WORKING PAPER Nº IDB-WP-1688

Disruption in the Classroom: Experimental Evidence from Ecuador

Pedro Carneiro
Yyannú Cruz-Aguayo
Francesca Salvati
Norbert Schady

Inter-American Development Bank
Social Protection and Labor Markets Division

March 2025

Disruption in the Classroom: Experimental Evidence from Ecuador

Pedro Carneiro*
Yyannú Cruz-Aguayo**
Francesca Salvati***
Norbert Schady****

*University College London, IFS, CEMMAP
**Inter-American Development Bank
***University of Essex, IFS
****World Bank

Inter-American Development Bank
Social Protection and Labor Markets Division

March 2025

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Disruption in the classroom: experimental evidence from Ecuador / Pedro Carneiro, Yyannú Cruz-Aguayo, Francesca Salvati, Norbert Schady.
p. cm. — (IDB Working Paper Series ; 1688)
Includes bibliographical references.
1. Problem children-Education-Ecuador. 2. Learning disabled children-Ecuador. 3. Academic achievement-Ecuador. I. Caneiro, Pedro. II. Cruz Aguayo, Yyanú. III. Salvati, Francesca. IV. Schady, Norbert. V. Inter-American Development Bank. Social Protection Labor Markets Division. VI. Series.
IDB-WP-1688

JEL Codes: I24, I25.

Keywords: Peer effects, behavioral problems.

http://www.iadb.org

Disruption in the classroom: Experimental evidence from Ecuador

Pedro Carneiro

University College London, IFS, CEMMAP

Yyannú Cruz-Aguayo

Inter-American Development Bank

Francesca Salvati

University of Essex, IFS

Norbert Schady

World Bank

October 31, 2024[1]

Abstract

We study how poorly-behaved children affect learning and other outcomes of their peers using data from a unique experiment in Ecuador. Within each school, students were randomly assigned to classrooms in every grade for seven consecutive grades, between kindergarten and 6th grade. Children with persistent behavioral problems lower the math and language achievement of their classmates. The more poorly-behaved children there are in a class, the larger is the negative effect on the achievement of their classmates. These negative impacts are larger for younger children, and they persist for at least two years after exposure to a poorly-behaved peer. We find indirect evidence that children with persistent behavioral difficulties are passed around schools.

[1] We gratefully acknowledge the support of the InterAmerican Development Bank. Carneiro also gratefully acknowledges the support of the ERC through grant ERC-2015-CoG-682349 and the Spencer Foundation.

1. Introduction

Across OECD countries, teachers spend on average 13 percent of their time (8 minutes per teaching hour) keeping order in the classroom (OECD 2019). If no learning occurs while teachers struggle to maintain order, the economic costs of classroom disruption could imply billions of dollars of foregone earnings in the U.S. alone.[2]

In this paper, we quantify the learning cost of classroom disruption using a unique experiment in Ecuador, a middle-income country in South America. Our analysis is motivated by an influential paper by Lazear (2001), who presents a model of how children who are disruptive can have large (negative) spillovers on the learning of their peers. One can think of different sources of disruption, and Lazear (2001) emphasizes two, which occur when (i) a student is misbehaving; or (ii) a student asks a question to which all classmates already know the answer. We investigate how these two potential sources of disruption affect learning in the classroom.

We find that the 2-3 percent of students with the most persistent problem behaviors reduce the learning of their peers. The magnitude of these negative spillovers increases with the number of misbehaved children in the classroom, and with the number of grades a student is exposed to them. In contrast, we find no negative spillovers from being exposed to classmates with very low achievement.

We use data from 202 schools in the coastal region of Ecuador. Every school had at least two classrooms per grade. A cohort of children entering kindergarten was randomly assigned to classrooms within schools. These children were then randomly reassigned to classrooms in every grade between 1st and 6th grades. Thus, children who did not switch schools were exposed to seven exogenous, orthogonal sets of peers, some of whom may have been particularly disruptive.[3] Compliance with the random assignment was almost perfect, 98.9 percent on average.

At the end of each grade, children were tested in math and language. Between 1st and 4th grades, data on child executive function (EF) was also collected. EF refers to a set of skills that allow individuals

---

[2] This can be shown with simple back-of-the-envelope calculations. Hanushek and Woessmann (2020) propose a "rough rule of thumb" that, on average, there is 0.3 SDs of learning per grade. If there is no learning when there is misbehavior in the classroom, and teachers spend 13 percent of their time managing misbehavior (the OECD average), learning lost would be ~0.04 SDs (0.13*0.3), or ~1.6 percentiles at the mean of a standard normal distribution. Using data from Project Star, Chetty et al. (2011) estimate that a 1 percentile increase in kindergarten test scores leads to a 0.83 percent increase in earnings, so a 1.6 percentile decline would imply a ~1.3 percent decline in earnings. To translate this into dollars, we use data from the U.S Bureau of Labor Statistics. These data show that, in the first quarter of 2024, there were 119.2 million full-time wage and salary earners in the U.S., making $1,139 per week on average (Bureau of Labor Statistics 2024). A 1.3 percent decline in earnings would therefore amount to a yearly decline of $770 per worker per year, and a total loss of earnings of approximately $92 billion.

[3] Of course, there is also variation in exposure to other measures of classroom quality, including the quality of teachers. However, by design, these are orthogonal to peer quality. Teachers were also assigned randomly to classrooms within schools and grades.

to plan, focus attention, remember instructions, and juggle multiple tasks. It includes working memory, inhibitory control, and cognitive flexibility (Center for the Developing Child 2019). Data on child depression, self-esteem, grit, and growth mindset was collected at the end of 6th grade.

To identify poorly-behaved children in the sample, we use information on their classroom behaviors directly reported by their teachers. At the end of each school year, teachers were asked to list 5 children in their classrooms with the most serious behavioral problems and, separately, 5 children who had the biggest difficulties learning. In this paper, we are interested in the effects of children with *persistent* behavioral problems or difficulty learning. For this reason, and to alleviate concerns about measurement error in teacher reports from a single grade, we classify a child as *poorly-behaved* if teachers reported them to be one of the worst-behaved children in their classroom in (all of) the three previous grades, and define *low-achieving* students in a comparable way. We show, however, that our point estimates are broadly similar across different definitions of who is a poorly behaved child.[4]

The design of the experiment allows us to address important empirical challenges and make new contributions to the literature. First, random assignment, with essentially perfect compliance, ensures that our results are not biased by purposeful placement of poorly-behaved children. While there is a large literature using random assignment to estimate peer effects in university education,[5] the number of papers based on random assignment in elementary, middle, and high school is much more limited. It includes Duflo et al. (2011) who estimate the effects of tracking in Kenya; Huang and Zhu (2020) and Xu et al. (2022) who study the effects of peers in middle school in China; Sojourner (2013), who analyzes the effects of peers in project STAR in Tennessee; and Zarate (2023), who uses random assignment of students in selective boarding schools in Peru. None of these focuses on the impacts of poorly-behaved children. The research on disruptive peers (Aizer 2008; Balestra et al. 2022; Carrell and Hoekstra 2010; Carrell et al. 2018; Figlio 2007; Kristoffersen et al. 2015; Santavirta and Sarzosa 2023) is non-experimental, with one exception (Zhao and Zhao 2021).

---

[4] The decision to use three (as opposed to two, four, or any other number) years of problem behaviors (low achievement) to define poorly behaved (low achieving) children is done to balance two goals: identifying persistently disruptive children and estimating the model with reasonable sample sizes. The larger the number of consecutive periods of problem behaviors used in this classification the more restrictive the definition of a poorly behaved child, and the more seriously disruptive the child is likely to be. This leads to fewer but more seriously disruptive children. With fewer disruptive children, fewer students will be exposed to a disruptive child in the classroom, meaning we have less variation to estimate effects of disruptive children on their peers' outcomes. Moreover, using a larger number of consecutive periods to define disruptive children means that there are fewer grades over which we can measure their impact, which affects the power of our estimates.

[5] Important papers include Booji et al. (2017); Carrell et al. (2009); De Giorgi and Pellizzari (2013); Feld and Zölitz (2017; 2022); Golsteyn et al. (2021); Lyle (2009); Sacerdote (2001), Shan and Zölitz (2024).

Second, since we observe repeated (exogenous) exposure to poorly-behaved children across different grades, and repeated assessments of children's skills, we can estimate dynamic models of learning with exposure to poorly-behaved peers. We can then assess whether the effects of disruption persist over time, and measure the cumulative impact of repeated exposure to disruption—specifically, are the impacts of poorly-behaved peers additive across grades or, rather, does the effect of exposure in a grade depend on whether children were exposed to disruptive children in earlier grades?

Third, since our measures are based directly on teacher reports (on the same children, but at different points in time) our analysis focuses precisely on the behaviors that teachers believe disrupt learning. Other papers have used proxies for potential disruptive behavior, such as boys with names commonly given to girls (Figlio 2007), exposure to domestic violence (Carrell and Hoekstra 2010; Carrell et al. 2018), whether children have been diagnosed with, and treated for, attention-deficit disorders or special needs (Aizer 2008; Balestra et al. 2022; Kristoffersen et al. 2015), children with alcoholic fathers (Zhao and Zhao 2021), or children who have been abused or neglected (Santavirta and Sarsoza 2023). While it is likely that these children exhibit disruptive behaviors in the classroom, which could lead to learning losses, there are other potential channels through which peer effects could occur, and disruptive behaviors are not directly observed in these studies.

Fourth, we can assess whether misbehavior affects the pattern of in- and out-transfers from schools. To the best of our knowledge, this is the first paper that reports (and quantifies) these effects. We note that parents, teachers, and school administrators may all have an incentive to encourage disruptive children to move schools, while receiving schools have little information on these children. Asymmetric information and coordination failures may lead to an outcome that is optimal for a school with one or multiple children with persistent behavioral problems, but socially suboptimal, especially if there are adjustment costs to moving schools (unless students who move find a better school match). We also note that reshuffling of this sort is likely to occur in other settings—for example, moving poorly-performing employees around departments in a large company if these workers cannot be fired.

We begin our empirical analysis by assessing the extent to which poorly-behaved students depress the learning outcomes of their classmates. Pooling information across grades 3 through 6, having one or more poorly-behaved children in a classroom lowers classmates' achievement by (on average) .019 SDs.[6]

[6] We begin with 3rd grade because of the definition of poorly-behaved children that we use. To classify a child as poorly-behaved or low-achieving, we need teacher reports from the three previous grades. For example, to classify a 3rd grade child as poorly-behaved, we use teacher assessments from kindergarten to 2nd grade. Third grade is the first grade for which we have three prior measures of student behaviors.

In principle, persistently low-achieving children could also have negative effects on the learning of their classmates if teachers spend an inordinate amount of time helping these children catch up. We find no evidence that this is the case. Although poorly-behaved and low-achieving students are treated in the same way in the Lazear (2001) model—both potentially disrupt the learning of their classmates—they are not equivalent in our sample: the former disrupt learning, while the latter do not. Moreover, with our data, we can rule out that being a slow learner is an important mechanism through which disruptive students affect their peers.

We analyze "dosage" effects and find these to be important. In the sample pooled across grades, having exactly one poorly-behaved student lowers others' achievement by .011 SDs; having exactly two such students has an effect of -.034 SDs; and three or more such children reduce classmates' achievement by .052 SDs. These impacts are relatively large. The effect of having at least three students with persistent behavioral problems is about one-half as large as that of having a one standard deviation better teacher, estimated for kindergarten teachers in this sample (Araujo et al. 2016).

Next, we show that the impacts of being exposed to poorly-behaved children are almost three times larger in 3rd and 4th grades, than in 5th and 6th grades.[7] These results are consistent with those in other papers, including papers using this same experiment, which suggest that younger children may be more sensitive to environmental influences than those who are somewhat older, even within elementary school.[8]

The fact that the same children are assessed repeatedly over time and experience random sequences of poorly-behaved peers allows us to study dynamic effects. We first show that the impact of poorly-behaved peers persists into the future for at least two additional grades.[9] We also find that the

[7] In principle, the fact that the costs of disruption are especially large in earlier grades could be either because the intensity of disruptive behavior is higher, or the sensitivity to disruption is higher in those grades. Using information provided by teachers on specific disruptive behaviors shown by poorly-behaved students, we find no evidence of more intense disruption happening in earlier grades.

[8] Carneiro et al. (2024) analyze the effect of within-classroom achievement rank on performance for children in this sample. More highly-ranked children have higher achievement than those with lower rank, with the largest effects found among children in 1st and 2nd grade. Aizer (2008) also finds evidence that the negative effects of peers with ADD are larger for younger children. She suggests that these may be driven by a higher rate of ADD diagnosis, and treatment (which improves these children's behavior), among somewhat older children. Although we have no data on the prevalence of ADD in our sample, the proportion of ADD children who are diagnosed is likely to be quite small, and the probability that they receive effective treatment even smaller.

[9] An important question is what implications this has in the long run. We cannot analyze long-term effects from our experiment, but a number of studies of young children, primarily in the U.S., have found that the effects on achievement of being in a high-quality preschool, or the impact of better teachers, tend to fade out quickly, but reappear in adulthood in the form of better labor market performance or a lower probability of criminal behavior. See, for example, Chetty et al. (2014), and Jacob et al. (2010) for estimates of the fade-out of the effects of teacher quality, measured by teacher value added. The same could be true about the impacts of disruptive peers, as in Carrell et al. (2018).

effects of poorly-behaved kids significantly affect achievement in end-of-primary (6th grade) assessments, even though disruption in 6th grade does not itself predict achievement in that grade.

We then investigate whether the impact of a poorly-behaved peer in a given grade depends on whether a child was exposed to a poorly-behaved peer in a prior grade—either because of diminishing (negative) returns to experiencing poorly-behaved peers in multiple grades, or alternatively, (negative) dynamic complementarities, which would imply that the total effect of a sequence of bad peers is larger than the sum of the individual effects. Although we cannot give a statistically precise answer to this question, we cannot reject that the effects are additive: that is, the impact of being exposed to a misbehaved peer in a given grade is the same regardless of whether or not a child was exposed to a misbehaved peer in a prior grade.

Turning to outcomes other than achievement, we find no impact of poorly-behaved children on classmates' executive function. On the other hand, there is some evidence that having peers with persistent behavioral difficulties negatively affects non-cognitive skills in 6th grade, although the effects are only significant when there are multiple such students in a classroom. Specifically, having three or more poorly-behaved peers reduces the composite measure of non-cognitive outcomes in 6th grade by .133 SDs.

