



ROCKWOOL Foundation Berlin

Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES

268/26

Disruption in the Classroom: Experimental Evidence from Ecuador

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

Disruption in the Classroom: Experimental Evidence from Ecuador

Authors

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

Reference

JEL Codes: I24, I25

Keywords: Disruption; Peers; Random Classroom Assignment

Recommended Citation: Pedro Carneiro, Yyannú Cruz-Aguayo, Francesca Salvati, Norbert Schady (2026): Disruption in the Classroom: Experimental Evidence from Ecuador. RFBerlin Discussion Paper No. 268/26

Access

Papers can be downloaded free of charge from the RFBerlin website: <https://www.rfberlin.com/discussion-papers>

Discussion Papers of RFBerlin are indexed on RePEc: <https://ideas.repec.org/s/crm/wpaper.html>

Disclaimer

Opinions and views expressed in this paper are those of the author(s) and not those of RFBerlin. Research disseminated in this discussion paper series may include views on policy, but RFBerlin takes no institutional policy positions. RFBerlin is an independent research institute.

RFBerlin Discussion Papers often represent preliminary or incomplete work and have not been peer-reviewed. Citation and use of research disseminated in this series should take into account the provisional nature of the work. Discussion papers are shared to encourage feedback and foster academic discussion.

All materials were provided by the authors, who are responsible for proper attribution and rights clearance. While every effort has been made to ensure proper attribution and accuracy, should any issues arise regarding authorship, citation, or rights, please contact RFBerlin to request a correction.

These materials may not be used for the development or training of artificial intelligence systems.

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com



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

November 26th, 2025¹

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 classroom, the larger is their negative impact 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. There are no dynamic complementarities in the impact of poorly-behaved peers: impacts are similar for children who were and were not previously exposed to poorly-behaved peers (and for high and low ability children). We find indirect evidence that children with persistent behavioral difficulties are passed around schools.

¹ We thank participants in numerous seminars for their comments. We gratefully acknowledge the support of the InterAmerican Development Bank. Carneiro also gratefully acknowledges the support of the ESRC through the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS (ES/T014334/1), 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.

To assess how consequential poorly-behaved peers are in terms of their impact on student learning, we study how the presence of a poorly-behaved peer impacts classroom value added relative to other peer characteristics. This analysis, which proceeds in two steps, follows a similar idea to that in Feng and Jaravel 2020, although with a different execution to adapt it to our setting.

First, we estimate the causal impact of being assigned to a classroom with a poorly-behaved peer on classroom value added (as in Araujo et al. 2016) for grades 3 to 6, which are the focus of our paper. After correcting for sampling error (using Bayesian shrinkage, as is standard in the literature), the standard deviation of classroom value added in these grades is 0.07 to 0.10 standard deviations (SD) of

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.

learning (test scores are standardized to have mean zero and standard deviation one in every grade), so just slightly smaller than comparable estimates from the USA (e.g., Koedel and Rockoff, 2015).

Then, we examine to what extent different student characteristics predict classroom value added. Pooling across grades, classrooms with children with persistent problem behaviors experience a drop in value added of 0.02 SD, especially in grades 3 and 4. It is interesting that children who are persistently low achievers (who are not the same children as those who are disruptive) have a much smaller impact on their peers, and we cannot reject that it is equal to zero. We do not find evidence of any other peer characteristics predicting classroom value added.³ Our focus in the remainder in the paper is therefore on the impact of poorly-behaved peers, and we leave a deeper analysis of other peer characteristics for future work.⁴

We also 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. 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. Similarly, we cannot reject that the impacts of disruptive peers are similar among children with different levels of ability (partly because they benefited from different levels of prior investments) at the start of the academic year, which again suggests that the effects of poorly-behaved peers are additive.

The main empirical challenge to our work and to other papers attempting to explore random group assignment to study peer effects is that of a zero population first stage, in the instrumental variables analogy of Angrist (2014). We argue below that this is unlikely to be a problem in our setting. We show that in more than half of the schools in the sample, in a given grade, there exist classrooms

³ It is possible however that linear-in-means peer effects are only visible once one accounts for rank effects, as suggested in Carneiro et al (2024), Cornelissen et al (2024) and Bertoni and Nisticò (2023).

⁴ We leave the study of other peer variables for another paper. There are two main reasons to focus on poorly-behaved peers here: 1) together with being in a classroom with a persistently low achieving peer, being in a classroom with a poorly-behaved peer has the most within school (across classroom) variation of all peer variables (table F3); 2) the presence of a poorly-behaved peer in the classroom is uncorrelated with the majority of other observable peer characteristics; in the few cases where these correlations are significant, they are low (see table F4), which is not true of most other peer variables we do not study as closely. 1) means that we can study this peer variable confidently, while 2) means that we can attribute the impacts of this variable primarily to poorly-behaved peers and not other peer characteristics, especially if are willing to assume that a low of correlation with other observable peer variables also suggests low correlation with other unobservable peer characteristics (Altonji, Elder and Taber 2005; Oster 2019). In addition, disruption in the classroom is an important policy issue, as discussed above.

with and without poorly-behaved peers. Furthermore, the number of poorly-behaved peers varies across classrooms in more than two thirds of the schools. Given school size, number of classrooms (and the implied classroom size), and the probability that a child is a poorly-behaved peer, the number of poorly-behaved peers per classroom (in each school) follows a hypergeometric distribution. Cutting in half the number of schools where there are classrooms with and without disruptive peers would require approximately doubling average school size, while reducing this number by 75% would require increasing average school size between 3 and 4 times. In sum, we would need to have much larger school (and classroom) sizes in our sample to significantly reduce the amount of variation we can explore to identify impacts of poorly-behaved peers.

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.⁵ 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 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 primarily interested in the effects of both children with *persistent* behavioral problems or those with *persistent* 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,

⁵ 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.

however, that our point estimates are broadly similar across different definitions of who is a poorly-behaved child.⁶

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,⁷ 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; Neidell and Waldfogel 2010; Santavirta and Sarzosa 2023) is non-experimental, with one exception (Zhao and Zhao 2021), who focus on children with alcoholic fathers to identify disruptive peers.

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

⁶ 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.

⁷ 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).

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 Sarzosa 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 estimating classroom effects as in Araujo et al (2016), for grades 3 through 6. We then assess the extent to which poorly-behaved students, low-achieving students, and several other classroom peer characteristics affect learning outcomes. Pooling information across grades 3 through 6, having one or more poorly-behaved children in a classroom lowers classmates' achievement by (on average) .02 SDs,⁸ regardless of whether we use classroom averages or individual achievement as outcomes. This estimate is similar in specifications with and without other peer characteristics. We also document that we have substantial (within school – cross classroom) variation to estimate the impact of disruptive peers.

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,

⁸ 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.

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 -.037 SDs; and three or more such children reduce classmates’ achievement by .056 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.⁹ 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.¹⁰

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.¹¹ We also find that the 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. Although we cannot give a statistically precise answer to this question, we cannot reject that the effects are additive.

⁹ 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.

¹⁰ 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.

¹¹ 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).

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

A. Data

The data we use come from an experiment in 202 schools in Ecuador, a middle-income country in South America.¹² 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 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

¹² 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.

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.¹³

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 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

¹³ 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).

a more stringent definition along both dimensions. The definition we choose trades off stringency and statistical power. We show in Appendix E that our point estimates are robust to deviations from this definition.

Table 1 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.¹⁴

Table 2 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.¹⁵

¹⁴ 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 $g+1$.

¹⁵ 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).

Turning to the comparison between poorly-behaved and other students, Table 2 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.¹⁶ 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).

B. Classroom Variation in the Characteristics of Randomly Assigned Peers

Since children are randomly assigned to classrooms within schools, in every grade they have a different set of randomly assigned peers. This is the central idea of our research design. However, does this random group design produce strong variation in peer attributes, or are peers quite similar across classrooms in the same school? Angrist (2014) argues that random group designs such as the one used in this paper may be affected by a problem analogous to a weak first stage in an instrumental variables framework, because as classroom size grows we should observe no difference in the characteristics of peers in different classrooms (i.e., there is a zero population first stage). Although Angrist (2014) examines primarily linear-in-means peer models, a version of this problem applies also to count data, which is more relevant to our study of the impact of disruptive peers.

Fortunately, our experiment generates considerable variation in exposure to poorly-behaved (and low-achieving) students. This can be seen in Table 3, which shows the average number of poorly-behaved students per school in each grade (column 1); average classroom and school size (columns 2 and 3, respectively); and the proportion of schools where there are classrooms with and without disruptive students (column 4). Appendix table D2 shows similar statistics concerning low-achieving students (since we devote so much attention in the paper to this variable as well).

¹⁶ 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.

Over two thirds of schools have at least one poorly-behaved student per grade (Appendix Table C1), and several schools have more than one such student, so on average there are between 1 and 2 poorly-behaved students per school in each grade. There are on average 36 to 39 students per classroom, which is large by US and European elementary school standards, but not large enough to eliminate finite sample imbalances in the number of poorly-behaved peers, who are (by our definition) unusual students. Below we discuss how large classroom sizes would need to be to eliminate meaningful within-school variation in the presence of poorly-behaved peers across classrooms. In about 55% schools in our sample we can find classrooms both with and without poorly-behaved students (which is the variation we use to estimate how student achievement varies with the presence of poorly-behaved peers in the classroom), and in between 63% and 71% of the schools there is within school variation not only in whether there is a poorly-behaved student in the classroom, but also in the number of poorly-behaved students in each classroom (which we use to estimate how student achievement varies with the number of poorly-behaved peers in the classroom).¹⁷ In Appendix C, we provide further detail on the distribution of poorly-behaved students across classrooms within schools.

The presence or number of poorly-behaved peers is not the only characteristic that varies randomly across classrooms. There are a number of other classroom level imbalances that result from the random assignment of students to classrooms within schools. Furthermore, within-school imbalances in other dimensions of classroom composition, such as peer ability or family background, could be correlated with the presence of a poorly-behaved peer, in which case we would not be able to say whether the impacts we detect are due to the poorly-behaved peer or to any other peer characteristic correlated with the presence of poorly-behaved peers in the classroom.

