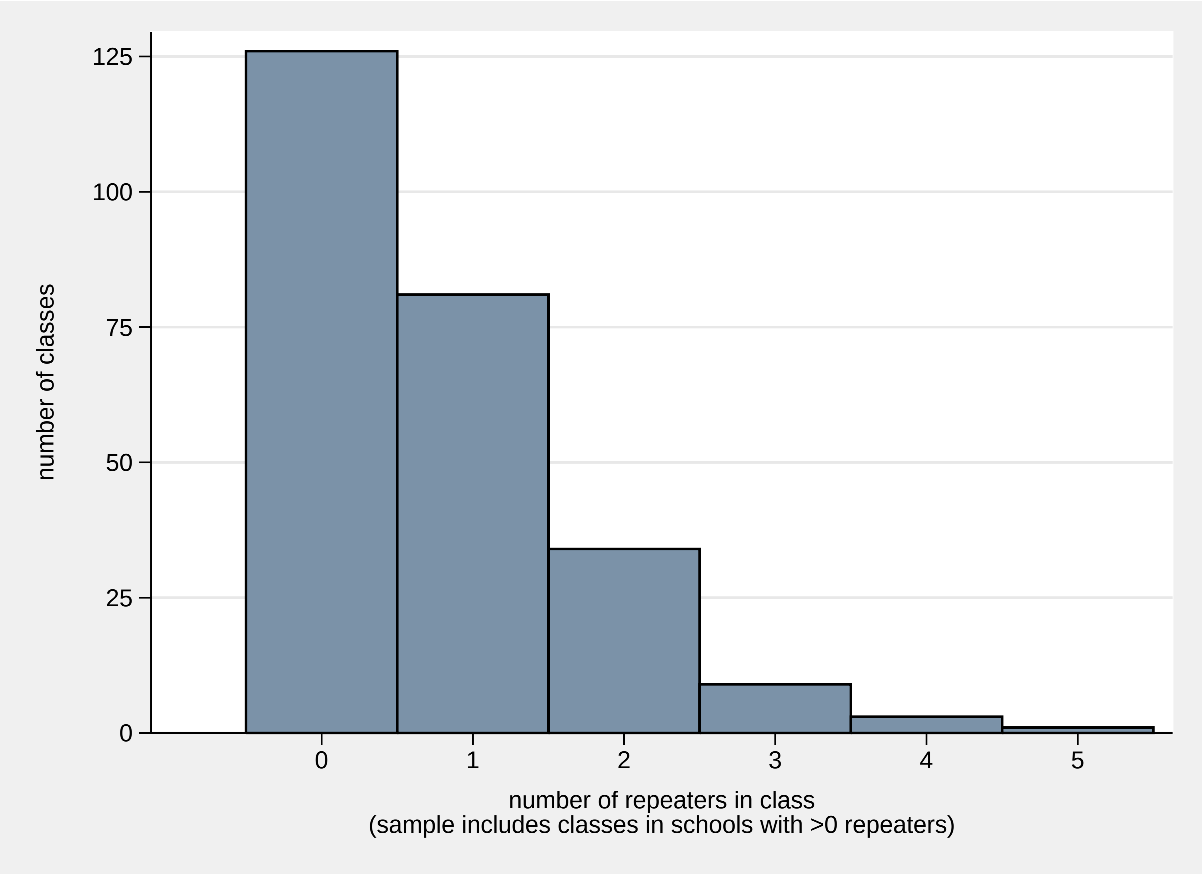
Schanzenbach, D.W. 2006. "What have researchers learned from Project STAR?" *Brookings Papers on Education Policy* 9:205–228.

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

Whitmore, D. 2005. "Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment." *American Economic Review*, pp. 199–204.

Word, E., J. Johnston, H.P. Bain, D.B. Fulton, C.M. Achilles, M.N. Lintz, J. Folger, and C. Breda. 1990. "The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Technical Report 1985-1990." Report, Tennessee State Department of Education.