Finally, we analyze the pattern of transfers in and out of our sample of schools. As we show, children are *not* more likely to leave a school if they are randomly assigned to classrooms with poorly-behaved children. On the other hand, poorly-behaved children are themselves more likely to transfer out of our sample of schools, and in-transfers are more likely to exhibit behavioral problems. Moreover, in-transfers are reported to be poorly-behaved by their teachers even three years after they first arrive in a school, suggesting that this is not simply an adjustment period associated with changing schools. We take this pattern as indirect evidence that the children with the worst behavioral problems are "passed around" schools—either because parents of misbehaving kids are looking for a better match for their child or because school administrators encourage them to leave. Understanding the mechanisms behind this reshuffling of children with behavioral problems is an important area for future research.

The remainder of the paper proceeds as follows. We discuss the setting, data, and experimental design in section 2. Section 3 presents our empirical specification. Section 4 discusses results. Section 5 concludes.

2. Data and experimental design

The data we use come from an experiment in 202 schools in Ecuador, a middle-income country in South America.[10] Schools have at least two classrooms per grade (most have exactly two). A cohort of children entering kindergarten was randomly assigned to classrooms (within schools) in the 2012 school year, and then randomly reassigned them to classrooms in every grade between 1st and 6th grade. Compliance with the assignment rules was very high—98.9 percent on average. We provide further details on the classroom assignment rules and compliance with randomization in Appendix A.

We have baseline data on maternal education, household wealth, whether a child attended preschool, and her vocabulary skills at the beginning of kindergarten. Data on math and language achievement was collected at the end of each grade between kindergarten and 6th grade. For both subjects, tests were a mixture of material that teachers were meant to have covered explicitly in class—for example, in math, addition or subtraction; material that would have been covered, but probably in a somewhat different format—for example, simple word problems; and material that would not have been covered at all in class but that has been shown to predict current and future math achievement—for example, the Siegler number line task (Siegler and Booth 2004). We aggregate responses in math and, separately, language, by Item Response Theory (IRT), and calculate an average achievement score that gives the same weight to math and language.[11]

Child executive function (EF) was assessed in every grade between kindergarten and 4th grade. EF includes a set of basic self-regulatory skills which involve various parts of the brain, but in particular the prefrontal cortex. Low levels of EF are associated with low levels of self-control and "externalizing" behavior, including disruption, aggression, and inability to sit still and pay attention (Séguin and Zelazo 2005). Executive function in childhood has also been shown to predict a variety of outcomes in adulthood, including performance in the labor market, involvement in criminal activities, and health status, even after controlling for socioeconomic status in childhood (Moffitt et al. 2011).

Executive function is generally thought of as having three domains: working memory, inhibitory control, and cognitive flexibility. We separately calculate scores for each of these domains, as well as an average EF score that gives the same weight to each component. In 6th grade, finally, data was collected on child depression, self-esteem, growth mindset, and grit. For each outcome, we aggregate responses by

[10] Araujo et al. (2016) discuss in detail the selection of schools in this study. They show that the characteristics of students and teachers in our sample are very similar to those of students and teachers in a nationally representative sample of schools in Ecuador.

[11] Our results are very similar if, instead, we calculate a simple sum of correct responses within blocks of questions on each test and give equal weight to each of these test sections (as in Araujo et al. 2016).

factor analysis, and also calculate an overall non-cognitive score that gives the same weight to each of the individual assessments. Further details on child assessments are provided in Appendix B.

At the end of each grade, teachers were asked to list the 5 children with the most severe behavioral problems and, separately, the 5 children with the lowest achievement in their class. In our main specification, we use these data to classify a child as *poorly-behaved* if teachers in the three previous grades reported them to be one of the worst-behaved children and define *low-achieving* students in a comparable way. We could have used alternative definitions along two dimensions: (i) take only the 4 lowest, 3 lowest, 2 lowest, or the single lowest-rated child in the classroom (rather than 5-lowest, as we do) to categorize children as poorly-behaved or low-achieving; or (ii) use fewer or more consecutive grades to categorize children (rather than 3, as we do). For example, we could have defined as poorly-behaved a child who was rated in the bottom 3 in the classroom for 4 consecutive years, which would be a more stringent definition along both dimensions. The definition we choose trades off stringency and statistical power. We show in Appendix D that our point estimates are robust to deviations from this definition.

Importantly, our experiment generates considerable variation in exposure to poorly-behaved (and low-achieving) students. This can be seen in Table 1, which shows the number of poorly-behaved students in each grade (column 1); the number of classrooms with poorly-behaved students and the number of total classrooms (columns 2 and 3, respectively); the proportion of classrooms with at least one disruptive student (column 4), and with one, two or three or more poorly-behaved students (columns 5 to 7). Appendix table C2 shows similar statistics concerning low-achieving students.

Table 2 shows how persistent is disruptive behavior. Recall that in order to classify a student as poorly-behaved in a given grade *g* we use information about his behavior exclusively in prior grades (*g-1* to *g-3*). In other words, we do not use any information in grade *g* to define someone as being poorly-behaved in that same grade. Therefore, it is useful to check whether, as expected, students who are defined to be poorly-behaved in grade *g* actually misbehave in that grade, and are rated by the grade *g* teacher as being among the worst behaved students in the classroom in that grade. Column 1 shows that this is indeed the case. In 64 percent to 73 percent of classrooms with a student who is classified to be poorly-behaved according to our definition, the teacher rates the poorly-behaved student as being among the worst-behaved students in the classroom in that grade. In other words, low ratings of behavior in past grades (used to define poorly-behaved students) are strong (but not perfect) predictors of low rating of behavior in the current grade. In columns 2 to 4 we show whether poorly-behaved students in grade *g* remain in the bottom five in terms of behavior in subsequent grades. Poorly-behaved students in grade *g* continue to have behavioral issues in subsequent grades, although less so over time. Nevertheless, the degree of persistence is remarkably large. These results justify the definition of poorly-behaved student we use in the paper.[12]

Table 3 provides summary statistics for children in our sample, comparing those who are classified as poorly-behaved in at least one grade, using teacher reports from the three previous grades, and those who are not. It shows that children who are not poorly-behaved were 5 years of age on the first day of kindergarten, on average, and half of them are girls. Mothers were in their early 30s and fathers in their mid-30s. About 70 percent of both parents had attained less than secondary education. Araujo et al. (2016) report that, at the beginning of kindergarten, the average receptive vocabulary score of children in the sample is 1.7 SDs below the level of children that were used to norm the test.[13]

Turning to the comparison between poorly-behaved and other students, Table 3 also shows the characteristics of poorly-behaved children. Children with persistent behavioral problems are overwhelmingly male—over 95 percent of them are boys. They have lower performance on math and language tests than other children, and lower levels of executive function. Poorly-behaved students also have worse depression scores, lower levels of self-esteem, lower levels of grit, and lower values for the measure of growth mindset than other children.[14] Broadly speaking, the socioeconomic status of poorly-behaved students is slightly worse compared to other children: poorly-behaved students are more likely to have fathers who attained less than secondary education, and household wealth is lower. Poorly-behaved students are more likely to have attended preschool than other children, a difference of about 15 percentage points. This may seem surprising, although we note there are several papers which show that prolonged time in daycare can have negative impacts on children's socio-emotional development (see for example Baker et al. 2008, 2019, or Fort et al. 2020).

3. Empirical specification

A. Main specification

[12] Recall that a child is classified as *poorly-behaved* in *g* if he was rated as being among the 5 worst-behaved students in the classroom at the end of grades *g-1*, *g-2* and *g-3*. One way to validate the informativeness of our measure is to check if children who we classify as disruptive in *g* (based on past information) also exhibit poor behaviors in *g+1*, and Table 2 shows that this is indeed the case. Much the same holds for children who are persistently low-achieving: between 65 percent and 71 percent of children who are listed as having the biggest difficulties learning in *g-2*, *g-1* and *g* are also listed as such by their teachers in *g+1*.

[13] To measure baseline receptive vocabulary, we use the *Test de Vocabulario en Imágenes Peabody* (TVIP) (Dunn et al. 1986), the Spanish-speaking version of the much-used Peabody Picture Vocabulary Test (PPVT). The TVIP has been used widely to measure development among Latin American children—see, for example Schady et al. (2015).

[14] For the comparisons in Table 3, we use *lagged* achievement and executive function. We cannot do this for the measures of depression, self-esteem, growth mindset and grit, as these were only collected in 6th grade.

Our main goal is to estimate whether child *i* in classroom *c*, grade *g*, and school *s* has lower achievement, executive function, or non-cognitive development after she was randomly assigned to classrooms with, or without, poorly-behaved students. For this purpose, we pool observations between grades 3 and 6 and run regressions of the following form:

$$Y_{i,c,g,s} = \beta D_{c,g,s} + \varphi_g\left(Y_{i,c,g-1,s}\right) + \theta_g X_{i,c,g,s} + \delta_{g,s} + \varepsilon_{i,c,g,s} \tag{3.1}$$

where $D_{c,g,s}$ is an indicator variable that takes on the value of one if there is one or more poorly-behaved students in a classroom; the function $\varphi(.)$ is a flexible formulation of lagged achievement (in particular, a fourth-order polynomial in lagged achievement); $X_{i,c,g,s}$ includes child age and gender, as well as an indicator variable for whether student *i* himself is a poorly-behaved student;[15] $\delta_{g,s}$ is a set of school-by-grade fixed effects; and $\varepsilon_{i,c,g,s}$ is a residual. In this model, β is restricted to be the same across all grades, but all other parameters are allowed to be grade-specific. Since in any grade *g* our definition of poorly-behaved student is based on their behavior in all of the three previous grades, we are able to evaluate effects of poorly-behaved students on their peers' outcomes between grades 3 and 6.

Other estimates we report are variants on this basic formulation. Specifically, we estimate (1) regressions that refer to outcomes measured one or two grades after a student was exposed to a poorly-behaved peer, not just those that refer to contemporaneous effects; (2) models in which the coefficients on β are allowed to vary by grade, rather than restricted to be the same across all grades; (3) models that allow effects to vary with the number of poorly-behaved students in a classroom; and (4) models that include separate indicator variables for poorly-behaved and low-achieving children. Standard errors are clustered at the classroom and student level when pooling data across grades, and at the classroom level in the individual grade regressions.

B. Executive function and non-cognitive skills

To estimate the effect of poorly-behaved students on non-cognitive skills in 6th grade we run a regression comparable to (3.1), replacing achievement in grade *g* with the relevant outcome. To estimate the effect of poorly-behaved students on executive function, we also use the model in (3.1), but use a fourth-order

[15] In our sample, new students enter a school and are incorporated in the randomization of students to classrooms in every grade. For these students, we do not know whether they are poorly-behaved or not as we do not have information on their past behaviors. Thus, among the controls, we also include an indicator variable which takes value one if student *i* is present in grade *g*, but does not have sufficient information on past behavior to be classified as being a poorly-behaved student.

polynomial in lagged executive function. These regressions use information in 3rd through 4th grades, where data on current and lagged executive function are available.

C. Dynamics

We now extend equation (3.1) to allow for interactions between exposure to poorly-behaved peers in different time periods. One simple alternative is to introduce an interaction between lagged achievement and exposure to a poorly-behaved peer:

$$Y_{i,c,g,s} = \beta_g D_{c,g,s} + \sigma_g D_{i,c,g,s} Y_{i,c,g-1,s} + \varphi_g\left(Y_{i,c,g-1,s}\right) + \theta_g X_{i,c,g,s} + \delta_{g,s} + \varepsilon_{i,c,g,s} \tag{3.2}$$

We could also allow for more flexible specifications by estimating a CES or a translog production function, as in Cunha et al. (2010) or Agostinelli and Wiswall (2024) for example. In addition, we could relax the assumption that the production of learning follows a first-order Markov process (as in Attanasio et al. 2020) and estimate:

$$Y_{i,c,g,s} = \beta_g D_{c,g,s} + \tau_g D_{i,c,g-1,s} + \sigma_g D_{i,c,g,s} D_{i,c,g-1,s} + \varphi_g\left(Y_{i,c,g-2,s}\right) + \theta_g X_{i,c,g,s} + \delta_{g,s} + \varepsilon_{i,c,g,s} \tag{3.3}$$

This allows the impact of peer exposure in the past to affect current learning over and beyond its impact on past learning.[16]

We also estimate the cumulative impact of poorly-behaved peers on achievement at the end of elementary school. To this end, we estimate the following model:

$$Y_{i,c,6,s} = \sum_{k=1}^{4} \lambda_k ND_{c,k,s} + \varphi_6\left(Y_{i,c,2,s}\right) + \theta_6 X_{i,c,6,s} + \delta_{6,s} + \varepsilon_{i,c,6,s} \tag{3.4}$$

where $ND_{c,k,s}$ ($k = 1, \dots, 4$) is an indicator variable which takes value 1 if the student was exposed to a poorly-behaved peer for *k* grades, between grades 3 and 6, and λ_k measures the impact on 6th grade (end of elementary school) test scores of having been in a classroom with a poorly-behaved peer for *k* grades between grades 3 and 6 (relative to never having been in a classroom with a poorly-behaved peer in any of these grades).

[16] To assess the sensitivity of our findings to different specifications, we also present estimates of the following equation in Appendix E: $Y_{i,c,g,s} = \beta_g D_{c,g,s} + \tau_g D_{i,c,g-1,s} + \sigma_g D_{i,c,g,s} D_{i,c,g-1,s} + \varphi_g\left(Y_{i,c,g-1,s}\right) + \varphi_g\left(Y_{i,c,g-2,s}\right) + \theta_g X_{i,c,g,s} + \delta_{g,s} + \varepsilon_{i,c,g,s}$, which is analogous to (3.3), but in addition to $Y_{i,c,g-2,s}$ we also control for $Y_{i,c,g-1,s}$.

4. Results

A. Main specification

Our first set of results is in Table 4, where the outcome of interest is the average of math and language achievement in each grade. The first row of Panel A shows estimates of equation (3.1) in each grade. The other rows of this panel correspond to variants of this equation where achievement is measured at a later point in time ($Y_{i,c,g+n,s}$, where *n=1,2,3*) than that when children were, or were not, exposed to a poorly-behaved classmate ($D_{c,g,s}$). In the first column of the panel the coefficient on $D_{c,g,s}$ in equation (3.1) is restricted to be the same for all grades ($\beta_g = \beta$), while in the remaining columns this restriction is relaxed. Each coefficient in the table corresponds to an entirely separate regression.