Below we present a brief analysis of the impact of different peer characteristics on average learning in the classroom. Although several of these characteristics are correlated among each other (such as average peer ability in math, language and executive function, or average parental education), they are mostly uncorrelated with the presence of poorly-behaved peers in the classroom. This is shown in table F4 in the Appendix, where we present the within-school correlation across multiple classroom peer characteristics. As a result, our estimated impacts of disruptive peers are not confounded by impacts of other imbalanced peer characteristics, at least among the ones we observe. This claim is further

¹⁷ As shown in Appendix Table C1, a fraction of schools does not provide variation to estimate our main parameters of interest, but it is still helpful to aid estimation of the coefficients on the remaining controls, namely lagged test scores. Results are similar if we restrict our estimates to the sample of schools for which we have relevant within-school variation in exposure to disruptive peers, as shown in Appendix Table C2.

supported by our analysis in section 4 which shows that the impact of disruptive peers is robust to the inclusion of other peer characteristics in the regression models.

3. Empirical specification

A. Classroom Effects

We begin our analysis by assessing which classroom-level peer characteristics matter for student achievement. To this end, we first estimate classroom value-added (VA) in each grade, (as done in Araujo et al (2016) and most of the literature), then we investigate which observable peer characteristics are predictive of VA. This is analogous to the exercise in Feng and Jaravel (2020) who examine predictors of patent examiner effects, although our implementation is quite different. To compute classroom effects, we exclude poorly-behaved students from the sample.¹⁸

In particular, for each grade (between grades 3 and 6) we estimate:

$$Y_{i,c,g,s} = \varphi_g(Y_{i,c,g-1,s}) + \theta_g X_{i,c,g,s} + \omega_{c,g,s} + \varepsilon_{i,c,g,s} \quad (3.1)$$

where $Y_{i,c,g,s}$ is the achievement of student i , in classroom c , in school s , at the end of grade g . The function $\varphi(\cdot)$ is a fourth-order polynomial in lagged achievement, $X_{i,c,g,s}$ includes child age and gender, $\omega_{c,g,s}$ is a set of classroom fixed effects; and $\varepsilon_{i,c,g,s}$ is a residual. Since random assignment only occurs within school, we demean our estimates of classroom VA, $\omega_{c,g,s}$, within school (see also Chetty et al. 2011, and Jacob and Lefgren 2008). Given our research design, these estimates (denoted by $\hat{\omega}_{c,g,s}$ in the remainder of the paper) correspond to the causal impacts of classrooms on student learning. The (within-school) standard deviations of classroom VA for the math and language index in grades 3-6, after accounting for sampling error (using a standard shrinking procedure; see, e.g., Araujo et al. 2016), are displayed in table F1 in the appendix. They are similar to those reported in Araujo et al. (2016) for Kindergarten, and of comparable magnitude to those reported in the literature for the USA (e.g., Jackson et al. 2014).

B. Peer and non-Peer Correlates of Classroom Effects

Once we have estimates of classroom VA, $\hat{\omega}_{c,g,s}$, we can investigate which classroom peer characteristics (Z_{cgs}) are good predictors of VA. In particular, we estimate the following models (one observation per classroom):

¹⁸ We obtain similar results when including poorly-behaved students to compute classroom effects.

$$\hat{\omega}_{c,g,s} = \pi_{gs} + Z_{cgs}\tau_g + v_{cgs} \quad (3.2)$$

where π_{gs} are school by grade fixed effects and v_{cgs} is the residual. Z_{cgs} includes the presence of a poorly-behaved student, the presence of a low-achieving student, the mean lagged student achievement in math and language, the standard deviation of lagged achievement in math and language (or different quantiles of the distribution of this variable), the mean lagged achievement in executive function, the standard deviation of lagged achievement in executive function (or different quantiles of the distribution of this variable), the mean student age, and the proportion of female students. In addition, we also include the following variables, constructed for students who were present in the sample at baseline (i.e., at the start of kindergarten): the mean family wealth at baseline, the standard deviation of family wealth at baseline, the proportion of students whose mother's education is below secondary, the proportion of students whose father's education is below secondary, the mean vocabulary score at baseline, the standard deviation of the vocabulary score at baseline, and the proportion of students who attended preschool. We estimate equation (3.2) using one regressor at a time and entering all the regressors simultaneously in the model. This equation can be estimated pooling data from all grades (and assuming that $\tau_g = \tau$ for all grades), which are the main estimates presented in the paper, or grade by grade, which we show in Appendix Table F1. When pooling data across grades we cluster the standard errors at the school level, allowing the residuals to be serially correlated within schools (because of persistent school level shocks).

C. Main specification

For most of the paper our focus 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:

$$Y_{i,c,g,s} = \beta D_{c,g,s} + \varphi_g(Y_{i,c,g-1,s}) + \theta_g X_{i,c,g,s} + \delta_{g,s} + \varepsilon_{i,c,g,s} \quad (3.3)$$

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 $\varphi(\cdot)$ 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;¹⁹ $\delta_{g,s}$ 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.

D. 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 (3.3), 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 (3.3), but use a fourth-order polynomial in lagged executive function. These regressions use information in 3rd through 4th grades, where data on current and lagged executive function are available.

E. Dynamics

We now extend equation (3.3) 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 peer²⁰:

$$Y_{i,c,g,s} = \beta_g D_{c,g,s} + \sigma_g D_{i,c,g,s} Y_{i,c,g-1,s} + \varphi_g(Y_{i,c,g-1,s}) + \theta_g X_{i,c,g,s} + \delta_{g,s} + \varepsilon_{i,c,g,s} \quad (3.4)$$

¹⁹ 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.

²⁰ Only around 1% of students in our sample change schools between grades. As such, we only control for the current school (by grade) fixed effect in the dynamics regressions as it captures a student's school over time.

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:

$$Y_{i,c,g,s} = \beta_g D_{c,g,s} + \tau_g D_{i,c,g-1,s} + \sigma_g D_{i,c,g,s} D_{i,c,g-1,s} + \varphi_g(Y_{i,c,g-2,s}) + \theta_g X_{i,c,g,s} + \delta_{g,s} + \varepsilon_{i,c,g,s} \quad (3.5)$$

This allows the impact of peer exposure in the past to affect current learning over and beyond its impact on past learning.²¹

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:

$$Y_{i,c,6,s} = \sum_{k=1}^4 \lambda_k ND_{c,k,s} + \varphi_6(Y_{i,c,2,s}) + \theta_6 X_{i,c,6,s} + \delta_{6,s} + \varepsilon_{i,c,6,s} \quad (3.6)$$

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). In specifications (3.5) and (3.6), which include both current and lagged peer disruption, we avoid controlling for a student's own current status as poorly-behaved peer, as this could be potentially influenced by past exposure to disruptive peers.

4. Results

A. Peer Correlates of (causal) Classroom Value Added

We estimate the causal impact of classrooms on an index of student achievement in math and language (classroom VA) in grades 3 to 6 using the procedure described in section 3A. We denote the resulting estimate as $\hat{\omega}_{c,g,s}$. After accounting for sampling error (“shrinking”), the standard deviation of classroom

²¹ To assess the sensitivity of our findings to different specifications, we also present estimates of the following equation in Appendix G: $Y_{i,c,g,s} = \beta_g D_{c,g,s} + \tau_g D_{i,c,g-1,s} + \sigma_g D_{i,c,g,s} D_{i,c,g-1,s} + \varphi_g(Y_{i,c,g-1,s}) + \varphi_g(Y_{i,c,g-2,s}) + \theta_g X_{i,c,g,s} + \delta_{g,s} + \varepsilon_{i,c,g,s}$, which is analogous to (3.5), but in addition to $Y_{i,c,g-2,s}$ we also control for $Y_{i,c,g-1,s}$.

VA ($\omega_{c,g,s}$) is .076, .089, .057 and .055, in grades 3, 4, 5 and 6 respectively (as shown in Appendix Table F1). These are smaller but comparable to estimates of classroom VA from the USA.

We then use equation (3.2) to examine the extent to which several peer characteristics at the classroom level predict classroom VA, $\hat{\omega}_{c,g,s}$, after controlling for school (by grade) fixed effects. We estimate models that pool all grades together, clustering the standard errors at the school level to account for persistent school shocks. Grade by grade estimates are presented in Appendix Table F1.

The peer variables we consider are the presence of a poorly-behaved student, the presence of a low-achieving student, the mean lagged student achievement in math and language, the standard deviation of lagged achievement in math and language (or different quantiles of the distribution of this variable), the mean lagged achievement in EF, the standard deviation of lagged achievement in EF (or different quantiles of the distribution of this variable), the mean student age, and the proportion of female students. In addition, we also consider the following variables, constructed for students who were present in the sample at baseline (i.e., at the start of kindergarten): the mean family wealth at baseline, the standard deviation of family wealth at baseline, the proportion of students whose mother's education is below secondary, the proportion of students whose father's education is below secondary, the mean vocabulary score at baseline, the standard deviation of the vocabulary score at baseline, and the proportion of students who were in preschool.

Table 4 presents the results of the model pooling data from all grades. In the first column we include one variable at a time, while in the second we include all of them simultaneously. Coefficients and standard errors are shown, as well as p-values for the test that each coefficient is equal to zero after accounting for multiple hypothesis testing, using the procedure in Romano and Wolf (2005). When we look at the variable-by-variable results there is only one variable for which we can reject the null hypothesis that its effect on classroom VA is equal to zero (after accounting for multiple hypothesis testing): the presence of a poorly-behaved peer. When we include all regressors simultaneously some of the coefficients change, but the one measuring the impact of a poorly-behaved peer on classroom VA barely moves. This is because, as shown above, there is little correlation between the presence of a poorly-behaved peer and all the other observable peer attributes we include here.²² Column 2 also shows how much of the variation in classroom VA is explained by the dummy for the presence of a poorly-behaved peer, as captured by the noise-corrected R2 (corrected using Bayesian shrinkage, as is standard

²² We can no longer reject that the coefficient on disruptive peers is equal to zero in the model including all peer variables simultaneously. Nevertheless, this is a regression that includes many controls which affects the precision with which each coefficient can be estimated. We note however that the estimated coefficient and standard error on disruptive peers are virtually identical to those estimated in the individual regressions.

in the literature). The poorly-behaved peer presence dummy explains 1.1% of the variation in classroom VA, compared to 2.2% explained by all other observed peer characteristics. In the remainder of the paper our focus is on further understanding the impact of being in a classroom with a poorly-behaved peer.

B. The Impact of Poorly-Behaved Peers

We now turn to student rather than classroom level data, since several of the relationships we want to study (such as the long-lasting impacts of exposure to poorly-behaved peers, or the impacts of being exposed to poorly-behaved peers in different sequences of grades, or heterogeneous impacts of exposure to poorly-behaved peers) make more sense as student-level relationships. Our first set of results is in Table 5, where the outcome of interest is the average of math and language achievement in each grade, and where we examine the impact of being exposed to a poorly-behaved student both contemporaneously (without lags, as in equation (3.3) above) and at various lags. In the latter (dynamic) specifications exposure to a poorly-behaved classmate ($D_{c,g,s}$) and achievement ($Y_{i,c,g+n,s}$, where $n=1,2,3$) are measured in different periods. In Panel A, we provide results pooling data across grades 3 to 6. Hence, in the first column of the panel the coefficient on $D_{c,g,s}$ in equation (3.3) is restricted to be the same for all grades ($\beta_g = \beta$). In Panel B, we show results by grade, where this restriction is relaxed. Each coefficient in the table corresponds to a separate regression.