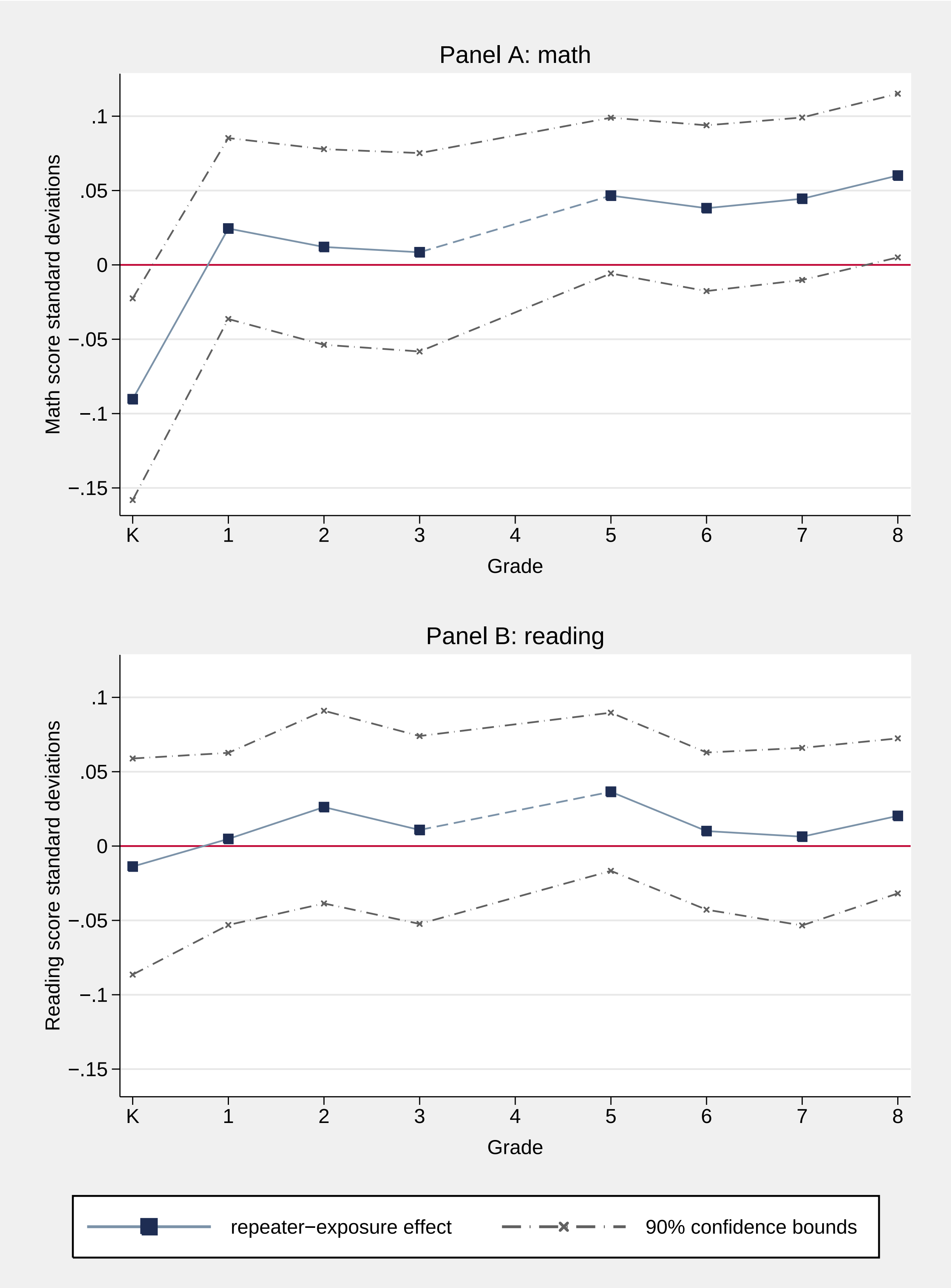
Figure 1
Distribution of repeaters across classes



Notes: 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).

Figure 2

Repeater exposure in kindergarten and post-kindergarten test scores



Notes: The figure plots regression coefficients and 90% confidence intervals obtained from 16 variations of the specifications in columns 2 and 4 of Table 2. The dependent variable in each regression is the math score (panel A) or reading score (panel B) in the year in which students were supposed to be in the grade indicated on the horizontal axis. 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 A for details. See the notes to Table 2 for information about additional controls included in each of the regressions.