Starting with the first row of the table, we see that, in the model that restricts coefficients to be the same across grades, having at least one poorly-behaved student in a class lowers the achievement of classmates by .019 SDs. The effects generally fall by grade, and we can reject the null that the average effect for 3rd and 4th graders, and that for 5th and 6th graders are the same (p-value: .052, reported in the last column of this row).[17]

There is mixed evidence that the impact of being exposed to poorly-behaved students in grade *g* persists in subsequent grades. In the grade-specific models, the impact of exposure to a poorly-behaved peer in 3rd grade on test scores in 3rd (lag 0), 4th (lag 1), 5th (lag 2) and 6th (lag 3) grades are -.032, -.017, -.012, and .002, respectively, and we can reject the null that these effects are the same (p-value: .019, reported in the last row of this panel). In the case of exposure to a poorly-behaved peer in 4th grade, the impacts are more persistent, although the number of lags we can consider is also smaller.[18]

Panel B turns to the comparison between the effects of poorly-behaved and low-achieving students, showing results of regressions in which we include an indicator for having a poorly-behaved student in the classroom and an indicator for having a low-achieving student in the classroom simultaneously. As discussed earlier, both categories of children could disrupt their peers' learning. Each column corresponds to a separate regression where these two variables are included simultaneously. The

[17] The fact that poorly-behaved students have larger, negative effects on their classmates in the earlier grades could occur either because when they are younger, poorly-behaved students engage in behaviors that are more disruptive than when they are older (for example, biting a classmate), or because older children are better able to pay attention than younger children even when there is classroom disruption, or some combination of both. Using information provided by teachers on specific disruptive behaviors done by poorly-behaved students, we find no evidence that poorly-behaved students engage in more intense disruptive behaviors in earlier grades, as shown in appendix figure B4.

[18] The pooled results in column (1) are less informative to study fadeout as they could confound differences in impacts by grade with the number of lags that can be included for the calculation.

coefficients on low-achieving peers are always close to zero. In the pooled regression, we can reject the null that the effects of poorly-behaved and low-achieving students are the same (p-value: .013).

Panel C asks whether the impact of poorly-behaved students increases with the number of such students in a classroom—what we refer to as dosage effects. For this purpose, we estimate a version of equation (3.1) that expands our explanatory variable, $D_{c,g,s}$, from a single indicator for whether there was at least one poorly-behaved child in the classroom, to three indicators for whether there were 1, 2, or 3 or more poorly-behaved children in the classroom. As in Panel B, each column corresponds to a separate regression where these three variables are included simultaneously. Across grades, there is clear evidence that having more poorly-behaved students in the classroom is worse than having fewer of them: in the estimates that pool across grades, having exactly 1, exactly 2, and 3 or more students with persistent behavioral difficulties lowers the learning of other children in the classroom by .011, .034 and .052 SDs, respectively, and we can reject the null that these effects are equal to each other (p-value .000).

In sum, Table 4 shows that students with persistent behavioral difficulties harm their classmates' achievement; that the effects are concentrated among younger students; that the negative impacts of poorly-behaved students on achievement persist for at least two years; that there are dosage effects (the more poorly-behaved children there are in a class, the larger is the negative effect on the achievement of their classmates); and that having students who are persistently low-achieving, as reported by their teachers, does not lower the learning outcomes of their classmates.

Appendix D presents robustness checks to different definitions of "poorly-behaved" students. We can change the definition of what constitutes a poorly-behaved student along two dimensions: (i) the teacher rating cutoff—so, considering as poorly-behaved only the 4 worst, 3 worst, 2 worst, or the single worst-rated child in the classroom, instead of using the 5 worst-behaved children, as we do in Table 4; (ii) the number of years considered—so, considering as poorly-behaved those ranked at the bottom of the classroom for different numbers of consecutive years. Using higher (lower) rank cutoffs or more (less) consecutive years to define poorly-behaved students makes our definition more (less) stringent. A more stringent definition implies fewer poorly-behaved students, and less variation in the number of classrooms exposed to a poorly-behaved student. Appendix tables D1 to D6 show that our main results are robust to the definition of "poorly-behaved" students we use.[19]

B. Executive function and non-cognitive skills

[19] Similarly, appendix tables D7 to D12 show that our results on the effects of low achieving students are robust to the definition of "low achieving" students we use.

We turn to other outcomes in Table 5. Each outcome is in a different column. Below the standard errors we report p-values computed based on the Romano-Wolf stepdown procedure using 5,000 bootstrap replications in square brackets (see Clarke et al. 2020; Romano and Wolf 2005). The first four columns correspond to executive function. We present results from models that restrict coefficients to be the same across grades 3 and 4 (we do not have measures of executive function for grades 5 and 6). The first column aggregates different measures of executive function, and individual impacts on inhibitory control, memory and attention, and cognitive flexibility are shown in columns 2 to 4. Similarly, column 5 aggregates non-cognitive skills into a single index, and results for individual components of this index are shown in the subsequent columns. Panel A corresponds to the estimates of equation (3.1), while Panel B considers an extension of equation (3.1) that accounts for dosage effects.

Panel A shows there is no evidence that having poorly-behaved peers in the classroom lowers the scores on the measures of classmates' inhibitory control, working memory, or cognitive flexibility, or on the composite measure of executive function. Panel B shows that this is the case for children exposed to a single, but also multiple, poorly-behaved students.

Other columns in the table focus on the effects of poorly-behaved peers on depression, self-esteem, growth mindset, and grit. In Panel A, the coefficients from these regressions are consistently negative, but they are not significant. Moreover, because we only collected data on these outcomes in 6^{th} grade, we cannot pool data across grades (as we do with achievement) to increase precision. That said, here too we find evidence of dosage effects, as can be seen in Panel B. In the regression that focuses on the non-cognitive aggregate, we find that having three or more poorly-behaved peers in the classroom reduces the score by .133 SDs. The clearest negative effect of multiple students with persistent behavioral problems is on growth mindset.

C. Dynamics

In Table 4 we documented two important results regarding the dynamics of skill formation: (i) the impacts of poorly-behaved peers are larger in 3^{rd} and 4^{th} grade than in 5^{th} or 6^{th} grade; and (ii) these impacts persist for two additional grades. The model on which these estimates are based (equation 3.1) is, however, additive in previous achievement and current exposure to a poorly-behaved peer. This implicitly assumes that poorly-behaved peers in different time periods are substitutes in the production of learning, and rules out by assumption the possibility that there are diminishing returns or, alternatively, dynamic complementarities in the impacts of poorly-behaved peers over time (as in Cunha and Heckman 2007). In the former case, the total impact of effects of exposure to poorly-behaved children in multiple grades would be smaller, and in the latter case, larger, than the sum of the individual (grade-specific) effects.

Estimates of equations (3.2) and (3.3) are shown in Panel A of Table 6. In both models we cannot reject that the model is additive (i.e., we cannot reject that σ_g is equal to zero in either model), although the relevant parameters are imprecisely estimated. This suggests that equation (3.1) (and our estimates in Table 4) is likely a good approximation to the process of skill accumulation in our dataset.[20]

Estimates of (3.4) are shown in panel B of Table 6. They suggest that end-of-6th-grade scores decline with the number of times a child was paired with a poorly-behaved peer in the classroom. However, because our estimates lack precision, we cannot reject that the parameters reported in this panel are statistically equal to each other.

D. Patterns of cross-school transfers

In Table 7, we turn to patterns of transfers in and out of our sample of schools. In Panel A we show the impact of having a poorly-behaved peer (column 1), or of being a poorly-behaved student (column 2), on the likelihood of leaving the sample between two consecutive grades. Children in our sample are no more likely to attrit when they are exposed to poorly-behaved students. Therefore, it is unlikely that our estimates are affected by selective attrition. On the other hand, the grade-on-grade attrition rate of poorly-behaved students is higher by 2.4 percentage points (relative to average grade-to-grade attrition of 8 percent). The converse is also true: children who move out of the school in a given grade are more likely to be classified as poorly behaved in the previous grade.

Note also that just as students from the schools in our sample are being *sent* to other schools, the schools in our sample are *receiving* students transferring from elsewhere. In panel B we estimate how the probability of being poorly-behaved differs between new entrants and children who were already in a given school. In column 1 of panel B, we regress an indicator for being listed by a teacher in a given grade as a child with behavioral problems on an indicator for being a new entrant into the sample. In-transfers are 0.7 percentage points more likely to be reported as having behavioral problems than other children.[21] Furthermore, new entrants in a given grade *g* are not only more likely to be poorly-behaved in that grade, but they are also more likely to be poorly-behaved in future grades. Column 2 of panel B of the table shows that, three years after they first arrived, in-transfers are 0.7 percentage points more likely to be reported as being persistently *poorly-behaved* than other children.

[20] In Appendix table E1 we present an additional specification where we include $Y_{i,c,g-1,s}$ in addition to $Y_{i,c,g-2,s}$.

[21] Hanushek et al. (2004) also find that children who move schools perform worse than other schools, and that movers reduce achievement in receiving classrooms.

We do not know why poorly-behaved children are more likely to move around schools. The decision could be voluntary, or a response to pressure from other parents, principals, and teachers. It is also not clear whether the effect of reshuffling disruptive children across schools is on aggregate positive (because children find a better school match) or negative (because disruptive children have a hard time adjusting to a new environment, potentially causing even more disruption in their new classrooms). Regardless, the reshuffling of poorly-behaved children means that we are likely to underestimate the effects of disruptive children on learning that we report in our paper.[22]

5. Conclusion

In this paper, we show that children with persistent behavioral problems lower the achievement of their peers. We document dosage effects—the more poorly-behaved children there are in a class, the larger is the negative effect on the achievement of their classmates. While in principle both *poorly-behaved* and *low-achieving* children could disrupt learning, in practice, only those with persistent behavioral problems lower the achievement of their classmates in the setting we study. The impact of poorly-behaved students on learning is larger in earlier than in later grades, and it persists for at least two grades. Repeated exposure to poorly-behaved peers accumulates additively, and leads to lower achievement at the end of elementary school.

The fact that children who have persistent behavioral problems have negative effects on their classmates raises important questions. How can policy-makers best ensure that any underlying medical conditions, like ADD, are diagnosed and treated? Should children with persistent behavioral problems be mainstreamed or placed in special needs classrooms? If children with persistently poor behavior are kept in regular classes, what are tools that teachers can use to effectively manage misbehavior?[23] Providing answers to these questions is difficult. Designing effective policies for children who have persistent behavioral problems is likely to be particularly challenging in developing countries, where resources are more limited.

---

[22] This is because a child is classified as *poorly-behaved* in *g* if she was rated as being among the 5 worst-behaved students in the classroom at the end of grades *g-1*, *g-2* and *g-3*. Therefore, no matter how disruptive his behavior, a new arrival in 4th, 5th, and 6th grades cannot be classified as poorly-behaved.

[23] Developing countries spend considerable resources on in-service training for teachers, but most programs appear to be ineffective (see Popova et al. 2022). A coaching program for 1st grade teachers, implemented in a different sample of urban schools in Ecuador, did not raise achievement, and may have worsened outcomes as teachers struggled to change their in-class behaviors (see Carneiro et al. 2022).

Aizer, A. 2008. "Peer Effects and Human Capital Accumulation: The Externalities of ADD." National Bureau of Economic Research Working Paper 14354.

Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo, and N. Schady. 2016. "Teacher Quality and Learning Outcomes in Kindergarten." *Quarterly Journal of Economics* 131(3): 1415-53.

Attanasio, O. P., R. Bernal, M. Giannola, and M. Nores. 2020. "Child Development in the Early Years: Parental Investment and the Changing Dynamics of Different Dimensions". NBER WP 27812.

Baker, M., J. Gruber, and K. Milligan. 2008. "Universal Child Care, Maternal Labor Supply, and Family Well-Being." *Journal of Political Economy* 116(4): 709-45.

----------. 2019. "The Long-Run Impacts of a Universal Child Care Program." *American Economic Journal: Economic Policy* 11(3): 1-26.

Balestra, S., B. Eugster, and H. Liebert. 2022. "Peers with Special Needs: Effects and Policies." *Review of Economics and Statistics* 104(3): 602-618.

Booij, A. S., E. Leuven, and H. Oosterbeek. 2017. "Ability Peer Effects in University: Evidence from a Randomized Experiment." *Review of Economic Studies* 84(2): 547–78.

Bureau of Labor Statistics. 2024. "News Release: Usual Weekly Earnings of Wage and Salary Workers, First Quarter 2024". Available at https://www.bls.gov/news.release/pdf/wkyeng.pdf. Accessed on April 24, 2024.

Carneiro, P., Y. Cruz-Aguayo, R. Intriago, J. Ponce, N. Schady, and S. Schodt. 2022. "When Promising Interventions Fail: Personalized Coaching for Teachers in a Middle-Income Country." *Journal of Public Economics Plus* 3.

Carneiro, P., Y. Cruz-Aguayo, F. Salvati, and N. Schady. 2024. "The Effect of Classroom Rank on Learning Throughout Elementary School: Experimental Evidence from Ecuador." Forthcoming, *Journal of Labor Economics*.

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(1): 211-28.

Carrell, S. E., M. L. Hoekstra, and E. Kuka. 2018. "The Long-Run effects of Disruptive Peers." *American Economic Review* 108(11): 3377-3415.

Carrell, S. E., R. L. Fullerton, and J. E. West 2009. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." *Journal of Labor Economics* 27(3): 439-464.

Center for the Developing Child. 2020. "Executive Function & Self-Regulation." Available at https://developingchild.harvard.edu/science/key-concepts/executive-function/. Accessed on March 15, 2020.

Chetty, R., J. Friedman, and J. Rockoff. 2014. "Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104(9): 2633–79.

Clarke, D., J. P. Romano, and M. Wolf. 2020. "The Romano–Wolf Multiple-Hypothesis Correction in Stata." *The Stata Journal* 20(4): 812-843.

Cunha, F., and J. J. Heckman. 2007. "The Technology of Skill Formation." *American Economic Review, Papers and Proceedings* 97(2): 31-47.

De Giorgi, G., and M. Pellizzari. 2013. "Understanding Social Interactions: Evidence from the Class Room." *Economic Journal* 124(579): 917–53.

Duflo, E., P. Dupas, and M. Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101(5): 1739–74.

Feld, J., and U. Zölitz. 2017. "Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects." *Journal of Labor Economics* 35(2): 387-428.

Feld, J., and U. Zölitz. 2022. "The Effect of Higher-Achieving Peers on Major Choices and Labor Market Outcomes." *Journal of Economic Behavior & Organization* 196: 200-219.

Figlio, D. N. 2007. "Boys Named Sue: Disruptive Children and Their Peers." *Education Finance and Policy* 2(4): 376-94.

Fort, M., A. Ichino, and G. Zanella. 2020. "Cognitive and Noncognitive Costs of Day Care at Age 0–2 for Children in Advantaged Families." *Journal of Political Economy* 128(1): 158-205.

Golsteyn, B. H., A. Non, and U. Zölitz. 2021. "The Impact of Peer Personality on Academic Achievement." *Journal of Political Economy* 129(4): 1052-99.