Starting with column 1 of Panel A, 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 .021 SDs. As shown in column 1 of Panel B, 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: .053, reported in the last row of this column).²³

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 -.034, -.018, -.013, and .002, respectively, and we can reject the null that these effects are the same (p-value: .020,

²³ 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.

reported in the last column 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.²⁴

In Panel C of Table 5, we report the number of schools in which only some classrooms within the school have a poorly behaved student, for grades 3 to 6, as well as the number of classrooms in such schools and the number of students enrolled in such schools. For the estimation of our main specification (3.3), these schools are those that provide identifying variation.

Table 6 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 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: .014).

Table 7 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 (3.3) 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, .037 and .056 SDs, respectively, and we can reject the null that these effects are equal to each other (p-value .000).

In Panel B of Table 7, we report the number of schools with differing numbers of poorly-behaved students across classrooms for grades 3 to 6, as well as the number of classrooms in such schools and the number of students enrolled in such schools. These schools contribute information for models which use variation in the number of poorly behaved students in the classroom.

In sum, Tables 5, 6 and 7 show 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

²⁴ 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.

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 E 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 5; (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 E1 to E6 show that our main results are robust to the definition of “poorly-behaved” students we use.²⁵

C. Executive function and non-cognitive skills

We turn to other outcomes in Table 8. 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 (3.3), while Panel B considers an extension of equation (3.3) 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.

²⁵ Similarly, appendix tables E7 to E12 show that our results on the effects of low achieving students are robust to the definition of “low achieving” students we use.

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 6th 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.

D. Dynamics

In Table 5 we documented two important results regarding the dynamics of skill formation: (i) the impacts of poorly-behaved peers are larger in 3rd and 4th grade than in 5th or 6th grade; and (ii) these impacts persist for two additional grades. The model on which these estimates are based (equation 3.3) 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 (3.4) and (3.5) are shown in Panel A of Table 9. 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 (3.3) (and our estimates in Table 5) is likely a good approximation to the process of skill accumulation in our dataset.²⁶

Estimates of (3.6) are shown in panel B of Table 9. 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.

E. Patterns of cross-school transfers

²⁶ In Appendix table E1 we present an additional specification where we include $Y_{i,c,g-1,s}$ in addition to $Y_{i,c,g-2,s}$.

In Table 10, 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.²⁷ 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 0.6 percentage points more likely to be reported as being persistently *poorly-behaved* than other children.

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).²⁸

²⁷ Hanushek et al. (2004) also find that children who move schools perform worse than other schools, and that movers reduce achievement in receiving classrooms.

²⁸ School assignment in Ecuador is based primarily on the distance between each student's residence and nearby schools, as in many other settings. A student is required to provide a utility bill (such as an electricity bill) that confirms the student's residence within the school's designated catchment area. Families can only select schools that are located within the zone determined by the address on the utility bill. Therefore, students cannot switch to another school at will unless it is in the same catchment area. Students in Ecuador are not relocated between schools due to behavioural issues. If disciplinary issues arise, they are typically managed by the Departamento de Consejería Estudiantil (DECE), the student counselling department. The DECE focuses on resolving issues through dialogue and agreements with parents, emphasizing support rather than expulsion.

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.²⁹

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. 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?³⁰ 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.

²⁹ 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.

³⁰ 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.
- Altonji, J. G., T. E. Elder, and C. R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113, no. 1: 151-184.
- Angrist, J. D. 2014. "The Perils of Peer Effects." *Labour Economics* 30: 98-108.
- 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. Nore. 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.
- Bertoni, M. and R. Nisticò. 2023. "Ordinal Rank and the Structure of Ability Peer Effects." *Journal of Public Economics*, 217, p.104797.
- 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. Intrigo, 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, N. Hilger, E. Saez, D.W. Schanzenbach and D. Yagan. 2011. “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR.” *The Quarterly journal of economics*, 126(4), pp.1593-1660.
- 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.
- Cornelissen, T., C. Dustmann and U. Schönberg. 2024. “Knowledge Spillover, Competition and Individual Careers.” Mimeo.
- 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 Classroom.” *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.
- Feng, J., and X. Jaravel. 2020. “Crafting Intellectual Property Rights: Implications for Patent Assertion Entities, Litigation, and Innovation.” *American Economic Journal: Applied Economics* 12(1): 140-181.
- 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.
- Jackson, C.K., J. Rockoff D. and Staiger. 2014. “Teacher Effects and Teacher-Related Policies.” *Annu. Rev. Econ.*, 6(1), pp.801-825.
- Jacob, B. and L. Lefgren. 2008. “Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education.” *Journal of Labor Economics*, 26(1), pp.101-136.
- Jacob, B., L. Lefgren, and D. P. Sims. 2010. “The Persistence of Teacher-Induced Learning.” *Journal of Human Resources* 45(4): 915-43.
- Koedel, C. and J.E. Rockoff, 2015. “Value-Added Modeling: A Review.” *Economics of Education Review*, 47, pp.180-195.
- 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* 116(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.
- Neidell, M., and J. Waldfogel. 2010. “Cognitive and Noncognitive Peer Effects in Early Education.” *The Review of Economics and Statistics* 92(3): 562–76.
- OECD. 2019. “TALOS 2018 Results: Teachers and School Leaders as Lifelong Learners”, Volume I, OECD.

- Oster, E., 2019. “Unobservable selection and coefficient stability: Theory and evidence.” *Journal of Business & Economic Statistics*, 37(2), pp.187-204.
- 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.

Tables

Table 1: 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 $g+3$
3 rd grade	.64	.64	.49	.41
4 th grade	.73	.77	.62	
5 th grade	.69	.72		
6 th grade	.66			

Note: The table shows descriptive statistics about the persistence in poor behavior. 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 $g-3$. 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 2: Characteristics of poorly-behaved students

	Poorly-behaved		Not poorly-behaved		Diff.	pvalue
	Mean	N	Mean	N		
Female	.047	829	.501	20,527	-.454	.000
Age at baseline (months)	68.0	547	67.5	10,584	.473	.014
Math	-.291	829	.002	18,976	-.293	.000
Language	-.432	829	.009	18,985	-.441	.000
Math+Language index	-.389	829	.005	18,976	-.394	.000
Executive function	-.250	829	.008	17,721	-.258	.000
Executive function: inhibitory control	-.122	829	-.002	17,721	-.121	.000
Executive function: memory and attention	-.238	829	.009	17,721	-.247	.000
Executive function: cognitive flexibility	-.107	829	.002	17,721	-.109	.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	.673	508	.674	9,905	-.001	.960
Father education less than secondary	.742	365	.693	7,885	.049	.045
Mother age	29.7	505	30.3	9,854	-.570	.049
Father age	34.5	351	34.6	7,715	-.087	.838
Wealth	-.040	542	.029	10,375	-.069	.105
Vocabulary at baseline (TVIP)	.000	526	.042	10,179	-.042	.334
Preschool	.717	548	.602	10,583	.115	.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. The sample only considers schools in which there are at least two classrooms.

Table 3: Distribution of poorly-behaved students

	(1) Average number of poorly-behaved students per school	(2) Average classroom size	(3) Average school size	(4) Proportion of schools in which only some classrooms have a poorly-behaved student
3 rd grade	1.39	37.0	89.7	.556
4 th grade	1.61	38.7	94.3	.529
5 th grade	2.06	36.8	88.3	.554
6 th grade	2.32	35.9	87.1	.536

Note: The table shows descriptive statistics about the distribution of poorly-behaved students across schools in every grade. 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 $g-3$. In column 1 we report the average number of poorly-behaved students within a school. Columns 2 and 3 report average classroom and school size. Column 4 reports the proportion of schools in which only some (but not all) classrooms have a poorly-behaved student. The total number of poorly-behaved students is 262 in grade 3, 304 in grade 4, 398 in grade 5, 450 in grade 6. The proportion of poorly-behaved students in the total sample is .015 in grade 3, .017 in grade 4, .023 in grade 5, .027 in grade 6. The sample only considers schools in which there are at least two classrooms.

Table 4: Effect of peer- and non-peer characteristics on classroom value added

	De-meaned classroom value added	
	Separate regressions	Joint regression
Has poorly-behaved student	-.022*** (.008) [.063]	-.021** (.009) [.128]
Has low-achieving student	-.009 (.010) [.946]	-.008 (.011) [.974]
Mean lagged math+language index	.006 (.028) [.998]	-.000 (.029) [.995]
SD lagged math+language index	-.004 (.035) [.998]	-.005 (.039) [.995]
Mean age	-.002 (.003) [.992]	-.002 (.003) [.985]
Proportion female	-.153 (.146) [.902]	-.065 (.149) [.989]
Mean wealth at baseline	-.014 (.021) [.983]	-.021 (.027) [.974]
SD wealth at baseline	.004 (.035) [.998]	.006 (.035) [.995]
Proportion mother education less than secondary	.013 (.038) [.997]	.014 (.043) [.995]
Proportion father education less than secondary	-.039 (.035) [.902]	-.050 (.036) [.766]
Mean vocabulary at baseline (TVIP)	-.000 (.021) [.998]	.024 (.026) [.949]
SD vocabulary at baseline (TVIP)	.034 (.025) [.795]	.054* (.031) [.445]
Proportion preschool	-.043 (.043) [.902]	-.044 (.042) [.899]
Observations	1,847	
SD of classroom value added	.087	
Corrected SD of classroom value added	.071	
P-value joint significance all peer characteristics		.435
P-value joint significance peer characteristics excluding poorly-behaved student dummy		.817
Noise-corrected R-squared		.033
Contribution of poorly-behaved student dummy to noise-corrected R-squared		.011

Notes: This table shows regressions of classroom value added on peer and non-peer classroom characteristics pooling across grades 3 to 6. To derive classroom value added and peer characteristics at the classroom level, we exclude poorly-behaved students from the sample. Column 1 reports estimated coefficients from regressions of classroom value added on each characteristic separately (hence, each coefficient corresponds to a separate regression). Column 2 reports estimated coefficients from a regression of classroom value added on all characteristics included jointly. All regressions control for school-by-grade fixed effects. Column 2 reports the noise-corrected R-squared of the regression to account for measurement error in classroom value added, as well as the proportion of the noise-corrected R-squared explained by the poorly-behaved student dummy alone. It also reports p-values for a test of joint significance of all peer characteristics, and a test of joint significance of all peer characteristics excluding the poorly-behaved dummy. In square brackets, we report p-values from the Romano and Wolf (2005) stepdown procedure based on 5000 bootstrap replications.