Table 1
Descriptive statistics

	Non-repeaters			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
Non-cognitive 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 non-repeating 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 1985-86 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 non-repeating students. The non-cognitive skill measures are indices summarizing teacher ratings of student behavior in four areas: effort, initiative, value, and discipline. All test scores and measures of non-cognitive skills are standardized to have mean 0 and standard deviation 1 across non-repeating students. High school GPA is measured on a scale from 0-100. 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.

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

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 non-repeating 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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3
Repeater exposure in kindergarten and non-cognitive skills

	Effort (1)	Initiative (2)	Value (3)	Discipline (4)
	Panel A: 4th grade			
Repeater exposure	0.104*	0.025	0.124**	0.142***
	(0.054)	(0.056)	(0.053)	(0.054)
Observations	1,628	1,628	1,628	1,628
	Panel B: 8th grade			
Repeater exposure	0.169***	0.105*	0.160***	0.194***
	(0.054)	(0.056)	(0.051)	(0.052)
Observations	1,731	1,731	1,731	1,731
	Panel C: summary index			
Repeater exposure	0.117***			
	(0.041)			
Observations	2,589			

Notes: The table reports estimates from regressions that relate students' non-cognitive skills in fourth and eighth grade to their exposure to repeaters in kindergarten. The outcome variables in panels A and B are indices summarizing teacher ratings of student behavior in four areas: effort, initiative, value, and discipline. The indices are standardized to have mean 0 and standard deviation 1 across non-repeating students in the estimation sample. The outcome variable in panel C is a summary index of non-cognitive skills that combines the available information from fourth and eighth grade for each student; see text for details on how this index is constructed. 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 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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

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)
Repeater exposure	0.552*	0.021*	0.033**	0.074***
	(0.308)	(0.013)	(0.015)	(0.028)
Observations	2,438	2,955	6,039	6,039

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 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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 5
Heterogeneous effects

	Kindergarten math score (1)	8th-grade math score (2)	Non-cog. index (3)	Long-term index (4)
	Panel A: heterogeneity by non-repeaters' 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 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. (both 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) and by class size in kindergarten (panel B). 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 descriptions of the outcome variables. Standard errors in parentheses allow for clustering at the kindergarten class level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 6

Repeater exposure, non-cognitive 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]
Non-cog. 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 non-cognitive 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 non-cognitive skills. All specifications control for students' demographic background, an indicator for small class in kindergarten, and kindergarten school fixed effects. Regressions include all non-repeating students for whom non-cognitive 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<0.10, ** p<0.05, *** p<0.01.

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 non-attritors						
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 non-repeating 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 non-repeating 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.

ONLINE APPENDIX TO:

The Long-Term Impacts of Low-Achieving Childhood Peers: Evidence from Project STAR

Jan Bietenbeck

Lund University and IZA

A. Data appendix

The Tennessee State Department of Education entrusted a consortium of researchers from four Tennessee universities and various state institutions with the planning and implementation of Project STAR. After the experiment ended, some researchers continued to collect data on outcomes of participating students. Finn et al. (2007) provide a detailed account of these data collection efforts. The Project STAR public use file, on which the empirical analysis in this paper is based, combines these data such that students can be followed throughout their scholastic careers until the end of high school. Additional data on test scores in grades 5-8 were generously provided to me by Diane Schanzenbach. In what follows, I discuss in detail how I constructed the outcome variables used in the empirical analysis.

Test scores. At the end of each school year from kindergarten through third grade, students in Project STAR were administered the grade-specific version of the Stanford Achievement Test. From fifth grade through eighth grade, students who were still residing in Tennessee took the Comprehensive Test of Basic Skills (CTBS) as part of a statewide testing program.[1] Both tests are standardized multiple-choice assessments with components in mathematics and reading and are graded centrally.

The public use file contains Stanford Achievement Test scores for all students who took these tests. However, it contains CTBS scores only for students who were on grade level, i.e. students who attended grade 5/6/7/8 in 1991/1992/1993/1994, respectively. This implies that test scores are not observed for a number of students who had been retained in grade by those years.[2] In contrast, the data supplied by Diane Schanzenbach contain CTBS scores for students who attended grades 5-8 in Tennessee in any year between 1990 and 1997. Test scores are provided as scale scores, which are

[1] An unrepresentative subsample of students took the CTBS also in fourth grade, see Finn et al. (2007). Due to the selective nature of this subsample, I chose not to use fourth-grade test scores in the empirical analysis.