Hanushek, E., and L. Woessmann. 2020. "The Economic Impacts of Learning Losses." Available at https://www.ams-forschungsnetzwerk.at/downloadpub/2020_OECD_The-economic-impacts-of-coronavirus-covid-19-learning-losses.pdf. Accessed on April 24, 2024.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin. 2004. "Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools." *Journal of Public Economics* 88: 1721-46.

Huang, B., and R. Zhu. 2020. "Peer Effects of Low-Ability Students in the Classroom: Evidence from China's Middle Schools." *Journal of Population Economics* 33(4), No. 6: 1343-80.

Jacob, B., L. Lefgren, and D. P. Sims. 2010. "The Persistence of Teacher-Induced Learning." *Journal of Human Resources* 45(4): 915-43.

Kristoffersen, J. H. G., M. V. Krægpøt, H. S. Nielsen, and M. Simonsen. 2015. "Disruptive School Peers and Student Outcomes." *Economics of Education Review* 45(4): 1-13.

Lazear, E. P. 2001. "Educational Production." *Quarterly Journal of Economics* 141(3): 777-803.

Lyle, D. S. 2009. "The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point." *American Economic Journal: Applied Economics* 1(4): 69–84.

Moffitt, T., L. Arseneault, D. Belsky, N. Dickson, R. Hancox, H. Harrington, R. Houts, R. Poulton, B. Roberts, S. Ross, M. Sears, W. M. Thomson, and A. Caspi. 2011. "A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety." *Proceedings of the National Academy of Sciences* 108: 2693–98.

OECD. 2019. "TALOS 2018 Results: Teachers and School Leaders as Lifelong Learners", Volume I, OECD.

Popova, A., D.K. Evans, M.E. Breeding, and V. Arancibia. 2022. "Teacher Professional Development around the World: The Gap between Evidence and Practice." *World Bank Research Observer* 37(1): 107-36.

Romano, J. P., and M. Wolf, M. 2005. "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica* 73(4): 1237-1282.

Sacerdote, B. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *The Quarterly Journal of Economics* 116(2): 681-704.

Santavirta, T., and M. Sarzosa. 2024. "Effects of Disruptive Peers in Endogenous Social Networks." Unpublished manuscript.

Schady, N., J. Behrman, M. C. Araujo, R. Azuero, R. Bernal, D. Bravo, F. Lopez-Boo, K. Macours, D. Marshall, C. Paxson, and R. Vakis. 2015. "Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries." *Journal of Human Resources* 50(2): 446-63.

Seguin, J. and P. Zelazo, "Executive Function in Early Physical Aggression," in Developmental Origins of Aggression, R. Tremblay, V. Hartup, and J. Archer, eds. (New York: Guilford Press, 2005).

Shan, X., and U. Zölitz. 2024. "Peers Affect Personality Development." forthcoming *Review of Economics and Statistics.*

Siegler, R.S., and J.L. Booth. 2004. "Development of Numerical Estimation in Young Children." *Child Development* 75(2): 428-44.

Sojourner, Aaron. 2012. "Identification of Peer Effects with Missing Peer Data: Evidence from Project STAR." *The Economic Journal* 123(June): 574-605.

Xu, D., Q. Zhang, and X. Zhou. 2022. "The Impact of Low-Ability Peers on Cognitive and Non-Cognitive Outcomes: Random Assignment Evidence on the Effects and the Operating Channels." *Journal of Human Resources* 57(2): 555-96.

Zárate, R. A. 2023. "Uncovering Peer Effects in Social and Academic Skills." *American Economic Journal: Applied Economics* 15(3): 35-79.

Zhao, L., and Z. Zhao. 2021. "Disruptive Peers in the Classroom and Students' Academic Outcomes: Evidence and Mechanisms." *Labour Economics* 68, Issue C.

Table 1: Distribution of poorly-behaved students across classrooms

	(1) Nr poorly-behaved students	(2) Nr classrooms with poorly-behaved student	(3) Total nr classrooms	(4) Proportion classrooms with poorly-behaved student	(5) Proportion classrooms with 1 poorly-behaved student	(6) Proportion classrooms with 2 poorly-behaved students	(7) Proportion classrooms with 3+ poorly-behaved students
3rd grade	299	206	470	.44	.33	.08	.03
4th grade	338	224	479	.47	.32	.11	.04
5th grade	439	276	485	.57	.36	.14	.07
6th grade	490	300	485	.62	.37	.17	.08

Note: The table shows descriptive statistics about the distribution of poorly-behaved students across classrooms in every grade. In any grade t between grades 3 and 6, a poorly-behaved student is a student who was ranked among the bottom 5 worst behaved students in the classroom according to the teacher in all grades between t-1, t-2 and t-3. In columns 1-2 we show how many students comply with this definition in each grade between 3rd and 6th grade, and the number of classrooms in which there is a poorly-behaved student according to this definition. In columns 3-7 we show the total number of classrooms in the sample, as well as the proportion of classrooms with a poorly-behaved student, and the proportion of classrooms with one, two, or three or more poorly-behaved students.

Table 2: Persistence in poor behavior

	(1) Proportion poorly-behaved students among bottom 5 in *g*	(2) Proportion poorly-behaved in *g+1*	(3) Proportion poorly-behaved in *g+2*	(4) Proportion poorly-behaved in *g+3*
3rd grade	0.64	0.65	0.49	0.40
4th grade	0.73	0.77	0.60	
5th grade	0.68	0.71		
6th grade	0.66			

Note: The table shows descriptive statistics about the persistence in poor behavior. In any grade t between grades 3 and 6, a poorly-behaved student is a student who was ranked among the bottom 5 worst behaved students in the classroom according to the teacher in all grades between g-1, g-2 and g-3. Each column shows the proportion of poorly-behaved students in any given grade g who are also poorly-behaved in grades g+1, g+2 and g+3 using our definition, conditional on attrition.

Table 3: Characteristics of poorly-behaved students

	Poorly-behaved		Not poorly-behaved			
	Mean	N	Mean	N	Diff.	p-value
Female	.049	881	.490	27,513	-.442	.000
Age at baseline (months)	68.1	591	67.6	14,942	.526	.011
Math	-.322	880	-.029	20,843	-.293	.000
Language	-.464	880	-.034	20,855	-.430	.000
Math+Language index	-.393	880	-.031	20,843	-.362	.000
Executive function	-.278	880	-.016	19,535	-.261	.000
Executive function: inhibitory control	-.141	880	-.017	19,535	-.125	.000
Executive function: memory and attention	-.263	880	-.014	19,535	-.249	.000
Executive function: cognitive flexibility	-.116	880	-.006	19,535	-.110	.000
Aggregate non-cognitive	-.244	356	.012	7,433	-.256	.000
Depression	-.193	356	.009	7,433	-.202	.000
Self-esteem	-.160	356	.008	7,433	-.167	.002
Growth mindset	-.224	356	.011	7,433	-.235	.000
Grit	-.181	356	.009	7,433	-.190	.000
Mother education less than secondary	.689	546	.685	13,146	.004	.846
Father education less than secondary	.756	390	.692	10,237	.065	.007
Mother age	29.6	543	30.1	13,067	-.496	.073
Father age	34.5	374	34.5	9,963	.039	.923
Wealth	-.070	584	.003	13,886	-.073	.084
Vocabulary at baseline (TVIP)	-.022	567	.001	13,729	-.023	.586
Preschool	.708	592	.562	15,039	.146	.000

Notes: This table shows characteristics of poorly-behaved students according to our main definition, and compares them to non-poorly-behaved students. In any grade g between grades 3 and 6, a poorly-behaved student is a student who was ranked among the bottom 5 worst behaved students in the classroom according to the teacher in all grades between g-1, g-2 and g-3. The table reports the mean of each variable for poorly-behaved and non-poorly-behaved students, as well as the difference in means between poorly-behaved students and non-poorly-behaved students, and the p-values testing whether the differences in means are equal to zero, pooling across grades 3 to 6. Data on executive function are only available up to grade 4.

Table 4: Effect of poorly-behaved students on student achievement

Panel A	Pooled	3rd grade	4th grade	5th grade	6th grade	p-value 3rd grade =4th grade =5th grade =6th grade	p-value avg 3rd and 4th grade = avg 5th and 6th grade
Has poorly-behaved student - Lag 0	-.019***	-.032***	-.028**	-.007	-.011	.255	.052
	(.005)	(.011)	(.012)	(.010)	(.009)		
Has poorly-behaved student - Lag 1	-.015***	-.017*	-.019**	-.010			
	(.005)	(.009)	(.008)	(.008)			
Has poorly-behaved student - Lag 2	-.019***	-.012	-.025***				
	(.006)	(.010)	(.008)				
Has poorly-behaved student - Lag 3	.002	.002					
	(.010)	(.010)					
p-value Lag 0 = Lag 1 = Lag 2 = Lag 3	.089	.019	.593	.720			

Panel B	Pooled	3rd grade	4th grade	5th grade	6th grade	p-value 3rd grade =4th grade =5th grade =6th grade	p-value avg 3rd and 4th grade = avg 5th and 6th grade
Has poorly-behaved student	-.019***	-.032***	-.028**	-.007	-.010	.244	.048
	(.005)	(.011)	(.012)	(.010)	(.009)		
Has low achieving student	-.001	-.007	.004	.001	-.001	.929	.911
	(.006)	(.011)	(.013)	(.009)	(.009)		
p-value poorly-behaved=low achieving	.013	.124	.069	.566	.482		

Panel C	Pooled	3rd grade	4th grade	5th grade	6th grade	p-value 3rd grade =4th grade =5th grade =6th grade	p-value avg 3rd and 4th grade = avg 5th and 6th grade
1 student	-.011**	-.027**	-.026*	.008	.001	.056	.007
	(.005)	(.012)	(.014)	(.010)	(.009)		
2 students	-.034***	-.047**	-.028	-.034**	-.025*	.817	.644
	(.007)	(.020)	(.018)	(.013)	(.013)		
3+students	-.052***	-.041	-.086**	-.030	-.061***	.529	.484
	(.013)	(.037)	(.033)	(.020)	(.019)		
p-value 1 student = 2 students = 3+ students	.000	.625	.173	.003	.002		

Notes: Panel A reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a poorly-behaved student for various lags of year of assignment to such a classroom (where lag 0 captures the contemporaneous effect), and for various grades. Panel B reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a poorly-behaved student and an indicator for being randomly assigned to a classroom with a low achieving student, for various grades. In any grade g between grades 3 and 6, a low achieving student is a student who was ranked among the bottom 5 worst achieving students in the classroom according to the teacher in all grades between g-1, g-2 and g-3. Panel C reports estimates from regressions of an index of math and language scores on indicators for the number of poorly-behaved students in the classroom (omitted category is 0), for various grades. Students can be assigned to classrooms with one, two, or three or more poorly-behaved students. In all panels, column 1 pools information across grades 3-6; and columns 2-5 report estimates from regressions by grade. In panel A, each regression controls for a fourth-order polynomial in lagged achievement, an indicator for a poorly-behaved student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. In panel B, each regression controls for a 4th-order polynomial in lagged achievement, an indicator for being a poorly-behaved or low achieving student, child gender and age, and school (by grade when pooling information across grades) fixed effects. In panel C, each regression controls for a fourth-order polynomial in lagged achievement, an indicator for being a poorly-behaved student, child age and gender, and school (by grade when pooling information across grades) fixed effects. Standard errors are clustered at the classroom and student level when pooling data, and at the classroom level in the regressions by grade.

Table 5: Effect of poorly-behaved students on student executive function and non-cognitive outcomes

Panel A	EF composite	Inhibitory control	Memory and attention	Cognitive flexibility	Aggregate non-cognitive	Depression	Self-esteem	Growth mindset	Grit
Has poorly-behaved student	.013	-.021*	.018	-.006	-.042	-.036	-.025	-.025	-.051*
	(.012)	(.012)	(.013)	(.013)	(.026)	(.027)	(.026)	(.024)	(.026)
	[.877]	[.442]	[.614]	[.982]	[.450]	[.641]	[.880]	[.856]	[.251]
Panel B	**EF composite**	**Inhibitory control**	**Memory and attention**	**Cognitive flexibility**	**Aggregate non-cognitive**	**Depression**	**Self-esteem**	**Growth mindset**	**Grit**
1 student	.020	-.025*	.030**	-.002	-.039	-.017	-.022	-.021	-.058**
	(.014)	(.013)	(.014)	(.014)	(.029)	(.028)	(.029)	(.027)	(.028)
	[.614]	[.347]	[.206]	[.997]	[.641]	[.938]	[.937]	[.937]	[.232]
2 students	-.002	-.014	-.007	-.018	-.017	-.081**	-.012	.008	-.013
	(.019)	(.023)	(.020)	(.020)	(.034)	(.039)	(.036)	(.035)	(.036)
	[.997]	[.982]	[.986]	[.926]	[.938]	[.246]	[.969]	[.969]	[.969]
3+students	-.026	.002	-.054*	-.016	-.133***	-.053	-.081**	-.144***	-.105**
	(.030)	(.039)	(.030)	(.029)	(.036)	(.054)	(.039)	(.037)	(.047)
	[.926]	[.997]	[.408]	[.982]	[.002]	[.880]	[.246]	[.001]	[.181]
p-value 1 student = 2 students = 3+ students	.172	.703	.005	.679	.008	.230	.235	.000	.186

Notes: Panel A reports estimates from regressions of composite executive function scores, executive function components and non-cognitive outcomes on an indicator for being randomly assigned to a classroom with a poorly-behaved student for various lags of year of assignment to a classroom with a poorly-behaved student. Data on executive function is only available up to grade 4, thus the executive function regressions pool information across grades 3-4. Data on non-cognitive outcomes are only available at the end of grade 6. In each regression, we regress the outcome on an indicator variable for being assigned to a classroom with a poorly-behaved student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a poorly-behaved student, child age and gender, and school (by grade when pooling) fixed effects. Panel B reports estimates from regressions of composite executive function scores, executive function components and non-cognitive outcomes on indicators for the number of poorly-behaved students in the classroom. In each regression, we regress the outcome on an indicator variable for being assigned to a classroom with different numbers of poorly-behaved students (omitted category is 0), controlling for a fourth-order polynomial in lagged achievement, an indicator for a poorly-behaved student, child age and gender, and school (by grade when pooling) fixed effects. Standard errors are clustered at the classroom and student level when pooling data, and at the classroom level in the regressions by grade. Below the standard errors in square brackets we report p-values computed according to the Romano and Wolf stepdown procedure (see Romano and Wolf, 2005 and Clarke, Romano and Wolf, 2020), using 5000 bootstrap replications.