Table 5: Effect of poorly-behaved students on student achievement

Panel A					
	Lag				
	0	1	2	3	Joint test 1
Has poorly-behaved student	-.021*** (.005)	-.016*** (.005)	-.020*** (.007)	.002 (.011)	.091
Panel B					
	Lag				
	0	1	2	3	Joint test 2
Has poorly-behaved student 3rd grade	-.034*** (.011)	-.018* (.010)	-.013 (.011)	.002 (.011)	.020
Has poorly-behaved student 4th grade	-.030** (.013)	-.021** (.009)	-.027*** (.009)		.585
Has poorly-behaved student 5th grade	-.008 (.010)	-.011 (.008)			.714
Has poorly-behaved student 6th grade	-.012 (.009)				
Joint test 3	.258				
Joint test 4	.053				
Panel C					
	3rd grade	4th grade	5th grade	6th grade	
Number of schools in which only some (but not all) classrooms have a poorly-behaved student	105	100	107	104	
Number of classrooms in such schools	271	264	275	269	
Number of students in such schools	10,071	10,107	10,204	9,573	

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 a classroom with a poorly-behaved student (where lag 0 captures the contemporaneous effect), pooling information across grades. 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 fixed effects. Joint test 1 shows the p-value for a joint test of equality of estimated coefficients for different lags of exposure to a poorly-behaved student. Standard errors are clustered at the classroom and student level. 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 for various lags of year of assignment to a classroom with a poorly-behaved student (where lag 0 captures the contemporaneous effect), for various grades. 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 fixed effects. Joint test 2 shows the p-value for a joint test of equality of estimated coefficients for different lags of exposure to a poorly-behaved student. Joint test 3 shows the p-value for a joint test of equality of estimated coefficients for contemporaneous effects across grades 3 to 6. Joint test 4 shows the p-value for a joint test of equality of average contemporaneous effects in grades 3 and 4 and in grades 5 and 6. Standard errors are clustered at the classroom level. Panel C reports information on schools, classrooms and students that provide identifying variation for the estimates in this table.

Table 6: Effect of poorly-behaved students and low-achievers on student achievement

	Pooled	3 rd grade	4 th grade	5 th grade	6 th grade	Joint test 1	Joint test 2
Has poorly-behaved student	-.021*** (.005)	-.034*** (.011)	-.030** (.013)	-.008 (.010)	-.011 (.009)	.248	.049
Has low achieving student	-.001 (.006)	-.007 (.012)	.004 (.014)	.001 (.010)	-.001 (.009)	.929	.911
Joint test 3	.014	.124	.069	.566	.482		

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 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 $g-3$. Column 1 pools information across grades 3-6; and columns 2-5 report estimates from regressions by grade. 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. Joint test 1 shows the p-value for a joint test of equality of estimated coefficients for contemporaneous effects across grades 3 to 6. Joint test 2 shows the p-value for a joint test of equality of average contemporaneous effects in grades 3 and 4 and in grades 5 and 6. Joint test 3 shows the p-value for a test of equality of estimated coefficients for exposure to a poorly-behaved peer and a low-achieving peer. Standard errors are clustered at the classroom and student level when pooling data, and at the classroom level in the regressions by grade.

Table 7: Dosage effects of poorly-behaved students on achievement

Panel A	Pooled	3 rd grade	4 th grade	5 th grade	6 th grade	Joint test 1	Joint test 2
1 student	-.011** (.006)	-.029** (.013)	-.028* (.015)	.009 (.011)	.001 (.010)	.056	.007
2 students	-.037*** (.008)	-.051** (.022)	-.030 (.020)	-.037** (.014)	-.027* (.014)	.822	.650
3+students	-.056*** (.014)	-.043 (.040)	-.092** (.036)	-.032 (.022)	-.066*** (.021)	.527	.491
Joint test 3	.000	.625	.173	.003	.002		
Joint test 4	.259	.805	.398	.011	.152		
Joint test 5	.242	.380	.846	.071	.032		
Joint test 6	.387	.314	.500	.184	.707		
Panel B		3 rd grade	4 th grade	5 th grade	6 th grade		
Number of schools with differing numbers of poorly-behaved students across classrooms		119	122	134	139		
Number of classrooms in such schools		303	314	337	352		
Number of students in such schools		11,320	12,006	12,541	12,646		

Notes: Panel A of this table 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 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. Joint test 1 shows the p-value for a joint test of equality of estimated coefficients for contemporaneous effects across grades 3 to 6. Joint test 2 shows the p-value for a joint test of equality of average contemporaneous effects in grades 3 and 4 and in grades 5 and 6. Joint test 3 shows the p-value for a test of equality of estimated coefficients for exposure to different numbers of poorly-behaved peers. Joint test 4 tests whether 2*the effect of one poorly-behaved peer is the same as the effect of two disruptive peers. Joint test 5 tests whether 3*the effect of one poorly-behaved student is the same as the effect of 3+ poorly-behaved students. Joint test 6 tests whether 2*the effect of two poorly-behaved peers is the same as the effect of 3+ poorly-behaved peers. Standard errors are clustered at the classroom and student level when pooling data, and at the classroom level in the regressions by grade. Panel B reports information on schools, classrooms and students that provide identifying variation for the estimates in this table.

Table 8: 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 (.012) [.877]	-.021* (.012) [.442]	.018 (.013) [.614]	-.006 (.013) [.982]	-.042 (.026) [.450]	-.036 (.027) [.641]	-.025 (.026) [.880]	-.025 (.024) [.856]	-.051* (.026) [.251]
Panel B	EF composite	Inhibitory control	Memory and attention	Cognitive flexibility	Aggregate non- cognitive	Depression	Self- esteem	Growth mindset	Grit
1 student	.020 (.014) [.614]	-.025* (.013) [.347]	.030** (.014) [.206]	-.002 (.014) [.997]	-.039 (.029) [.641]	-.017 (.028) [.938]	-.022 (.029) [.937]	-.021 (.027) [.937]	-.058** (.028) [.232]
2 students	-.002 (.019) [.997]	-.014 (.023) [.982]	-.007 (.020) [.986]	-.018 (.020) [.926]	-.017 (.034) [.938]	-.081** (.039) [.246]	-.012 (.036) [.969]	.008 (.035) [.969]	-.013 (.036) [.969]
3+students	-.026 (.030) [.926]	.002 (.039) [.997]	-.054* (.030) [.408]	-.016 (.029) [.982]	-.133*** (.036) [.002]	-.053 (.054) [.880]	-.081** (.039) [.246]	-.144*** (.037) [.001]	-.105** (.047) [.181]
Joint test 1	.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. Joint test 1 shows the p-value for a test of equality of estimated coefficients for exposure to different numbers of poorly-behaved peers. 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. Grade 3: number of schools in which only some classrooms have a poorly-behaved student: 105; number of classrooms in such schools: 271; number of students in such schools: 10,071. Grade 4: number of schools in which only some classrooms have a poorly-behaved student: 100; number of classrooms in such schools: 264; number of students in such schools: 10,107. Grade 5: number of schools in which only some classrooms have a poorly-behaved student: 107; number of classrooms in such schools: 275; number of students in such schools: 10,204. Grade 6: number of schools in which only some classrooms have a poorly-behaved student: 104; number of classrooms in such schools: 269; number of students in such schools: 9,573.

Table 9: Dynamic effects of poorly-behaved students on student achievement

Table 3: Dynamic effects of poorly behaved students on student achievement				
Panel A	(1)		(2)	
Has poorly behaved student in g	-.021***		-.019**	
	(.005)		(.009)	
Has poorly behaved student in $g-1$			-.020**	
			(.009)	
Has poorly behaved student in g * Has poorly behaved student in $g-1$			.010	
			(.012)	
Has poorly behaved student in g * Lagged ability	.007			
	(.004)			
Panel B	Once	Twice	Three times	Four times
Has poorly behaved student	-.028	-.051**	-.070***	-.066**
	(.021)	(.021)	(.023)	(.027)

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 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 A, 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 $g-1$, 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 $g-1$, 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 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, child age and gender. Standard errors are clustered at the classroom level. Grade 3: number of schools in which only some classrooms have a poorly-behaved student: 105; number of classrooms in such schools: 271; number of students in such schools: 10,071. Grade 4: number of schools in which only some classrooms have a poorly-behaved student: 100; number of classrooms in such schools: 264; number of students in such schools: 10,107. Grade 5: number of schools in which only some classrooms have a poorly-behaved student: 107; number of classrooms in such schools: 275; number of students in such schools: 10,204. Grade 6: number of schools in which only some classrooms have a poorly-behaved student: 104; number of classrooms in such schools: 269; number of students in such schools: 9,573.

Table 10: 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. Columns 1 and 2: 45,188 student-grade observations. 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. Column 1: 101,879 student-grade observations; column 2: 41,008 student-grade observations.

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.³¹ 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

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

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

³¹ 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.