[2] Note that students who were retained in grade at any point between kindergarten and third grade dropped out of the STAR cohort and therefore did not write the subsequent Stanford Achievement Tests. However, these students did write the CTBS in later grades as long as they stayed in Tennessee.

comparable across grade levels (Finn et al., 2007). In order to increase sample size, I define test scores for a given grade level as scores obtained in the school year in which participating students were supposed to be in that grade (e.g., eighth-grade scores are defined as scores obtained in 1994, even though some students were attending seventh grade in that year). Results are however robust to using only the test scores available in the public use file. I standardize all test scores to have mean 0 and standard deviation 1 across non-repeating kindergarten students in the estimation sample.

Non-cognitive skills. I obtain fourth-grade non-cognitive skill measures from a questionnaire administered to teachers of a random sample of participating students in November 1989. The questionnaire asked teachers to rate how often each student had engaged in 31 different behaviors over the last two to three months. Ratings were recorded on a scale from 1 ("never") to 5 ("always"), and ratings of 28 of these behaviors were consolidated into four indices. The effort index includes items such as whether a student is persistent when confronted with difficult problems, whether she completes her homework, and whether she gets discouraged easily when encountering an obstacle in schoolwork. The initiative index is based on such items as whether a student participates actively in classroom discussions, whether she does more than just the assigned work, and whether she often asks questions. The value index measures how much a student appreciates the school learning environment. Finally, the discipline index captures such characteristics as whether a student often acts restless, whether she needs reprimanding, and whether she interferes with peers' work.[3]

During the 1993-94 school year, eighth-grade math and English teachers of a different random subset of participants were asked about student behaviors on a similar though shorter questionnaire. Thirteen of these behaviors were again consolidated into four indices measuring each student's effort, initiative, value, and discipline. I first average these indices across math and English for each student, and then normalize each of the eight

[3]Note that what the paper refers to as the "discipline index" is the inverse of the "index of non-participatory behavior" in the original data. See Finn et al. (2007) for a complete listing of the behaviors included in each of the indices.

fourth- and eighth-grade indices by subtracting its mean and dividing by its standard deviation (computed across non-repeating students in the estimation sample). Finally, I construct the summary index of non-cognitive skills by averaging the available normalized indices for each student and normalizing the resulting composite.

High school grade point average and graduation. Most students in Project STAR graduated from high school in 1998, and transcripts were gathered from selected high schools in 1999 and 2000. High schools were chosen for data collection based on the likelihood that STAR participants would attend them given the locations of students' last known middle schools. Course grades from transcripts were transferred to a scale from 0-100 if necessary, and separate GPAs for math, science, and foreign languages were computed and are available in the data. The empirical analysis in this paper uses overall GPA, defined as the average of the these three subject-specific GPAs, as an outcome variable.

Information on high school graduation was also derived from transcripts and cross-checked with data from the Tennessee State Department of Education in ambiguous cases. Nevertheless, graduation status could not be determined with certainty for all students. In these cases, which comprise 7% of the non-repeating students in the estimation sample, the data collectors made a best guess whether a student "probably graduated" or "probably dropped out" based on the available course grades, information on attendance, and additional information from the Tennessee State Department of Education. The variable used in the empirical analysis codes 2,378 students who graduated, 98 students who probably graduated, and 82 students who received a General Educational Development certificate as graduates, and 296 students who dropped out and 101 students who probably dropped out as dropouts.

College-test taking and summary index of long-term educational attainment. ACT/SAT test taking was recorded by Krueger and Whitmore (2001), who matched all students in STAR to the administrative records of the two companies responsible for these tests in 1998. The outcome variable used in the empirical analysis is an indicator that takes value 1 if a student took either of these college entrance exams in 1998 and 0 otherwise. The summary index of long-term educational attainment combines information on high school GPA and graduation and college-test taking by first standardizing each of these variables to have mean 0 and standard deviation 1 across non-repeating students in the estimation sample. The average of these standardized variables is then normalized by subtracting its mean and dividing by its standard deviation.