Table 6: Dynamic effects of poorly-behaved students on student achievement

Panel A	(1)	(2)
Has poorly behaved student in *g*	-.020***	-.017**
	(.005)	(.009)
Has poorly behaved student in *g*-1		-.019**
		(.008)
Has poorly behaved student in *g* * Has poorly behaved student in *g*-1		.010
		(.011)
Has poorly behaved student in *g* * Lagged ability	.007	
	(.004)	

Panel B	Once	Twice	Three times	Four times
Has poorly behaved student	-.024	-.041**	-.051**	-.028
	(.020)	(.020)	(.022)	(.025)

Notes: This table reports estimates of dynamic effects of poorly-behaved students on their peer's math and language achievement. Panel A, column 1 reports estimates of a regression of achievement at the end of grade g on an indicator for having a poorly-behaved student in the classroom in grade g and in grade g-1, and an interaction between the two indicators, pooling information across grades. The regression controls for a fourth-order polynomial in twice lagged achievement, an indicator for being a poorly-behaved student in g, an indicator for being a poorly-behaved student in g-1, child age and gender, as well as school-by-grade fixed effects. Standard errors are clustered at the student and classroom level, using the sequence of classrooms in grades g-1 and g. Panel A, column 3 reports estimates of a regression of achievement at the end of grade g on an indicator for having a poorly-behaved student in the classrom in grade g, and an interaction between the indicator and lagged ability. The regression controls for a fourth-order polynomial in lagged achievement, an indicator for being a poorly-behaved student in g, child age and gender, and school-by-grade fixed effects. Standard errors are clustered at the student and classroom level. Panel B shows cumulative effects of being assigned to a classroom with a poorly-behaved student once, twice, three times or four times over time, on math and language achievement at the end of 6th grade. The regression controls for a fourth-order polynomial in achievement at the end of second grade, an indicator for ever being a poorly-behaved student between grades 3 and 6, child age and gender. Standard errors are clustered at the classroom level.

Table 7: Poorly-behaved students and attrition

Panel A	Attritor from *g* to *g*+1	
	(1)	(2)
Has poorly-behaved student	.001	
	(.003)	
Is poorly-behaved student		.024**
		(.010)
Panel B	Bottom 5 worst behaved	Poorly-behaved
	(1)	(2)
New entrant in *g*	.007**	
	(.003)	
New entrant in *g*-3		.006**
		(.003)

Notes: Panel A, column 1 shows results from a regression of an indicator variable for being an attritor between any grades g and g+1 on an indicator for having a poorly-behaved student in the classroom in g, pooling information across grades. The regression controls for a fourth-order polynomial in lagged ability, an indicator for being a poorly-behaved student, child age and gender, and school-by-grade fixed effects. Column 2 shows results from a regression of an indicator variable for being an attritor between any grades g and g+1 on an indicator for being a poorly-behaved student in g, pooling information across grades. The regression controls for a fourth-order polynomial in lagged ability, child age and gender, as well as school-by-grade fixed effects. Panel B, column 1 reports estimates from a regression of an indicator for being among the 5 worst behaved students in the classroom on an indicator for being a new entrant in any given grade, pooling information across grades 1-6. We regress the outcome on an indicator for being a new entrant in that grade, child age and gender, and school-by-grade fixed effects. Column 2 reports estimates from a regression of an indicator for being poorly-behaved in a given grade g according to our main definition on an indicator for being a new entrant in grade g-3, pooling information across grades 4-6. Standard errors are clustered at the classroom and student level throughout.

Appendix A

An important assumption underlying our empirical strategy is that poorly-behaved students are not purposefully matched to classrooms, due to random assignment of children to classrooms within schools in every year.[24] Random assignment is closely monitored, and compliance was very high, 98.9 percent on average. In this appendix, we present tests of random assignment using a methodology developed in Jochmans (2023).

First, we briefly discuss the procedure outlined in Jochmans (2023). Consider our setting, in which we observe data on S schools, and each school has $n_1, \dots, n_s$ students. Within each school, children are assigned to a classroom—and therefore their peer group—every year. Let $x_{s,i}$ be an observable characteristic of child i in school s. If assignment to peer groups is random, $x_{s,i}$ will be uncorrelated with $x_{s,j}$, for all j belonging to the set of $i's$ classroom peers. Let $\bar{x}_{s,j}$ be the average value of characteristic x among student $i's$ peers. The procedure tests whether the correlation in a within-school regression of $x_{s,i}$ on $\bar{x}_{s,i}$ is statistically significantly different from zero (a methodology first proposed in Sacerdote (2001)), introducing a bias correction for the inclusion of group fixed effects (in our case, schools). It is important to control for school fixed effects, as randomization happens within schools, but there may be selection into a school based on individual characteristics. Jochmans (2023) shows that a fixed-effects regression of $x_{s,i}$ on $\bar{x}_{s,i}$ will yield biased estimates due to inconsistency of the within-group estimator. The proposed corrected estimator is given by

$$ts = \frac{\sum_{s=1}^{S} \sum_{i=1}^{n_s} \tilde{x}_{s,i} \left(\bar{x}_{s,j} + \frac{x_{s,i}}{n_s - 1} \right)}{\sqrt{\sum_{s=1}^{S} \left(\sum_{i=1}^{n_s} \tilde{x}_{s,i} \left(\bar{x}_{s,j} + \frac{x_{s,i}}{n_s - 1} \right) \right)^2}} \qquad (A.1)$$

where $\tilde{x}_{s,i}$ is the deviation of $x_{s,i}$ from its within-school mean. The null hypothesis is thus absence of correlation between $i's$ characteristics and those of her peers. To test the random assignment in our setting, we implement this procedure by testing for the presence of correlation between child $i's$ scores measured at the end of grade $g-1$ and the average end-of-grade scores in $g-1$ of the classroom peers assigned to

[24] We use the word "random" as shorthand but, as discussed at length in Araujo et al. (2016), strictly speaking random assignment only occurred in 3rd through 6th grade. In the other grades, the assignment rules were as-good-as-random. Specifically, the assignment rules we implemented were as follows: In kindergarten, all children in each school were ordered by their last name and first name, and were then assigned to teachers in alternating order; in 1st grade, they were ordered by their date of birth, from oldest to youngest, and were then assigned to teachers in alternating order; in 2nd grade, they were divided by gender, ordered by their first name and last name, and then assigned in alternating order; in 3rd through 6th grades, they were divided by gender and then randomly assigned to one or another classroom.

her in a given grade g. We do so for each grade, for math and language achievement as well as executive function. The results are shown in tables A1, A2 and A3 . Note that, to check random assignment in kindergarten, we use TVIP scores collected at baseline. Our results show that we cannot reject the null hypothesis that there is no correlation between child $i's$ achievement and that of her classroom peers. Hence, we conclude that random assignment was successful in our setting.

Table A1: Testing for random assignment of children to classrooms, math

	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Test statistic	1.36	-.550	1.04	.104	-.749	.304	.720
P-value	.174	.583	.299	.917	.454	.761	.471

Notes: In this table, we report results for tests of random assignment of children to classrooms within schools using a methodology proposed by Jochmans (2023). The null hypothesis is absence of correlation between a child's math ability measured at the end of the previous grade and the average math ability of classroom peers assigned to her at the beginning of a given grade, conditional on school. To check random assignment in kindergarten, we use TVIP scores collected at baseline.

Table A2: Testing for random assignment of children to classrooms, language

	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Test statistic	1.36	-2.89	-.674	.231	-.383	-.780	-.084
P-value	.174	.004	.501	.818	.702	.435	.933

Notes: In this table, we report results for tests of random assignment of children to classrooms within schools using a methodology proposed by Jochmans (2023). The null hypothesis is absence of correlation between a child's language ability measured at the end of the previous grade and the average language ability of classroom peers assigned to her at the beginning of a given grade, conditional on school. To check random assignment in kindergarten, we use TVIP scores collected at baseline.

Table A3: Testing for random assignment of children to classrooms, EF

	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4
Test statistic	1.36	.161	-.083	-.988	-1.04
P-value	.174	.872	.934	.323	.299

Notes: In this table, we report results for tests of random assignment of children to classrooms within schools using a methodology proposed by Jochmans (2023). The null hypothesis is absence of correlation between a child's executive function score measured at the end of the previous grade and the average executive function score of classroom peers assigned to her at the beginning of a given grade, conditional on school. To check random assignment in kindergarten, we use TVIP scores collected at baseline. We only collected data on executive function up to fourth grade.

Appendix B

This appendix presents additional information on test scores, executive function, non-cognitive skills and disruptive behaviors. Figure B1 presents the univariate densities of our achievement measures, separately by grade. The figure shows that most of the distributions appear to have a reasonable spread and are generally symmetric. One clear exception is math achievement in kindergarten, which is left-censored.

Figure B2 presents comparable densities for executive function. It shows that the distributions of inhibitory control and cognitive flexibility are often highly skewed. This is not surprising given the nature of the tests. As an example, we describe the executive function tests we applied in kindergarten.

In the inhibitory control test, kindergarten children were quickly shown a series of 14 flash cards that had either a sun or a moon and were asked to say the word "day" when they saw the moon and "night" when they saw the sun. Just over half (50.8 percent) of all children made no mistake on this test, so there is a concentration of mass at the highest value, while very few children (1.6 percent) answered all prompts incorrectly.

The cognitive flexibility test we applied in kindergarten worked as follows. Children were handed a series of picture cards, one by one. Cards had either a truck or a star, in red or blue. The enumerator asked the child to sort cards by *color*, or by *shape*. Specifically, in the first half of the test, the enumerator asked the child to play the "colors" game, handed her cards, indicating their color, and asked the child to place them in the correct pile ("this is a red card: where does it go?"). After 10 cards, the enumerator told the child that they would switch to the "shapes" game, and reminded the child that, in this game, trucks should be placed in one pile and stars in another. The enumerator then handed the child cards, indicating the shapes on the card, and asked her to place them in the correct pile ("this is a star: where does it go?"). In both the first and the second part of the test, if the child made three consecutive mistakes, the enumerator paused the test, reminded her what game they were playing ("remember we are playing the shapes game; in the shapes game, all trucks go in this pile, and all stars in this other pile"), and handed the child a new card with the corresponding instruction. A small proportion of children in kindergarten (7.5 percent) did not understand the game, despite repeated examples, and were given a score of 0; just under half of all children (47 percent) answered all prompts correctly in both the "colors" and "shapes" parts of the test; and just over a quarter (27.3 percent) of all children made no mistakes in the first part of the test (the "colors" game), but incorrectly classified every card in the second part of the test (the "shapes" game). These children were unable to switch rules, despite repeated promptings from the enumerator. The distribution of scores for this test therefore has a concentration of mass at two points, with much less mass at other points.

The working memory test had two parts. In the first part, children were given 2 minutes to find as many sequences of dog, house, and ball, in that order, on a sheet that has rows of dogs, houses, and balls in various possible sequences. The score on this part of the test is the number of correct sequences found by the child. In the second part of the test, the enumerator recited strings of numbers, and asked the child to repeat them, in the same order or backwards. Figure B2 shows that the aggregate working memory score is distributed smoothly, with little evidence of a concentration of mass at particular values.

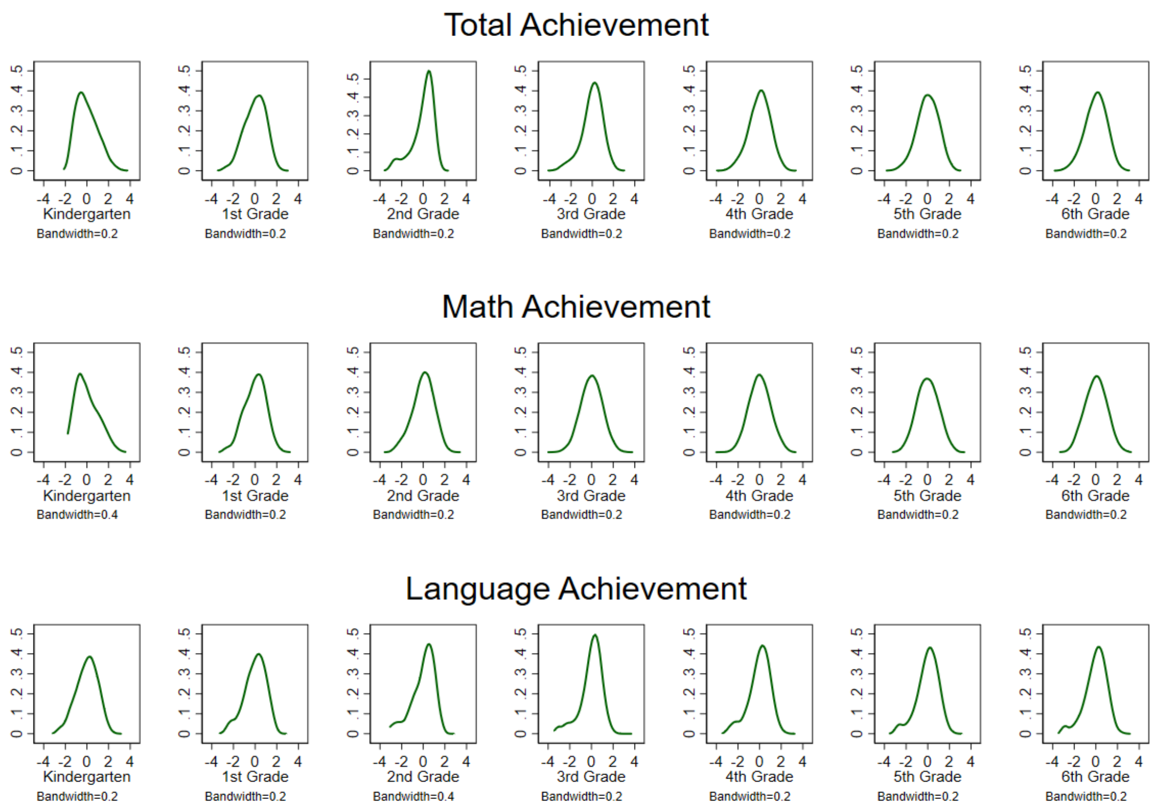
In practice the correlations of the scores across the three dimensions in our sample are low—in the range of 0.21 to 0.32 between cognitive flexibility and working memory, between 0.17 and 0.33 between working memory and inhibitory control, and in the range of 0.12 to 0.15 between cognitive flexibility and inhibitory control—see Appendix Table B1.[25] When the scores across the three dimensions are averaged, the distributions of the total executive function score are generally smooth and symmetric.