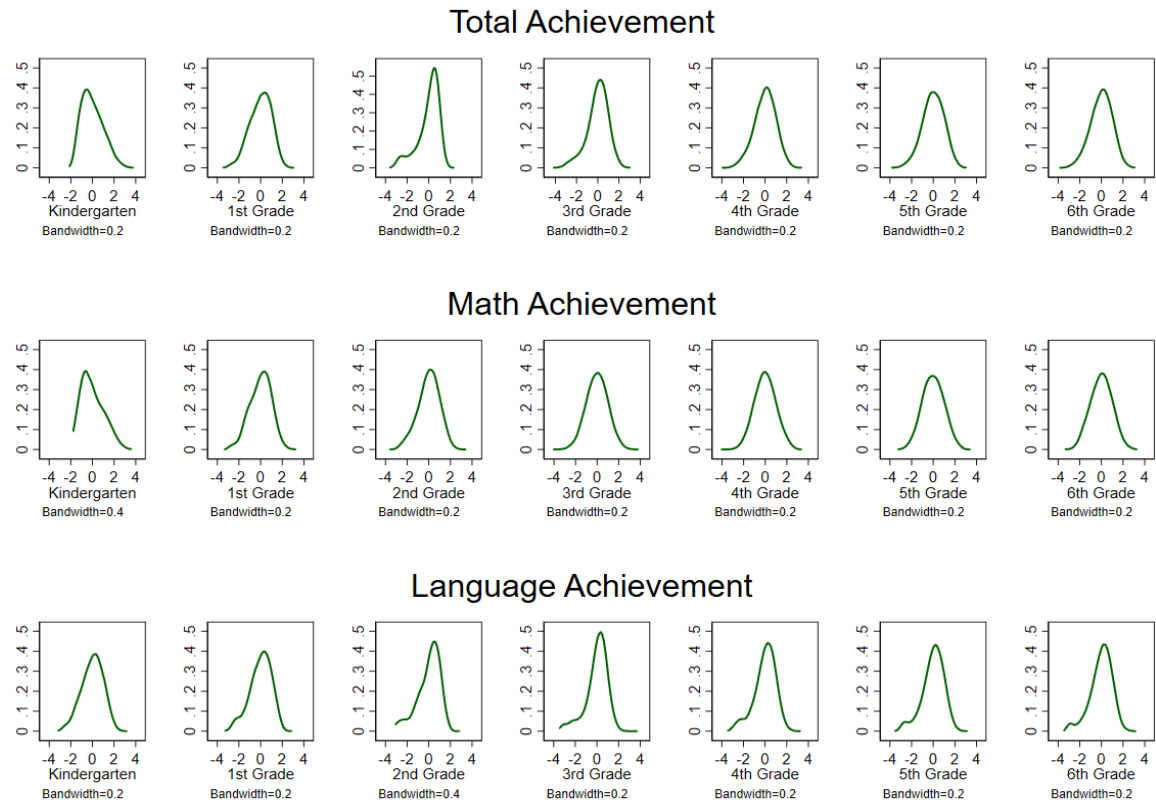
In practice the correlations of the scores across the three dimensions in our sample are low—in the range of 0.21 to 0.32 between cognitive flexibility and working memory, between 0.17 and 0.33 between working memory and inhibitory control, and in the range of 0.12 to 0.15 between cognitive flexibility and inhibitory control—see Appendix Table B1.³² 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

³² 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.

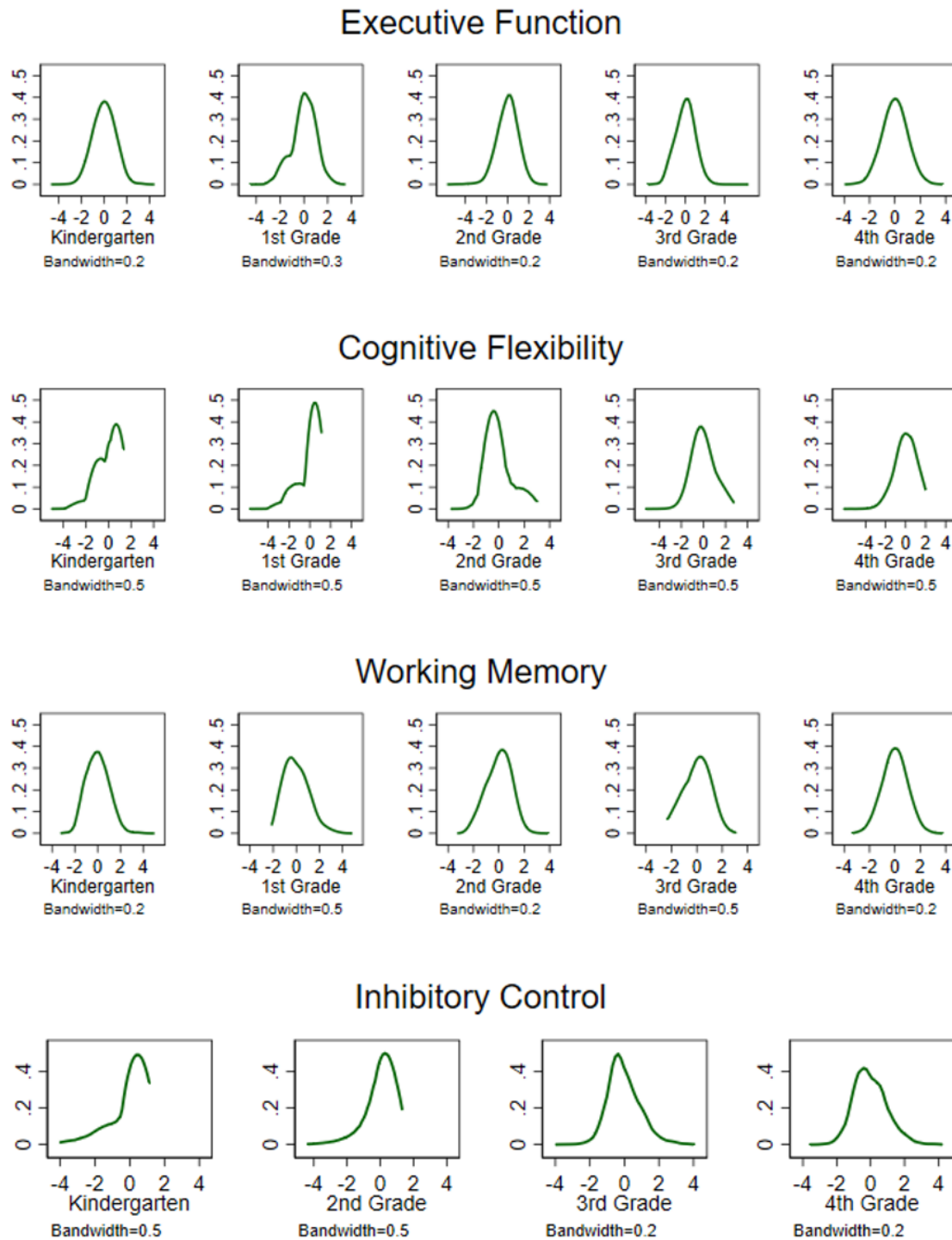
the average number of disruptive behaviors done always, as well as confidence intervals at the 95 percent level.

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

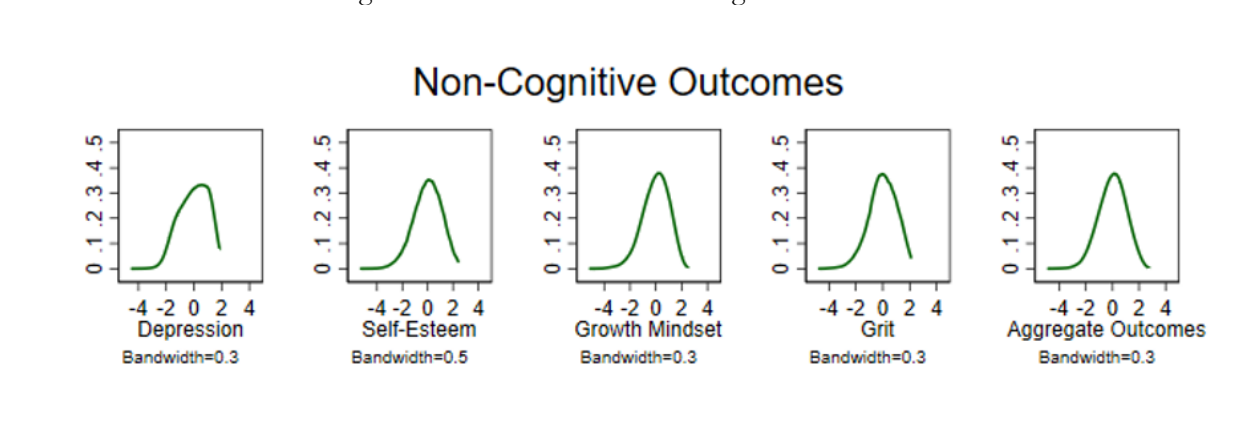
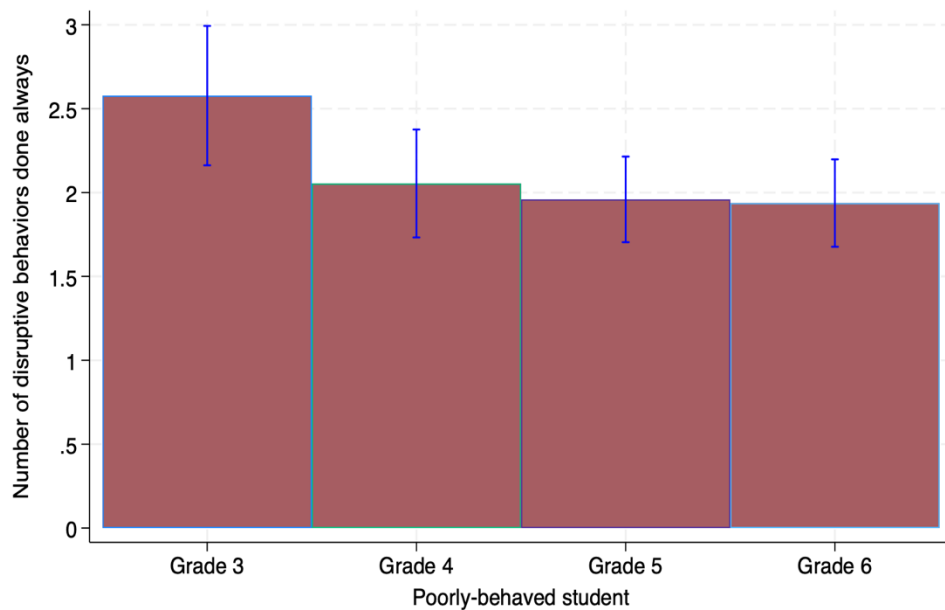


Figure B4: Behaviors of poorly-behaved students



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
	1 st Grade	
Working Memory		0.23
	2 nd Grade	
Cognitive Flexibility	0.15	
Working Memory	0.25	0.24
	3 rd Grade	
Cognitive Flexibility	0.12	
Working Memory	0.17	0.21
	4 th 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

This appendix presents additional information on the distribution of poorly-behaved students across classrooms within schools. We document the extent of variation in poorly-behaved peers across classrooms induced by the experiment, accounting for the stratified nature of randomization, and how well this variation matches what we should expect from the design.

In our setting, students are randomly assigned to classrooms within a school in every grade between Kindergarten and 6th grade. Table C3 shows the distribution of schools across cells characterized by the number of classrooms in the school and the number of poorly-behaved students in the school according to our main definition. We document that most schools have exactly two classrooms and one poorly-behaved student.