References

Finn, J.D., J. Boyd-Zaharias, R.M. Fish, and S.B. Gerber. 2007. "Project STAR and Beyond: Database User's Guide." Report, HEROS Incorporated.

Krueger, A.B., and D.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.

B. Appendix tables

Table B.1
Randomization tests

	Male	Black	Free lunch	Age in years	Old for grade
	(1)	(2)	(3)	(4)	(5)
	Panel A: controlling for school fixed effects				
Repeater exposure	–0.005	–0.001	0.004	0.001	–0.003
	(0.015)	(0.007)	(0.015)	(0.009)	(0.005)
	Panel B: controlling for school fixed effects and class size				
Repeater exposure	–0.006	–0.002	0.001	0.004	–0.001
	(0.015)	(0.007)	(0.015)	(0.009)	(0.005)
Obs. (both panels)	6,039	6,039	6,039	6,039	6,039

Notes: The table reports estimates from regressions that relate non-repeating students' demographic characteristics to their exposure to repeaters 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. Standard errors in parentheses allow for clustering at the kindergarten class level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table B.2

Results for individual items of non-cognitive skill indices

	coefficient	std. error
4th-grade effort index		
(+) pays attention in class	0.087	(0.058)
(+) completes homework	0.096*	(0.057)
(+) works well with other children	0.150***	(0.055)
(+) completes assigned seatwork	0.126**	(0.055)
(+) is persistent when facing difficult problems	0.067	(0.058)
(+) makes sincere effort	0.065	(0.053)
(+) tries to finish even difficult assignments	0.050	(0.049)
(–) loses or misplaces materials	–0.071	(0.059)
(–) comes late to class	–0.093**	(0.046)
(–) doesn't seem to know what's going on in class	–0.086	(0.056)
(–) no independent initiative, needs constant help	–0.107**	(0.053)
(–) prefers easy problems	–0.002	(0.052)
(–) is easily discouraged, gives up easily	–0.072	(0.052)
4th-grade initiative index		
(+) tries to do work thoroughly and well	0.114*	(0.061)
(+) participates actively in discussions	–0.011	(0.057)
(+) does more than just the assigned work	0.057	(0.056)
(+) asks questions	–0.004	(0.058)
(+) raises hand to answer a question	0.011	(0.054)
(+) seeks reference material on her/his own	0.043	(0.051)
(+) discusses subject with teacher outside of class	–0.020	(0.057)
(–) is withdrawn	0.028	(0.060)
4th-grade value index		
(+) thinks school is important	0.144**	(0.058)
(–) is critical of peers who do well in school	–0.061	(0.053)
(–) criticizes the subject matter	–0.082	(0.055)
4th-grade discipline index		
(–) acts restless, unable to sit still	–0.145***	(0.049)
(–) needs reprimanding	–0.097*	(0.056)
(–) annoys or interferes with peers' work	–0.139***	(0.053)
(–) talks with classmates too much	–0.131**	(0.057)
8th-grade effort index		
(+) pays attention in class	0.133**	(0.055)
(+) completes assigned seatwork	0.103*	(0.054)
(+) is persistent when facing difficult problems	0.117**	(0.048)
(–) loses or misplaces materials	–0.219***	(0.057)
(–) comes late to class	–0.110**	(0.055)

(continued on next page)

(continued from previous page)

	coefficient	std. error
8th-grade initiative index		
(+) participates actively in discussions	0.064	(0.054)
(+) does more than just the assigned work	0.122**	(0.057)
(+) discusses subject with teacher outside of class	0.075	(0.054)
8th-grade value index		
(–) is critical of peers who do well in school	–0.103*	(0.053)
(–) criticizes the subject matter	–0.180***	(0.050)
8th-grade discipline index		
(–) needs reprimanding	–0.184***	(0.053)
(–) annoys or interferes with peers' work	–0.187***	(0.052)
(–) is verbally or physically abusive to teacher	–0.116**	(0.053)

Notes: The table reports estimates of the effects of exposure to repeaters in kindergarten on the individual items which make up the non-cognitive skill indices that are used as outcomes in Table 3. Items in the left column that are marked (+) enter the respective index positively, while items marked (–) enter it negatively. All items are standardized to have mean 0 and standard deviation 1 across non-repeating students in the estimation sample. The middle and rightmost columns report the coefficient estimates and standard errors for the repeater-exposure dummy, respectively. Regressions include the 1,628 (1,731) students observed with non-cognitive skills in 4th (8th) grade. 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.