Figure B3, shows univariate densities of the four non-cognitive measures we applied in 6th grade. The figure shows that the distribution of the depression and grit scores appear to be right-censored. The distribution for the aggregate measure of non-cognitive outcomes, on the other hand, is smooth and symmetric. Table B2 shows that the different non-cognitive outcomes are positively correlated, although the correlations are far from unity—they range from 0.20 (between depression and grit) to 0.49 (between growth mindset and self-esteem). Figure B4 shows the average number of disruptive behaviors done always by poorly-behaved students according to our main definition. At the end of every grade, we ask teachers whether the worst 5 behaved students in any given grade t engaged in the following disruptive behaviors "never", "sometimes", "often" or "always" in that grade: the student is easily annoyed/frustrated; when spoken to, the student answered with bad manners; the student was disobedient and didn't respect classroom rules; the student behaved badly intentionally, for example by harassing other children; the student blamed other children for their own mistakes; the student was rancorous towards other children; the student quarreled with other children; the student hit/kicked/bit other children; the student intentionally broke toys or other objects; the student frequently interrupted the class; the student was extremely restless. For each group of poorly-behaved students according to our main definition, we report

[25] The fact that these correlations are very low is likely to be a result of both measurement error and differences across the constructs that each domain measures.

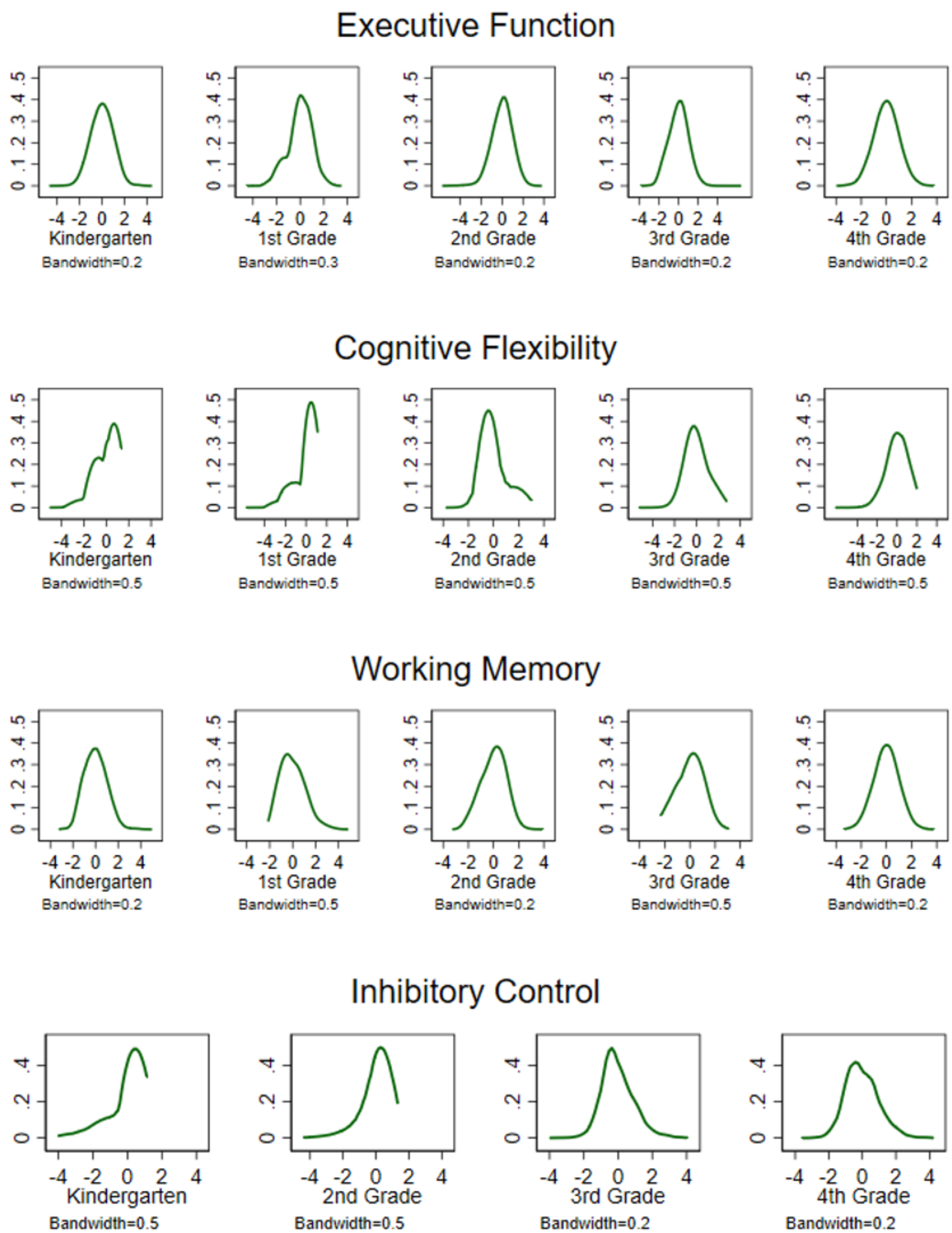
the average number of disruptive behaviors done always, as well as confidence intervals at the 95 percent level.

Figure B1: Distributions of achievement, by grade



Notes: The figure shows univariate densities of achievement, in z-scores, by grade.

Figure B2: Distributions of executive function, by grade



Notes: The figure shows univariate densities of executive function, in z-scores, by grade.

Figure B3: Distributions of non-cognitive outcomes

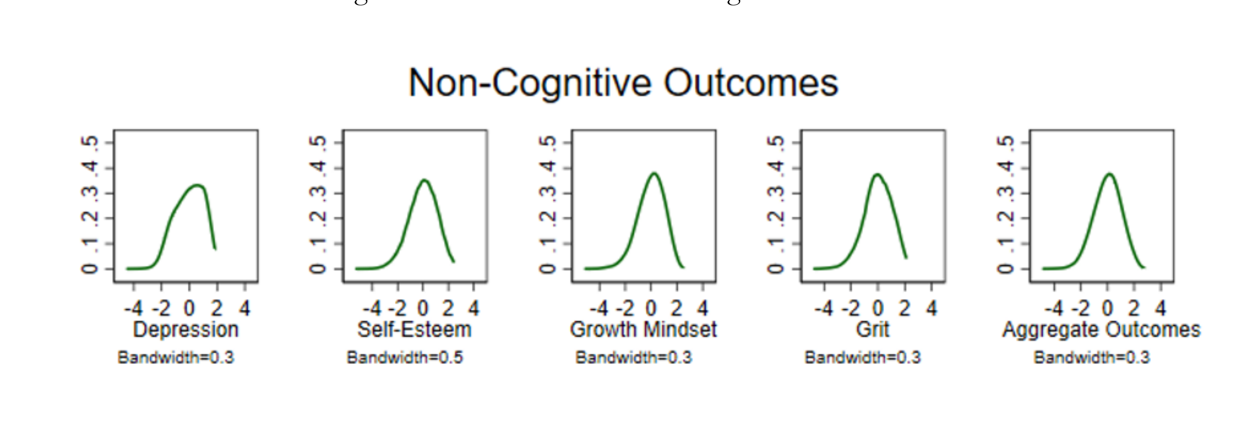
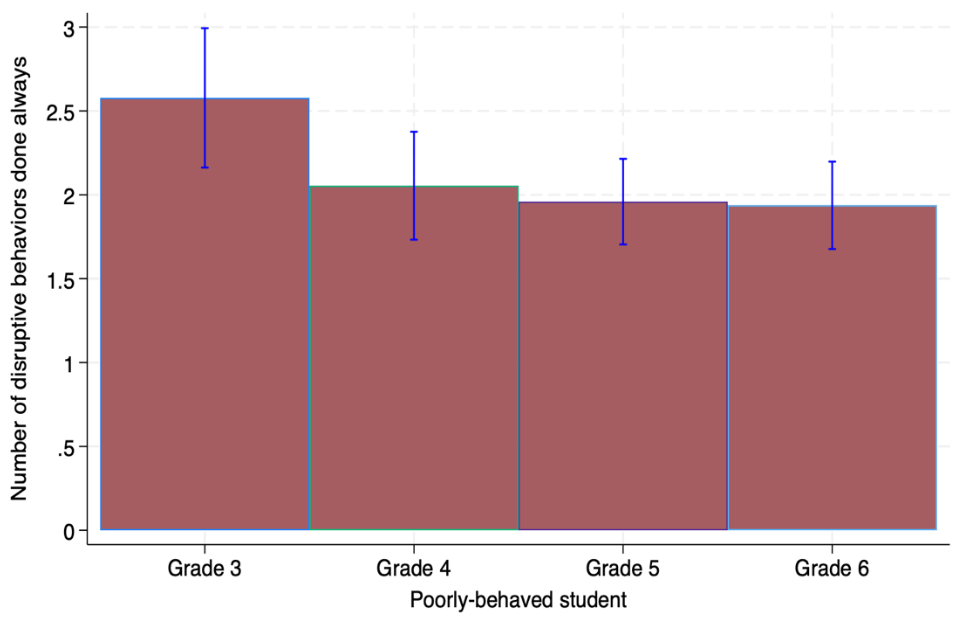


Figure B4: Behaviors of poorly-behaved students



Notes: This figure shows the average number of disruptive behaviors shown always by poorly-behaved students. We ask teachers whether the worst 5 behaved students in any given grade g engaged in the following behaviors "never", "sometimes", "often" or "always" that grade: the student is easily annoyed/frustrated; when spoken to, the student answered with bad manners; the student was disobedient and didn't respect classroom rules; the student behaved badly intentionally, for example by harassing other children; the student blamed other children for their own mistakes; the student was rancorous towards other children; the student quarreled with other children; the student hit/kicked/bit other children; the student intentionally broke toys or other objects; the student frequently interrupted the class; the student was extremely restless. For each group of poorly-behaved students according to our main definition, we report the average number of disruptive behaviors shown always in each grade. Confidence intervals at the 95 percent level.

Table B1: Correlations across dimensions in executive function

	Inhibitory Control	Cognitive Flexibility
	Kindergarten	
Cognitive Flexibility	0.13	
Working Memory	0.22	0.29
	1st Grade	
Working Memory		0.23
	2nd Grade	
Cognitive Flexibility	0.15	
Working Memory	0.25	0.24
	3rd Grade	
Cognitive Flexibility	0.12	
Working Memory	0.17	0.21
	4th Grade	
Cognitive Flexibility	0.15	
Working Memory	0.33	0.32
	Pooled	
Cognitive Flexibility	0.14	
Working Memory	0.24	0.26

Notes: The table reports the pairwise correlations between executive function dimensions. All the correlations are significant at the 1 percent level.

Table B2: Correlations across non-cognitive outcomes

	Depression	Self- Esteem	Growth Mindset
Self- Esteem	0.24		
Growth Mindset	0.26	0.49	
Grit	0.20	0.45	0.38

Notes: Table presents the results from pairwise correlations between non-cognitive outcomes collected in 6th grade. All the correlations are significant at the 1 percent level.

Appendix C

Table C1 presents results from a balancing exercise in which we test for the presence of correlation between pre-determined child characteristics and the presence of a poorly-behaved student in the classroom at the beginning of a given grade. These pre-determined characteristics are the student's math and language index measured at the end of the previous grade, child age and gender, and a factor score capturing family-level characteristics at baseline, constructed using factor analysis on variables measuring mother's education, father's education and household wealth.

Table C2 shows the distribution of low-achieving students across classrooms.

Table C1: Balance test

	Math+Language index	Age	Female	Family factor
Has poorly-behaved student	-.011	-.032	-.003	-.025
	(.010)	(.083)	(.002)	(.020)

Notes: This table presents estimated coefficients from a balancing exercise testing for the presence of correlation between pre-determined child characteristics and the presence of a poorly-behaved student in the classroom at the beginning of a given grade. Column 1 uses the math and language index measured at the end of the previous grade, column 2 uses child age at the beginning of the grade, column 3 uses child gender, column 4 uses a factor score capturing family-level characteristics at baseline (constructed using factor analysis on variables measuring mother's education, father's education and household wealth) as the outcome.

Table C2: Distribution of low-achieving students across classrooms

	(1) Nr low-achieving students	(2) Nr classrooms with low-achieving student	(3) Total nr classrooms	(4) Proportion classrooms with low-achieving student
3rd grade	319	194	470	41.3
4th grade	428	249	479	52.0
5th grade	554	306	485	63.1
6th grade	573	319	485	65.8

Note: The table shows descriptive statistics about the distribution of low-achieving students across classrooms in every grade. In any grade g between grades 3 and 6, a low-achieving student is a student who was ranked among the bottom 5 lowest achieving students in the classroom according to the teacher in all grades between g-1, g-2 and g-3. In columns 1-2 we show how many students comply with this definition in each grade between 3rd and 6th grade, and the number of classrooms in which there is a low-achieving student according to this definition. In columns 3-4 we show the total number of classrooms in the sample, as well as the proportion of classrooms with a low-achieving student.

Appendix D

Tables D1 to D6 present robustness checks of the effects of poorly-behaved students on their peers' math and language achievement to different definitions of "poorly-behaved" students. As discussed in the main body of the paper, our main specification defines a poorly behaved student in any grade *g* as a student who was reported to be among the 5 worst behaved students in the classroom by their teacher in all of the three previous grades. We can change the definition of what constitutes a poorly-behaved student along two dimensions: 1) the teacher rating cutoff, i.e., considering as disruptive those who are within the 5 worst-behaved children in the classroom, or take only the 4 worst, 3 worst, 2 worst, or the one worst-rated child in the classroom; 2) the number of years considered, i.e., considering as disruptive those ranked at the bottom of the classroom in terms of behavior for different numbers of consecutive years. Using higher (lower) rank cutoffs or more (less) consecutive years to define poorly-behaved students makes our definition more (less) stringent. A more stringent definition implies fewer poorly-behaved students, and less variation in the amount of classrooms exposed to a poorly-behaved student. Tables D1, D2 and D3 show estimated effects of poorly-behaved students on their peers' academic achievement, using information on teacher rankings of student behavior in all of the two previous grades, all of the three previous grades, and all of the four previous grades, respectively. In addition to changing the teacher rating cutoff or the number of years considered to define a poorly-behaved student, we can also change the set of grades we consider to define a poorly-behaved student. While in our main specification and appendix tables D1, D2 and D3 we use information from previous grades to define a poorly-behaved student in any given grade, we can also define poorly-behaved students by using information on student behavior collected in the first two, first three, or first four grades of elementary school to evaluate the impact of poorly-behaved students on their peers' achievement in later grades. Tables D4, D5 and D6 show estimated effects of poorly-behaved students on their peers' academic achievement using information on teacher rankings of student behavior in the first two grades, in the first three grades, and in the first four grades of elementary school, respectively.