Within each school, the count of poorly-behaved students assigned to a classroom should follow a hypergeometric distribution with school-specific parameters. Figures C1 to C4 show the predicted distribution of poorly behaved students across classrooms according to the hypergeometric probability distribution, compared to the actual distribution of poorly behaved students across classrooms, for grades 3 to 6. To construct these figures, in each grade, we group schools in cells characterised by the number of classrooms in the school and the number of poorly behaved students in the school, as shown in Table C3. Then, for each classroom in each cell, we compute the predicted distribution of successes according to the hypergeometric distribution, i.e., the hypergeometric probability of k successes out of a sample of size n , from a population of size N containing K elements that have the attribute of interest – where N = school population, K = number of poorly behaved students in a given school, draw size n = number of students in a given classroom, where success is defined as obtaining an element with the attribute of interest, i.e., the poorly behaved student. Then, we average values across schools in that cell. This gives us the predicted probability of successes according to the hypergeometric probability distribution, using the specific parameters for each school in that cell. We then compare these to the actual distribution of successes within each cell. Each panel in each figure reports the number of schools in that cell. We only report values for cells that have at least two schools. The figures show that the actual distribution of successes in our sample follows the predicted distribution quite closely across all grades.

An alternative way of showing this is to compute, for each school in each (number of classrooms, number of poorly behaved students) cell, the probability of each possible configuration of how poorly behaved students can be allocated across classrooms within that school, given school size, total number of poorly behaved students within the school, and the size of each classroom within the school. For example, in a school with two classrooms and two poorly behaved students, the possible configurations

are either (2,0) (i.e. both poorly behaved students are allocated to the same classroom within the school) or (1,1) (i.e. each classroom within the school is allocated one poorly behaved student). To compute these probabilities, we use the following algorithm. Suppose a school of size N has two classrooms of size p and q respectively, and a total of K poorly behaved students. To find the probability of any possible configuration of how poorly behaved students can be allocated across classrooms within that school (a, b) we compute:

$$\frac{\binom{N-K}{p-a} \binom{K}{a} \binom{N-K-p+a}{q-b} \binom{K-a}{b}}{\binom{N}{p} \binom{N-p}{q}}$$

We then compare these predicted probabilities to the actual distribution of configurations within each cell. The results are shown in Figure C5. Each panel in this figure reports the number of schools in that cell. We only report values for cells that have at least two schools, and with two classrooms (as those are the majority of schools in our sample). The figure shows that the actual distribution of classroom configurations within schools in our sample follows the predicted distribution quite closely.

Table C1 shows the proportion of schools without poorly behaved students, the proportion of schools in which all classrooms within the school have at least one poorly behaved student, and the proportion of schools in which only some classrooms within the school have a poorly behaved student, for grades 3 to 6. For the estimation of our main specification, which regresses an outcome of student i in classroom c in grade g and school s on an indicator variable for being in the classroom with a poorly behaved student and other controls, only the latter group of schools provides variation. We show that, in all grades, more than half of the schools in the sample provides this variation. Table C2 shows estimates of our main model in the body of the paper, but restricting the sample to only schools in which only some classrooms within the school have a poorly-behaved student in them (as reported in Table C1). The results are robust to this sample restriction.

Tables C4 to C7 report the distribution of poorly behaved students across classrooms within (number of classroom, number of poorly behaved students) school cells for each grade between grades 3 and 6. For each cell, we report the proportion of schools with a given configuration of distribution of poorly-behaved students across classrooms within the school, i.e. we count how many schools have a given configuration of distribution of poorly-behaved students across classrooms within the school, and we divide that number by the total number of schools in the sample. The estimation of our dosage

specification regresses an outcome of student i in classroom c in grade g and school s on the number of poorly behaved students in the classroom and other controls. The tables show that, within each cell, most schools provide variation to estimate this specification, as only a few cases have the same number of poorly behaved students assigned to all classrooms within the school.

Table C1: Distribution of poorly-behaved students across classrooms

	Grade			
Panel A	3	4	5	6
Proportion of schools without poorly behaved students	.280	.270	.161	.141
Proportion of schools in which all classrooms have at least one poorly behaved student	.164	.201	.276	.323
Proportion of schools in which only some classrooms have a poorly behaved student	.556	.529	.563	.536
Panel B	3	4	5	6
Proportion of schools in which all classrooms have the same number of poorly-behaved students	.370	.354	.297	.286
Average class size	37.2	38.6	36.7	36.0

Notes: Panel A of this table reports the proportion of schools that have at least two classrooms and that are either without poorly behaved students, or such that all classrooms have at least one poorly behaved student, or such that only some classrooms within the school have a poorly behaved student, for grades 3 to 6. Panel B of this table reports the proportion of schools in which all classrooms have the same number of poorly-behaved students for grades 3 to 6.

Table C2: Main model estimates using restricted sample of schools

	Grade			
	3	4	5	6
Has poorly-behaved student	-0.037***	-0.030**	-0.007	-0.014
	(0.011)	(0.013)	(0.010)	(0.009)

Notes: This table reports results from our main model, but restricting the sample to only schools in which only some classrooms within the school have a poorly-behaved student in them.

Table C3: Distribution of schools in number of classrooms/number of poorly behaved students cells

Table 63: Distribution of schools in number of classrooms, number of poorly behaved students cells												
Number of classrooms		Number of poorly behaved students										
Grade 3	0	1	2	3	4	5	6	7	8	9	10	
2	45	46	26	14	1	1	0	0	0	0	0	
3	6	12	13	6	1	0	1	1	0	0	0	
4	2	2	3	2	3	0	1	0	0	0	0	
5	0	0	0	0	0	0	0	0	0	0	0	
6	0	0	1	2	0	0	0	0	0	0	0	
Grade 4	0	1	2	3	4	5	6	7	8	9	10	
2	40	43	27	7	9	4	0	0	0	0	0	
3	10	8	12	5	4	4	1	0	0	0	0	
4	1	3	0	3	1	0	1	0	0	0	0	
5	0	1	0	0	1	1	0	0	0	0	0	
6	0	0	1	1	1	0	0	0	0	0	0	
Grade 5	0	1	2	3	4	5	6	7	8	9	10	
2	24	46	35	14	9	5	2	0	0	0	0	
3	7	8	9	6	8	2	4	0	0	0	0	
4	0	1	2	2	1	2	1	1	0	0	0	
5	0	0	0	0	1	1	0	0	0	0	0	
6	0	0	0	0	0	2	0	0	0	0	0	
Grade 6	0	1	2	3	4	5	6	7	8	9	10	
2	24	40	33	16	16	2	3	0	0	0	0	
3	2	9	7	11	7	5	4	0	0	0	0	
4	0	1	1	0	5	1	0	1	0	0	1	
5	0	0	0	1	1	0	0	1	0	0	0	
6	0	0	0	0	1	0	0	0	1	0	0	

Notes: This table shows the number of schools in each cell characterised by the number of classrooms in the school and the number of poorly behaved students in the school, between grades 3 and 6.

Table C4: Distribution of poorly-behaved students within schools, by classroom configurations, grade 3

Nr classrooms	Nr poorly behaved students													
	1	2	3	4	5	6	7							
	Configura tion	Proport ion	Configura tion	Proport ion	Configura tion	Proport ion	Configura tion	Proport ion	Configura tion	Proport ion	Configura tion	Proport ion	Configura tion	Proport ion
2	(1,0)	.243	(2,0)	.063	(3,0)	.026	(3,1)	.005	(4,1)	.005	NA		NA	
			(1,1)	.074	(2,1)	.048								
3	(1,0,0)	.063	(2,0,0)	.021	(3,0,0)	.011	(2,2,0)	.005	NA		(3,2,1)	.005	(3,2,2)	.005
			(1,1,0)	.048	(2,1,0)	.005								
					(1,1,1)	.016								
4	(1,0,0,0)	.011	(2,0,0,0)	.005	(2,1,0,0)	.005	(3,1,0,0)	.005	NA		(3,1,1,1)	.005	NA	
			(1,1,0,0)	.011	(1,1,1,0)	.005	(2,1,1,0)	.011						
6	NA		(1,1,0,0,0,0)	.005	(2,1,0,0,0,0)	.005	NA		NA		NA		NA	
					(1,1,1,0,0,0)	.005								

Notes: This table reports the distribution of poorly behaved students across classrooms within schools for grade 3. We group schools in cells characterised by the number of classrooms in the school and the number of poorly behaved students in the school. Then, for each cell, we report how many schools have a given configuration of distribution of poorly-behaved students across classrooms within the school, divided by the total number of schools in the sample. For example, in a school with two classrooms and two poorly behaved students, the possible configurations are either (2,0) (i.e. both poorly behaved students are allocated to the same classroom within the school) or (1,1) (i.e. each classroom within the school is allocated one poorly behaved student). Where "NA" is reported, it means that there are no schools in that particular cell. For cells with at least one school, we only report configurations for which there is at least one school with that configuration.

Table C5: Distribution of poorly-behaved students within schools, by classroom configurations, grade 4

Nr classrooms	Nr poorly behaved students											
	1	2	3	4	5	6						
	Configuration	Proportion	Configuration	Proportion	Configuration	Proportion	Configuration	Proportion	Configuration	Proportion	Configuration	Proportion
2	(1,0)	.228	(2,0)	.074	(3,0)	.005	(4,0)	.005	(4,1)	.016	NA	
			(1,1)	.069	(2,1)	.032	(3,1)	.032	(3,2)	.005		
							(2,2)	.011				
3	(1,0,0)	.042	(2,0,0)	.011	(3,0,0)	.005	(3,1,0)	.005	(3,1,1)	.005	(4,1,1)	.005
			(1,1,0)	.053	(2,1,0)	.016	(2,2,0)	.011	(2,2,1)	.016		
					(1,1,1)	.005	(2,1,1)	.005				
4	(1,0,0,0)	.016	NA		(2,1,0,0)	.016	(2,1,1,0)	.005	NA		(3,2,1,0)	.005
5	(1,0,0,0,0)	.005	NA		NA		(2,1,1,0,0)	.005	(2,1,1,1,0)	.005	NA	
6	NA		(2,0,0,0,0,0)	.005	(1,1,1,0,0,0)	.005	(2,1,1,0,0,0)	.005	NA		NA	

Notes: This table reports the distribution of poorly behaved students across classrooms within schools for grade 4. We group schools in cells characterised by the number of classrooms in the school and the number of poorly behaved students in the school. Then, for each cell, we report how many schools have a given configuration of distribution of poorly-behaved students across classrooms within the school, divided by the total number of schools in the sample. For example, in a school with two classrooms and two poorly behaved students, the possible configurations are either (2,0) (i.e. both poorly behaved students are allocated to the same classroom within the school) or (1,1) (i.e. each classroom within the school is allocated one poorly behaved student). Where "NA" is reported, it means that there are no schools in that particular cell. For cells with at least one school, we only report configurations for which there is at least one school with that configuration.