Table B.3
Results for school stayers

	Kindergarten math score (1)	8th-grade math score (2)	Non-cog. index (3)	Long-term index (4)
Repeater exposure	–0.101**	0.080**	0.155***	0.098**
	(0.051)	(0.039)	(0.044)	(0.043)
Observations	2,625	2,350	2,076	2,776

Notes: The table reports estimates from regressions for a sample of students who stayed in the same school for the entire duration of Project STAR (i.e. from kindergarten through third grade; N=2,776 non-repeating students). 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 a description 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.

Table B.4
Effects on special education/instruction status

	Kindergarten		First grade	
	Special education (1)	Special instruction (2)	Special education (3)	Special instruction (4)
Repeater exposure	–0.002	–0.003	–0.002	–0.020
	(0.005)	(0.007)	(0.003)	(0.014)
Observations	6,038	6,038	4,311	4,311

Notes: The table reports estimates of the effect of exposure to repeaters in kindergarten on indicators for being classified as a special education or special instruction student in kindergarten and first grade. 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 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.

Table B.5

Results for a sample excluding special education/instruction repeaters

	Kindergarten math score (1)	8th-grade math score (2)	Non-cog. index (3)	Long-term index (4)
Repeater exposure	–0.109**	0.065*	0.121**	0.064**
	(0.048)	(0.038)	(0.048)	(0.032)
Observations	5,012	3,871	2,266	5,402

Notes: The table reports estimates from regressions for a sample that excludes classes with repeaters classified as special education students or pulled out for special instruction. 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 a description 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.

Table B.6
Alternative measures of repeater exposure

	Kindergarten math score (1)	8th-grade math score (2)	Non-cog. 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.

Table B.7

Controlling for average demographic characteristics of classmates

	Kindergarten math score (1)	8th-grade math score (2)	Non-cog. index (3)	Long-term index (4)
	Panel A: controlling for the share of male classmates			
Repeater exposure	–0.087**	0.060*	0.118***	0.075***
	(0.041)	(0.033)	(0.041)	(0.028)
	Panel B: controlling for the share of black classmates			
Repeater exposure	–0.090**	0.061*	0.118***	0.074***
	(0.041)	(0.033)	(0.041)	(0.028)
	Panel C: controlling for the share of classmates on free lunch			
Repeater exposure	–0.089**	0.062*	0.117***	0.072**
	(0.041)	(0.034)	(0.041)	(0.028)
	Panel D: controlling for the average age of classmates			
Repeater exposure	–0.116***	0.043	0.095**	0.064**
	(0.044)	(0.034)	(0.045)	(0.030)
	Panel E: controlling for the share of old-for-grade classmates			
Repeater exposure	–0.086	0.075*	0.136***	0.080**
	(0.053)	(0.038)	(0.051)	(0.032)
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 controlling for average demographic characteristics (indicated in the heading of each panel) of students' kindergarten classmates. Average demographic characteristics are computed as leave-out means for each student, i.e. as the average across all of her kindergarten classmates (including repeaters) but excluding herself. 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 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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table B.8

Robustness to relative measurement of non-cognitive skills

	Effort (1)	Initiative (2)	Value (3)	Discipline (4)
	Panel A: 4th grade, no repeaters in class			
Repeater exposure	0.072	–0.006	0.031	0.081
	(0.070)	(0.074)	(0.071)	(0.068)
Observations	1,037	1,037	1,037	1,037
	Panel B: 8th grade, no repeaters in school			
Repeater exposure	0.161*	0.074	0.170**	0.227***
	(0.082)	(0.086)	(0.071)	(0.084)
Observations	866	866	866	866

Notes: The table reports estimates from regressions that probe for measurement of non-cognitive skills relative to repeaters. In panel A (panel B), the sample is restricted to students whose fourth-grade class (eighth-grade school) did not contain any of the 193 original kindergarten repeaters. 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 the notes to Table 3 for descriptions 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. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.