In each table, we report results for different teacher ratings used to define a poorly-behaved student, as well as the total number of poorly-behaved students according to each definition, and the percent of classrooms with a poorly-behaved student according to each definition.

Table D1: Effect of poorly-behaved students on student achievement, using grades t-1 and t-2 to define poorly-behaved students

	Rank 1					
	Pooled	2014	2015	2016	2017	2018
Has poorly-behaved student	-.003	.002	.002	-.010	.003	-.011
	(.005)	(.017)	(.014)	(.015)	(.011)	(.010)
Number of poorly-behaved students		87	97	94	97	127
Percent classrooms with poorly behaved students		18.06	17.09	16.81	17.94	21.86
	Rank 2 or worse					
	Pooled	2014	2015	2016	2017	2018
Has poorly-behaved student	-.003	-.007	.003	.005	.004	-.014*
	(.004)	(.012)	(.012)	(.011)	(.009)	(.008)
Number of poorly-behaved students		206	254	244	254	316
Percent classrooms with poorly behaved students		35.91	39.96	39.29	40.82	48.66
	Rank 3 or worse					
	Pooled	2014	2015	2016	2017	2018
Has poorly-behaved student	-.006	-.014	-.002	.001	.000	-.013
	(.005)	(.012)	(.011)	(.012)	(.010)	(.009)
Number of poorly-behaved students		326	402	414	457	556
Percent classrooms with poorly behaved students		51.61	53.63	57.35	59.59	69.69
	Rank 4 or worse					
	Pooled	2014	2015	2016	2017	2018
Has poorly-behaved student	-.014**	-.020	-.008	-.022	-.008	-.012
	(.006)	(.013)	(.012)	(.014)	(.011)	(.013)
Number of poorly-behaved students		437	537	591	645	797
Percent classrooms with poorly behaved students		61.51	64.74	69.75	71.75	84.74
	Rank 5 or worse					
	Pooled	2014	2015	2016	2017	2018
Has poorly-behaved student	-.011*	-.021	.001	-.019	-.014	.009
	(.006)	(.015)	(.012)	(.014)	(.012)	(.016)
Number of poorly-behaved students		533	658	771	839	1050
Percent classrooms with poorly behaved students		66.45	72.22	78.57	79.38	91.13

Notes: This table reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a poorly-behaved student for various definitions of a poorly-behaved students. We classify a child as poorly-behaved if teachers reported them to be among the worst-behaved children in their classroom in (all of) grades t-1 and t-2 of elementary school. Each panel uses a different rank cutoff to define a poorly behaved student. In each regression, we regress the math and language scores index on an indicator variable for being assigned to a classroom with a poorly-behaved student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a poorly-behaved student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. Standard errors are clustered at the classroom level in the grade-by-grade regressions, and at the classroom and student level in the pooled regressions.

Table D2: Effect of poorly-behaved students on student achievement, using grades t-1, t-2 and t-3 to define poorly-behaved students

	Rank 1				
	Pooled	2015	2016	2017	2018
Has poorly-behaved student	-.027***	-.021	-.034	-.019	-.032*
	(.010)	(.021)	(.026)	(.018)	(.018)
Number of poorly-behaved students		33	32	34	34
Percent classrooms with poorly behaved students		6.00	6.11	6.63	6.60

	Rank 2 or worse				
	Pooled	2015	2016	2017	2018
Has poorly-behaved student	-.011*	-.002	-.015	-.001	-.021**
	(.006)	(.015)	(.016)	(.011)	(.010)
Number of poorly-behaved students		89	102	111	125
Percent classrooms with poorly behaved students		15.63	17.68	19.88	22.89

	Rank 3 or worse				
	Pooled	2015	2016	2017	2018
Has poorly-behaved student	-.014***	-.016	-.022*	.001	-.020**
	(.005)	(.012)	(.013)	(.009)	(.009)
Number of poorly-behaved students		156	179	223	246
Percent classrooms with poorly behaved students		26.55	28.42	36.65	39.18

	Rank 4 or worse				
	Pooled	2015	2016	2017	2018
Has poorly-behaved student	-.022***	-.032***	-.026**	-.009	-.019**
	(.005)	(.011)	(.013)	(.010)	(.009)
Number of poorly-behaved students		231	256	325	369
Percent classrooms with poorly behaved students		36.19	37.05	46.79	51.75

	Rank 5 or worse				
	Pooled	2015	2016	2017	2018
Has poorly-behaved student	-.019***	-.032***	-.028**	-.007	-.011
	(.005)	(.011)	(.012)	(.010)	(.009)
Number of poorly-behaved students		299	338	439	490
Percent classrooms with poorly behaved students		43.90	46.95	56.94	61.65

Notes: This table reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a poorly-behaved student for various definitions of a poorly-behaved students. We classify a child as poorly-behaved if teachers reported them to be among the worst-behaved children in their classroom in (all of) grades t-1, t-2 and t-3 of elementary school. Each panel uses a different rank cutoff to define a poorly behaved student. In each regression, we regress the math and language scores index on an indicator variable for being assigned to a classroom with a poorly-behaved student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a poorly-behaved student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. Standard errors are clustered at the classroom level in the grade-by-grade regressions, and at the classroom and student level in the pooled regressions.

Table D3: Effect of poorly-behaved students on student achievement, using grades t-1, t-2, t-3 and t-4 to define poorly-behaved students

	Rank 1			
	Pooled	2016	2017	2018
Has poorly-behaved student	-.037*	-.066	-.032	-.011
	(.020)	(.044)	(.022)	(.032)
Number of poorly-behaved students		10	14	13
Percent classrooms with poorly behaved students		1.89	2.49	2.48

	Rank 2 or worse			
	Pooled	2016	2017	2018
Has poorly-behaved student	-.018	-.063**	.010	-.006
	(.011)	(.025)	(.016)	(.013)
Number of poorly-behaved students		40	51	60
Percent classrooms with poorly behaved students		7.37	8.30	11.59

	Rank 3 or worse			
	Pooled	2016	2017	2018
Has poorly-behaved student	-.017**	-.036**	.000	-.019*
	(.008)	(.018)	(.011)	(.010)
Number of poorly-behaved students		79	114	126
Percent classrooms with poorly behaved students		13.89	19.09	22.77

	Rank 4 or worse			
	Pooled	2016	2017	2018
Has poorly-behaved student	-.011*	-.006	-.017	-.011
	(.006)	(.014)	(.011)	(.009)
Number of poorly-behaved students		130	167	210
Percent classrooms with poorly behaved students		22.53	25.73	33.54

	Rank 5 or worse			
	Pooled	2016	2017	2018
Has poorly-behaved student	-.010*	-.009	-.010	-.011
	(.006)	(.014)	(.010)	(.009)
Number of poorly-behaved students		173	230	284
Percent classrooms with poorly behaved students		29.47	32.78	42.24

Notes: This table reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a poorly-behaved student for various definitions of a poorly-behaved students. We classify a child as poorly-behaved if teachers reported them to be among the worst-behaved children in their classroom in (all of) grades t-1, t-2, t-3 and t-4 of elementary school. Each panel uses a different rank cutoff to define a poorly behaved student. In each regression, we regress the math and language scores index on an indicator variable for being assigned to a classroom with a poorly-behaved student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a poorly-behaved student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. Standard errors are clustered at the classroom level in the grade-by-grade regressions, and at the classroom and student level in the pooled regressions.

Table D4: Effect of poorly-behaved students on student achievement, using first two grades to define poorly-behaved students

	Rank 1					
	Pooled	2014	2015	2016	2017	2018
Has poorly-behaved student	-.006	.002	-.015	-.018	-.003	.002
	(.007)	(.017)	(.015)	(.017)	(.012)	(.013)
Number of poorly-behaved students			87			
Percent classrooms with poorly behaved students		18.06	14.93	12.55	11.16	9.90
	Rank 2 or worse					
	Pooled	2014	2015	2016	2017	2018
Has poorly-behaved student	-.009*	-.007	-.006	-.024*	-.010	.000
	(.005)	(.012)	(.011)	(.013)	(.009)	(.010)
Number of poorly-behaved students			206			
Percent classrooms with poorly behaved students		35.91	31.56	28.66	26.03	24.95
	Rank 3 or worse					
	Pooled	2014	2015	2016	2017	2018
Has poorly-behaved student	-.012**	-.014	-.022**	-.015	-.001	-.007
	(.005)	(.012)	(.010)	(.013)	(.009)	(.009)
Number of poorly-behaved students			326			
Percent classrooms with poorly behaved students		51.61	44.56	42.26	38.84	35.88
	Rank 4 or worse					
	Pooled	2014	2015	2016	2017	2018
Has poorly-behaved student	-.015***	-.020	-.033***	-.019	-.004	-.003
	(.005)	(.013)	(.010)	(.012)	(.009)	(.009)
Number of poorly-behaved students			437			
Percent classrooms with poorly behaved students		61.51	52.67	51.26	46.90	44.74
	Rank 5 or worse					
	Pooled	2014	2015	2016	2017	2018
Has poorly-behaved student	-.014***	-.021	-.030***	-.023*	-.007	.003
	(.005)	(.015)	(.011)	(.013)	(.009)	(.008)
Number of poorly-behaved students			533			
Percent classrooms with poorly behaved students		66.45	58.00	57.74	51.86	50.72

Notes: This table reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a poorly-behaved student for various definitions of a poorly-behaved students. We classify a child as poorly-behaved if teachers reported them to be among the worst-behaved children in their classroom in (all of) the first two grades of elementary school. Each panel uses a different rank cutoff to define a poorly behaved student. In each regression, we regress the math and language scores index on an indicator variable for being assigned to a classroom with a poorly-behaved student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a poorly-behaved student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. Standard errors are clustered at the classroom level in the grade-by-grade regressions, and at the classroom and student level in the pooled regressions.

Table D5: Effect of poorly-behaved students on student achievement, using first three grades to define poorly-behaved students

	Pooled	2015	2016	2017	2018
			Rank 1		
	Pooled	2015	2016	2017	2018
Has poorly-behaved student	-.018	-.021	-.038	.001	-.007
	(.012)	(.021)	(.026)	(.014)	(.021)
Number of poorly-behaved students			33		
Percent classrooms with poorly behaved students		6.00	4.63	4.14	3.73
			Rank 2 or worse		
	Pooled	2015	2016	2017	2018
Has poorly-behaved student	-.009	-.002	-.032*	-.007	.006
	(.007)	(.015)	(.018)	(.013)	(.013)
Number of poorly-behaved students			89		
Percent classrooms with poorly behaved students		15.63	13.89	12.01	10.97
			Rank 3 or worse		
	Pooled	2015	2016	2017	2018
Has poorly-behaved student	-.018***	-.016	-.025*	-.023**	-.006
	(.006)	(.012)	(.014)	(.010)	(.011)
Number of poorly-behaved students			156		
Percent classrooms with poorly behaved students		26.55	23.58	21.12	18.84
			Rank 4 or worse		
	Pooled	2015	2016	2017	2018
Has poorly-behaved student	-.015***	-.032***	-.014	-.021**	.008
	(.005)	(.011)	(.012)	(.010)	(.009)
Number of poorly-behaved students			231		
Percent classrooms with poorly behaved students		36.19	33.68	29.61	27.33
			Rank 5 or worse		
	Pooled	2015	2016	2017	2018
Has poorly-behaved student	-.014***	-.032***	-.026**	-.018**	.019**
	(.005)	(.011)	(.012)	(.009)	(.009)
Number of poorly-behaved students			299		
Percent classrooms with poorly behaved students		43.90	42.32	36.02	34.99

Notes: This table reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a poorly-behaved student for various definitions of a poorly-behaved students. We classify a child as poorly-behaved if teachers reported them to be among the worst-behaved children in their classroom in (all of) the first three grades of elementary school. Each panel uses a different rank cutoff to define a poorly behaved student. In each regression, we regress the math and language scores index on an indicator variable for being assigned to a classroom with a poorly-behaved student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a poorly-behaved student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. Standard errors are clustered at the classroom level in the grade-by-grade regressions, and at the classroom and student level in the pooled regressions.

Table D6: Effect of poorly-behaved students on student achievement, using first four grades to define poorly-behaved students

	Rank 1			
	Pooled	2016	2017	2018
Has poorly-behaved student	-.032	-.066	-.027	.008
	(.023)	(.044)	(.018)	(.031)
Number of poorly-behaved students		10		
Percent classrooms with poorly behaved students		1.89	1.66	1.45

	Rank 2 or worse			
	Pooled	2016	2017	2018
Has poorly-behaved student	-.011	-.063**	.014	.027*
	(.013)	(.025)	(.019)	(.017)
Number of poorly-behaved students		40		
Percent classrooms with poorly behaved students		7.37	6.00	5.59

	Rank 3 or worse			
	Pooled	2016	2017	2018
Has poorly-behaved student	-.019**	-.036**	-.011	-.006
	(.009)	(.018)	(.014)	(.014)
Number of poorly-behaved students		79		
Percent classrooms with poorly behaved students		13.89	12.01	10.35

	Rank 4 or worse			
	Pooled	2016	2017	2018
Has poorly-behaved student	-.006	-.006	-.016	.005
	(.007)	(.014)	(.012)	(.011)
Number of poorly-behaved students		130		
Percent classrooms with poorly behaved students		22.53	20.08	18.01

	Rank 5 or worse			
	Pooled	2016	2017	2018
Has poorly-behaved student	-.006	-.009	-.020*	.012
	(.006)	(.014)	(.011)	(.010)
Number of poorly-behaved students		173		
Percent classrooms with poorly behaved students		29.47	25.05	23.19

Notes: This table reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a poorly-behaved student for various definitions of a poorly-behaved students. We classify a child as poorly-behaved if teachers reported them to be among the worst-behaved children in their classroom in (all of) the first four grades of elementary school. Each panel uses a different rank cutoff to define a poorly behaved student. In each regression, we regress the math and language scores index on an indicator variable for being assigned to a classroom with a poorly-behaved student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a poorly-behaved student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. Standard errors are clustered at the classroom level in the grade-by-grade regressions, and at the classroom and student level in the pooled regressions.