Table C6: Distribution of poorly-behaved students within schools, by classroom configurations, grade 5

Nr classrooms	Nr poorly behaved students													
	1	2	3	4	5	6	7							
	Configura tion	Proport ion	Configura tion	Proport ion	Configura tion	Proport ion	Configura tion	Proport ion	Configura tion	Proport ion	Configura tion	Proport ion	Configura tion	Proport ion
2	(1,0)	.238	(2,0)	.078	(3,0)	.016	(3,1)	.021	(4,1)	.005	(5,1)	.005	NA	
			(1,1)	.104	(2,1)	.057	(2,2)	.026	(3,2)	.021	(3,3)	.005		
3	(1,0,0)	.041	(2,0,0)	.016	(3,0,0)	.016	(3,1,0)	.026	(3,2,0)	.005	(3,3,0)	.010	NA	
			(1,1,0)	.031	(2,1,0)	.010	(2,2,0)	.005	(3,1,1)	.005	(3,2,1)	.005		
					(1,1,1)	.005	(2,1,1)	.010			(2,2,2)	.005		
4	(1,0,0,0)	.005	(2,0,0,0)	.005	(2,1,0,0)	.010	(2,1,1,0)	.005	(4,1,0,0)	.005	(3,2,1,0)	.005	(3,2,1,1)	.005
			(1,1,0,0)	.005					(2,1,1,1)	.005				
5	NA		NA		NA		(1,1,1,1,0)	.005	(2,2,1,0,0)	.005	NA		NA	
									(1,1,1,1,1)	.005				
6	NA		NA		NA		NA		(3,1,1,0,0,0)	.005	NA		NA	
									(2,2,1,0,0,0)	.005				

Notes: This table reports the distribution of poorly behaved students across classrooms within schools for grade 5. We group schools in cells characterised by the number of classrooms in the school and the number of poorly behaved students in the school. Then, for each cell, we report how many schools have a given configuration of distribution of poorly-behaved students across classrooms within the school, divided by the total number of schools in the sample. For example, in a school with two classrooms and two poorly behaved students, the possible configurations are either (2,0) (i.e. both poorly behaved students are allocated to the same classroom within the school) or (1,1) (i.e. each classroom within the school is allocated one poorly behaved student). Where "NA" is reported, it means that there are no schools in that particular cell. For cells with at least one school, we only report configurations for which there is at least one school with that configuration.

Table C7: Distribution of poorly-behaved students within schools, by classroom configurations, grade 6

Nr poorly behaved students																		
Nr classr ooms	1		2		3		4		5		6		7		8		10	
	Config uration	Prop ortio n	Config uration	Prop ortio n	Config uration	Prop ortio n	Config uration	Prop ortio n	Config uration	Prop ortio n	Config uration	Prop ortio n	Config uration	Prop ortio n	Config uration	Prop ortio n	Config uration	Prop ortio n
2	(1,0)	.206	(2,0)	.088	(3,0)	.021	(4,0)	.005	(4,1)	.005	(3,3)	.015	NA		NA		NA	
			(1,1)	.082	(2,1)	.062	(3,1)	.052	(3,2)	.005								
							(2,2)	.026										
3	(1,0,0)	.046	(2,0,0)	.015	(3,0,0)	.005	(3,1,0)	.005	(4,1,0)	.005	(4,1,1)	.010	NA		NA		NA	
			(1,1,0)	.021	(2,1,0)	.041	(2,2,0)	.015	(3,1,1)	.005	(2,2,2)	.010						
					(1,1,1)	.010	(2,1,1)	.015	(2,2,1)	.015								
4	(1,0,0,0)	.005	(1,1,0,0)	.005	NA		(3,1,0,0)	.005	(2,1,1,1)	.005	NA		(3,3,1,0)	.005	NA		(3,3,2,2)	.005
							(2,2,0,0)	.005										
							(2,1,1,0)	.010										
							(1,1,1,1)	.005										
5	NA		NA		(1,1,1,0)	.005	(1,1,1,1)	.005	NA		NA		(2,2,2,1)	.005	NA		NA	
					,0)		,0)						,0)					
6	NA		NA		NA		(1,1,1,1)	.005	NA		NA		NA		(4,2,1,1)	.005	NA	
							,0,0)								,0,0)			

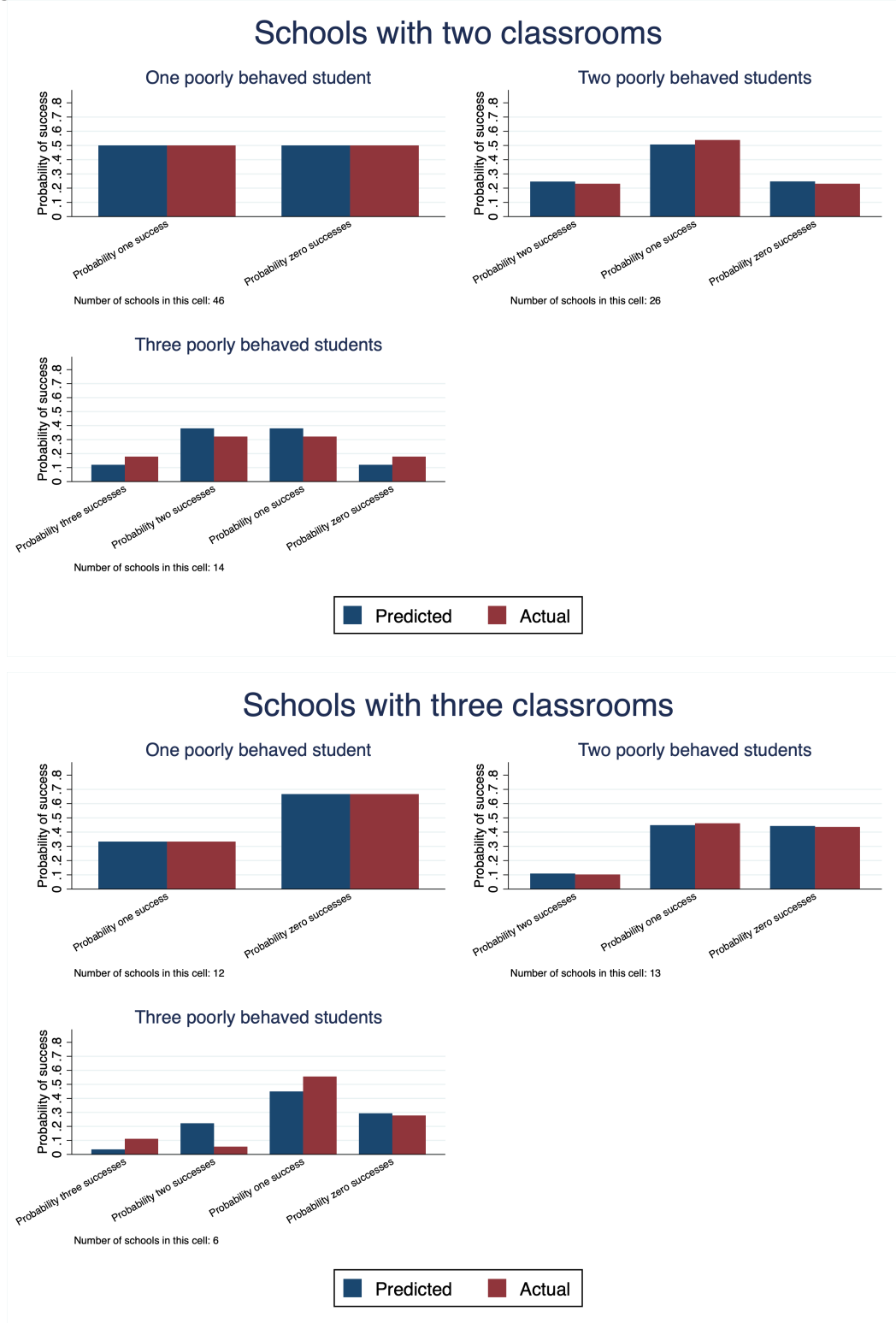
Notes: This table reports the distribution of poorly behaved students across classrooms within schools for grade 6. We group schools in cells characterised by the number of classrooms in the school and the number of poorly behaved students in the school. Then, for each cell, we report how many schools have a given configuration of distribution of poorly-behaved students across classrooms within the school, divided by the total number of schools in the sample. For example, in a school with two classrooms and two poorly behaved students, the possible configurations are either (2,0) (i.e. both poorly behaved students are allocated to the same classroom within the school) or (1,1) (i.e. each classroom within the school is allocated one poorly behaved student). Where "NA" is reported, it means that there are no schools in that particular cell. For cells with at least one school, we only report configurations for which there is at least one school with that configuration.

Table C8: Simulations of distribution of poorly-behaved students across classrooms