Table D7: Effect of low achieving students on student achievement, using grades t-1 and t-2 to define low achieving students

	Pooled	2014	2015	2016	2017	2018
	Rank 1					
Has low achieving student	-.011*	-.017	-.013	.023	-.034***	-.016
	(.006)	(.017)	(.013)	(.017)	(.012)	(.010)
Number of low achieving students		52	81	74	71	92
Percent classrooms with low achieving students		9.25	14.04	12.94	13.20	17.32
	Rank 2 or worse					
	Pooled	2014	2015	2016	2017	2018
Has low achieving student	-.006	.000	-.009	.016	-.018**	-.014*
	(.005)	(.013)	(.011)	(.012)	(.008)	(.008)
Number of low achieving students		179	236	247	240	283
Percent classrooms with low achieving students		29.68	32.98	36.95	38.97	42.89
	Rank 3 or worse					
	Pooled	2014	2015	2016	2017	2018
Has low achieving student	-.010*	-.005	-.016	.009	-.019**	-.014
	(.005)	(.012)	(.011)	(.013)	(.009)	(.009)
Number of low achieving students		311	422	452	471	532
Percent classrooms with low achieving students		44.95	52.13	56.78	61.44	67.84
	Rank 4 or worse					
	Pooled	2014	2015	2016	2017	2018
Has low achieving student	-.006	.025*	-.021	.003	-.013	-.034***
	(.006)	(.013)	(.013)	(.014)	(.012)	(.012)
Number of low achieving students		467	634	681	733	810
Percent classrooms with low achieving students		58.28	65.74	68.68	75.88	84.12
	Rank 5 or worse					
	Pooled	2014	2015	2016	2017	2018
Has low achieving student	.002	.024	-.009	.011	-.002	-.029*
	(.007)	(.015)	(.015)	(.018)	(.014)	(.017)
Number of low achieving students		612	834	940	988	1107
Percent classrooms with low achieving students		67.96	73.62	80.38	81.86	91.55

Notes: This table reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a low achieving student for various definitions of a low achieving student. We classify a child as low achieving if teachers reported them to be among the lowest achieving children in their classroom in (all of) grades t-1 and t-2 of elementary school. Each panel uses a different rank cutoff to define a low achieving student. In each regression, we regress the math and language scores index on an indicator variable for being assigned to a classroom with a low achieving student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a low achieving student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. Standard errors are clustered at the classroom level in the grade-by-grade regressions, and at the classroom and student level in the pooled regressions.

Table D8: Effect of low achieving students on student achievement, using grades t-1, t-2 and t-3 to define low achieving students

	Rank 1				
	Pooled	2015	2016	2017	2018
Has low achieving student	-.007	-.025	.048	-.037*	-.022
	(.014)	(.020)	(.032)	(.023)	(.020)
Number of low achieving students		11	18	19	17
Percent classrooms with low achieving students		1.91	3.34	3.30	3.30
	Rank 2 or worse				
	Pooled	2015	2016	2017	2018
Has low achieving student	-.003	.006	.009	-.018	-.003
	(.007)	(.016)	(.015)	(.011)	(.011)
Number of low achieving students		56	78	96	101
Percent classrooms with low achieving students		9.15	12.94	17.53	17.32
	Rank 3 or worse				
	Pooled	2015	2016	2017	2018
Has low achieving student	-.012**	-.020	-.021	-.017*	.002
	(.005)	(.014)	(.013)	(.009)	(.008)
Number of low achieving students		125	172	204	244
Percent classrooms with low achieving students		19.15	26.1	32.78	38.35
	Rank 4 or worse				
	Pooled	2015	2016	2017	2018
Has low achieving student	-.009	-.007	-.013	-.010	-.005
	(.005)	(.012)	(.013)	(.009)	(.008)
Number of low achieving students		222	297	358	395
Percent classrooms with low achieving students		32.55	40.08	48.87	55.05
	Rank 5 or worse				
	Pooled	2015	2016	2017	2018
Has low achieving student	-.002	-.007	.003	.001	-.004
	(.006)	(.012)	(.014)	(.009)	(.009)
Number of low achieving students		319	428	554	573
Percent classrooms with low achieving students		41.28	51.98	63.09	65.77

Notes: This table reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a low achieving student for various definitions of a low achieving student. We classify a child as low achieving if teachers reported them to be among the lowest achieving children in their classroom in (all of) grades t-1, t-2 and t-3 of elementary school. Each panel uses a different rank cutoff to define a low achieving student. In each regression, we regress the math and language scores index on an indicator variable for being assigned to a classroom with a low achieving student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a low achieving student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. Standard errors are clustered at the classroom level in the grade-by-grade regressions, and at the classroom and student level in the pooled regressions.

Table D9: Effect of low achieving students on student achievement, using grades t-1, t-2, t-3 and t-4 to define low achieving students

	Rank 1			
	Pooled	2016	2017	2018
Has low achieving student	-.026	.070	-.062**	-.061**
	(.025)	(.057)	(.030)	(.026)
Number of low achieving students		4	4	7
Percent classrooms with low achieving students		0.84	0.62	1.44
	Rank 2 or worse			
	Pooled	2016	2017	2018
Has low achieving student	.017	.053*	.006	.008
	(.011)	(.028)	(.019)	(.012)
Number of low achieving students		22	32	49
Percent classrooms with low achieving students		3.97	5.77	8.45
	Rank 3 or worse			
	Pooled	2016	2017	2018
Has low achieving student	-.001	-.018	-.017	.025**
	(.008)	(.020)	(.012)	(.011)
Number of low achieving students		65	88	112
Percent classrooms with low achieving students		10.65	15.67	18.56
	Rank 4 or worse			
	Pooled	2016	2017	2018
Has low achieving student	.006	.014	-.010	.016*
	(.007)	(.015)	(.010)	(.009)
Number of low achieving students		122	173	207
Percent classrooms with low achieving students		18.37	28.25	32.99
	Rank 5 or worse			
	Pooled	2016	2017	2018
Has low achieving student	.008	.014	-.004	.015*
	(.006)	(.014)	(.009)	(.008)
Number of low achieving students		177	282	345
Percent classrooms with low achieving students		26.1	41.44	47.42

Notes: This table reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a low achieving student for various definitions of a low achieving student. We classify a child as low achieving if teachers reported them to be among the lowest achieving children in their classroom in (all of) grades t-1, t-2, t-3 and t-4 of elementary school. Each panel uses a different rank cutoff to define a low achieving student. In each regression, we regress the math and language scores index on an indicator variable for being assigned to a classroom with a low achieving student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a low achieving student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. Standard errors are clustered at the classroom level in the grade-by-grade regressions, and at the classroom and student level in the pooled regressions.

Table D10: Effect of low achieving students on student achievement, using first two grades to define low achieving students

	Rank 1					
	Pooled	2014	2015	2016	2017	2018
Has low achieving student	-.023**	-.017	-.043**	.017	-.055**	-.017
	(.009)	(.017)	(.020)	(.019)	(.023)	(.017)
Number of low achieving students	52					
Percent classrooms with low achieving students		9.25	6.60	6.26	5.77	4.74
	Rank 2 or worse					
	Pooled	2014	2015	2016	2017	2018
Has low achieving student	-.001	.000	.008	-.001	-.023**	.007
	(.006)	(.013)	(.013)	(.014)	(.012)	(.012)
Number of low achieving students	179					
Percent classrooms with low achieving students		29.68	21.06	20.88	18.76	16.49
	Rank 3 or worse					
	Pooled	2014	2015	2016	2017	2018
Has low achieving student	-.006	-.005	-.006	-.009	-.026**	.013
	(.005)	(.012)	(.012)	(.012)	(.010)	(.011)
Number of low achieving students	311					
Percent classrooms with low achieving students		44.95	35.53	32.15	30.52	27.42
	Rank 4 or worse					
	Pooled	2014	2015	2016	2017	2018
Has low achieving student	.004	.025*	-.001	-.003	-.025***	.022**
	(.005)	(.013)	(.011)	(.012)	(.010)	(.009)
Number of low achieving students	467					
Percent classrooms with low achieving students		58.28	47.87	42.80	43.51	37.73
	Rank 5 or worse					
	Pooled	2014	2015	2016	2017	2018
Has low achieving student	.003	.024	-.012	.002	-.020**	.016*
	(.005)	(.015)	(.011)	(.012)	(.010)	(.009)
Number of low achieving students	612					
Percent classrooms with low achieving students		67.96	59.15	52.40	51.96	46.60

Notes: This table reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a low achieving student for various definitions of a low achieving student. We classify a child as low achieving if teachers reported them to be among the lowest achieving children in their classroom in (all of) the first two grades of elementary school. Each panel uses a different rank cutoff to define a low achieving student. In each regression, we regress the math and language scores index on an indicator variable for being assigned to a classroom with a low achieving student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a low achieving student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. Standard errors are clustered at the classroom level in the grade-by-grade regressions, and at the classroom and student level in the pooled regressions.

Table D11: Effect of low achieving students on student achievement, using first three grades to define low achieving students

	Pooled	2015	2016	2017	2018
			Rank 1		
	Pooled	2015	2016	2017	2018
Has low achieving student	-.019	-.025	-.002	-.000	-.056***
	(.015)	(.020)	(.036)	(.021)	(.019)
Number of low achieving students			11		
Percent classrooms with low achieving students		1.91	1.88	1.44	1.24
			Rank 2 or worse		
	Pooled	2015	2016	2017	2018
Has low achieving student	-.000	.006	.014	.008	-.036**
	(.008)	(.016)	(.018)	(.016)	(.017)
Number of low achieving students			56		
Percent classrooms with low achieving students		9.15	8.77	7.63	5.98
			Rank 3 or worse		
	Pooled	2015	2016	2017	2018
Has low achieving student	-.004	-.020	-.008	.010	.009
	(.007)	(.014)	(.014)	(.014)	(.013)
Number of low achieving students			125		
Percent classrooms with low achieving students		19.15	16.70	15.46	13.81
			Rank 4 or worse		
	Pooled	2015	2016	2017	2018
Has low achieving student	.006	-.007	.006	.005	.020**
	(.006)	(.012)	(.012)	(.011)	(.010)
Number of low achieving students			222		
Percent classrooms with low achieving students		32.55	28.18	27.42	23.09
			Rank 5 or worse		
	Pooled	2015	2016	2017	2018
Has low achieving student	.004	-.007	.010	-.014	.027***
	(.006)	(.012)	(.012)	(.010)	(.009)
Number of low achieving students			319		
Percent classrooms with low achieving students		41.28	37.58	35.88	32.16

Notes: This table reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a low achieving student for various definitions of a low achieving student. We classify a child as low achieving if teachers reported them to be among the lowest achieving children in their classroom in (all of) the first three grades of elementary school. Each panel uses a different rank cutoff to define a low achieving student. In each regression, we regress the math and language scores index on an indicator variable for being assigned to a classroom with a low achieving student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a low achieving student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. Standard errors are clustered at the classroom level in the grade-by-grade regressions, and at the classroom and student level in the pooled regressions.

Table D12: Effect of low achieving students on student achievement, using first four grades to define low achieving students

	Pooled	2016	2017	2018
		Rank 1		
	Pooled	2016	2017	2018
Has low achieving student	.029	.070	.024	-.036
	(.032)	(.057)	(.040)	(.032)
Number of low achieving students		4		
Percent classrooms with low achieving students		0.84	0.62	0.41
		Rank 2 or worse		
	Pooled	2016	2017	2018
Has low achieving student	-.001	.053*	-.007	-.061***
	(.015)	(.028)	(.022)	(.021)
Number of low achieving students		22		
Percent classrooms with low achieving students		3.97	3.51	2.68
		Rank 3 or worse		
	Pooled	2016	2017	2018
Has low achieving student	-.017*	-.018	-.030*	-.002
	(.010)	(.020)	(.016)	(.014)
Number of low achieving students		65		
Percent classrooms with low achieving students		10.65	9.48	8.45
		Rank 4 or worse		
	Pooled	2016	2017	2018
Has low achieving student	.005	.014	-.011	.013
	(.008)	(.015)	(.013)	(.011)
Number of low achieving students		122		
Percent classrooms with low achieving students		18.37	17.53	14.85
		Rank 5 or worse		
	Pooled	2016	2017	2018
Has low achieving student	.009	.014	-.017	.031***
	(.007)	(.014)	(.011)	(.010)
Number of low achieving students		177		
Percent classrooms with low achieving students		26.10	24.54	20.41

Notes: This table reports estimates from regressions of an index of math and language scores on an indicator for being randomly assigned to a classroom with a low achieving student for various definitions of a low achieving student. We classify a child as low achieving if teachers reported them to be among the lowest achieving children in their classroom in (all of) the first four grades of elementary school. Each panel uses a different rank cutoff to define a low achieving student. In each regression, we regress the math and language scores index on an indicator variable for being assigned to a classroom with a low achieving student, controlling for a fourth-order polynomial in lagged achievement, an indicator for a low achieving student, child age and gender, and school (by grade, when pooling data across grades) fixed effects. Standard errors are clustered at the classroom level in the grade-by-grade regressions, and at the classroom and student level in the pooled regressions.

Appendix E

Table E1: Dynamic effects of poorly-behaved students on student achievement

	(1)	(2)	(3)
Has poorly behaved student in *g*	-.012*	-.014**	-.016**
	(.007)	(.007)	(.008)
Has poorly behaved student in *g*-1	-.014**	.000	-.003
	(.005)	(.005)	(.007)
Has poorly behaved student in *g* * Has poorly behaved student in *g*-1			.005
			(.010)
Controls for	Twice-lagged achievement	Once-lagged achievement and twice-lagged achievement	Once-lagged achievement and twice-lagged achievement

Notes: This table reports estimates of dynamic effects of poorly-behaved students on their peer's math and language achievement. Column 1 reports estimates of a regression of achievement at the end of grade g on an indicator for having a poorly-behaved student in the classroom in grade g and in grade g-1, pooling information across grades. The regression controls for a fourth-order polynomial in twice lagged achievement, an indicator for being a poorly-behaved student in g, an indicator for being a poorly-behaved student in g-1, child age and gender, as well as school-by-grade fixed effects. Standard errors are clustered at the student and classroom level, using the sequence of classrooms in grades g-1 and g. Column 2 reports estimates of a regression of achievement at the end of grade g on an indicator for having a poorly-behaved student in the classroom in grade g and in grade g-1 pooling information across grades. The regression controls for a fourth-order polynomial in lagged achievement, a fourth-order polynomial in twice lagged achievement, an indicator for being a poorly-behaved student in g, an indicator for being a poorly-behaved student in g-1, child age and gender, as well as school-by-grade fixed effects. Standard errors are clustered at the student and classroom level, using the sequence of classrooms in grades g-1 and g. Column 3 reports estimates of a regression of achievement at the end of grade g on an indicator for having a poorly-behaved student in the classroom in grade g and in grade g-1, and an interaction between the two indicators, pooling information across grades. The regression controls for a fourth-order polynomial in lagged achievement, a fourth-order polynomial in twice lagged achievement, an indicator for being a poorly-behaved student in g, an indicator for being a poorly-behaved student in g-1, child age and gender, as well as school-by-grade fixed effects. Standard errors are clustered at the student and classroom level, using the sequence of classrooms in grades g-1 and g.