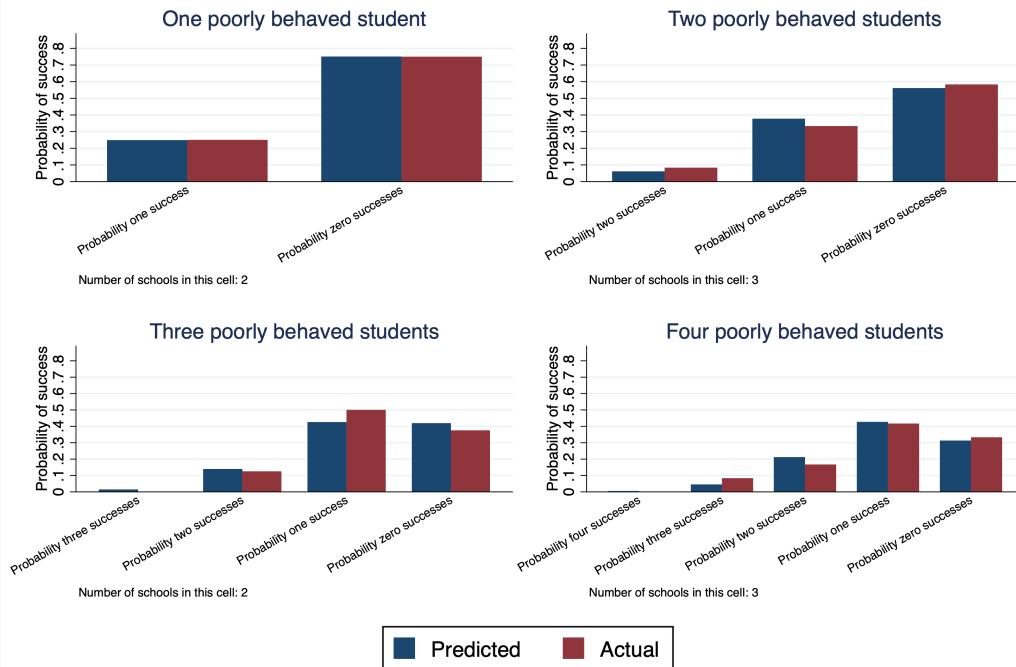
	Grade			
Double school size	3	4	5	6
Proportion of schools without poorly behaved students	.323	.323	.202	.155
Proportion of schools in which all classrooms have at least one poorly behaved student	.349	.402	.523	.598
Proportion of schools in which only some classrooms have a poorly behaved student	.328	.275	.275	.247
Average classroom size	74	77	74	72
3x school size	3	4	5	6
Proportion of schools without poorly behaved students	.307	.280	.181	.144
Proportion of schools in which all classrooms have at least one poorly behaved student	.508	.519	.674	.706
Proportion of schools in which only some classrooms have a poorly behaved student	.185	.201	.145	.149
Average classroom size	111	116	110	108
5x school size	3	4	5	6
Proportion of schools without poorly behaved students	.286	.270	.161	.139
Proportion of schools in which all classrooms have at least one poorly behaved student	.624	.635	.777	.794
Proportion of schools in which only some classrooms have a poorly behaved student	.090	.095	.062	.067
Average classroom size	185	193	184	180
10x school size	3	4	5	6
Proportion of schools without poorly behaved students	.280	.270	.161	.134
Proportion of schools in which all classrooms have at least one poorly behaved student	.698	.720	.819	.856
Proportion of schools in which only some classrooms have a poorly behaved student	.021	.011	.021	.010
Average classroom size	370	387	368	359
15x school size	3	4	5	6
Proportion of schools without poorly behaved students	.280	.270	.161	.134
Proportion of schools in which all classrooms have at least one poorly behaved student	.720	.730	.839	.866
Proportion of schools in which only some classrooms have a poorly behaved student	0	0	0	0
Average classroom size	555	580	551	539

Notes: This table reports results from simulations that explore how identifying variation would change with school size. To simulate each dataset, we increase school size by a given factor (2, 3, 5, 10, 15), keeping constant the number of classrooms within each school. Students are randomly assigned poorly-behaved student status based on the proportion of poorly-behaved students within the school observed in the real data. Then, for each simulated dataset, we report the proportion of schools that have at least two classrooms and that are either without poorly behaved students, or such that all classrooms have at least one poorly behaved student, or such that only some classrooms within the school have a poorly behaved student, for grades 3 to 6. We also report average classroom size in the simulated datasets.

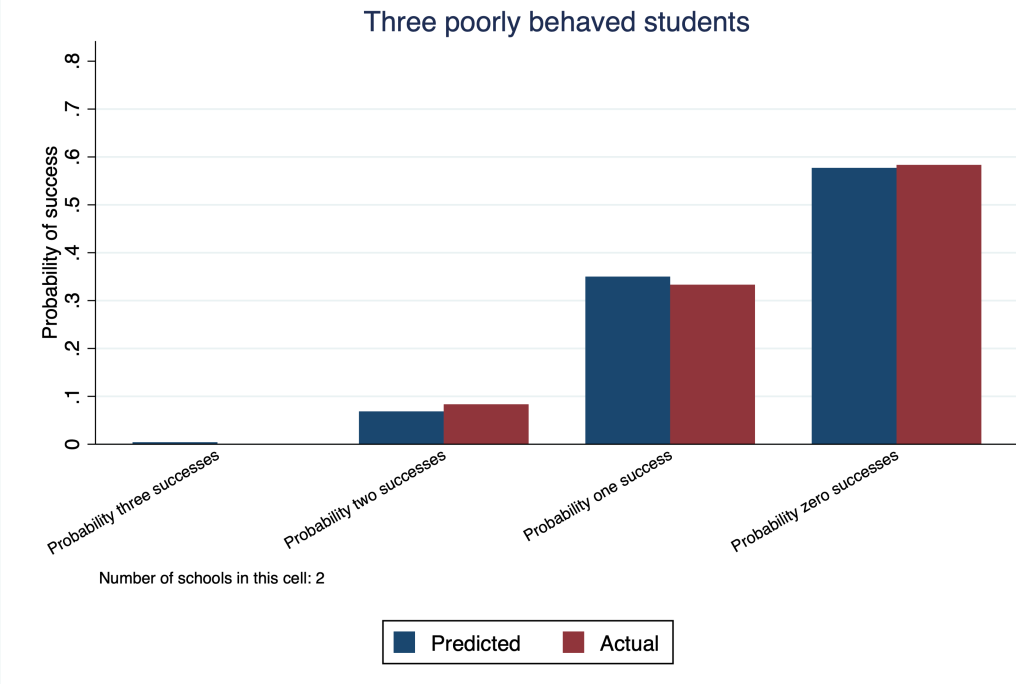
Figure C1: Predicted and actual distribution of poorly-behaved students across classrooms, grade 3



Schools with four classrooms



Schools with six classrooms

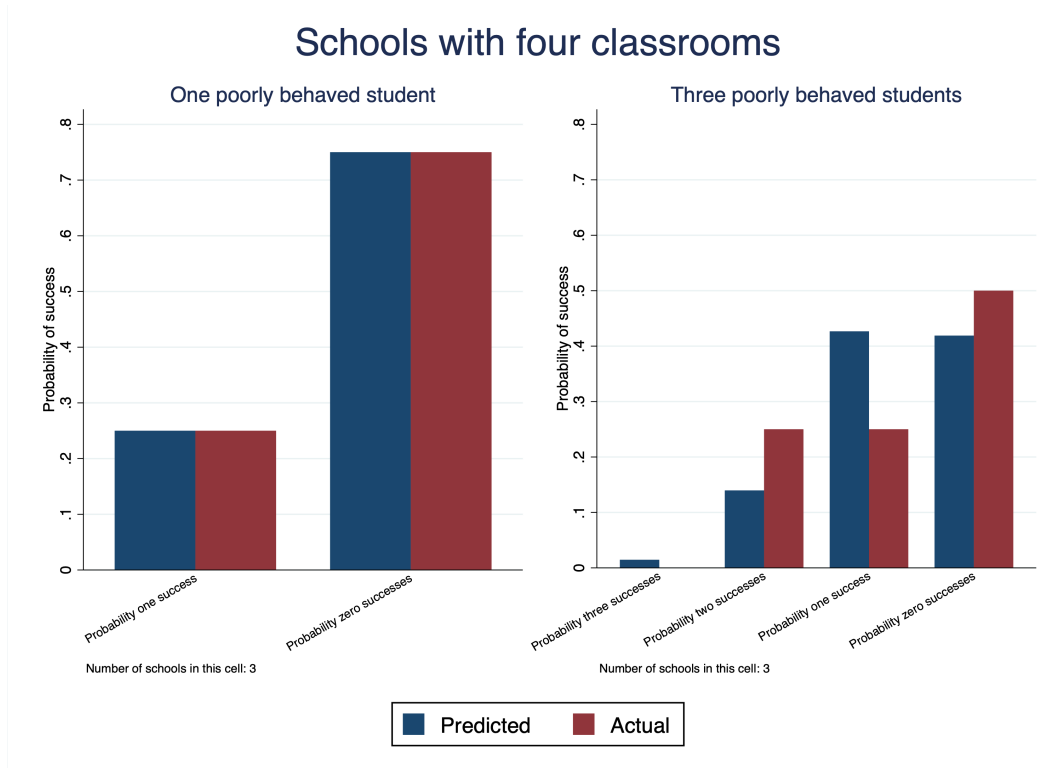


Notes: This figure shows the predicted distribution of poorly behaved student across classrooms according to the hypergeometric probability distribution, compared to the actual distribution of poorly behaved students across classrooms for grade 3. To construct this figure, we group schools in cells characterised by number of classrooms in the school and number of poorly behaved students in the

school. Then, for each classroom in each cell, we compute the predicted distribution of successes (i.e., poorly-behaved students in a classroom) according to the hypergeometric distribution, using the specific parameters for each school in that cell. We then compare these to the actual distribution of successes within each cell. Each panel reports the number of schools in that cell at the bottom of the bar chart. We only report values for cells that have at least two schools.

Figure C2: Predicted and actual distribution of poorly-behaved students across classrooms, grade 4



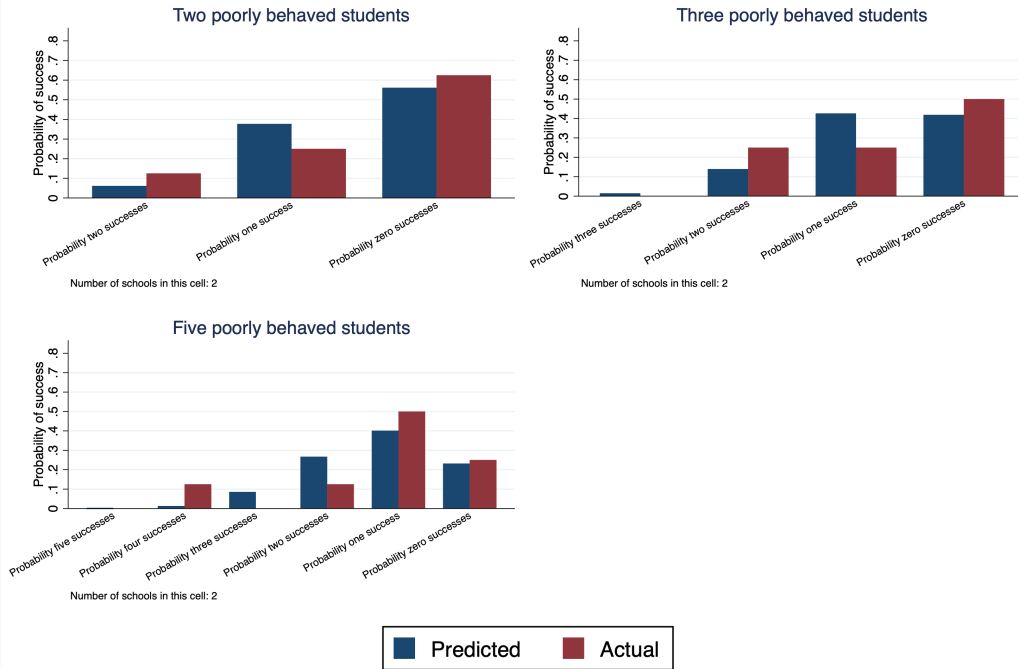


Notes: This figure shows the predicted distribution of poorly behaved student across classrooms according to the hypergeometric probability distribution, compared to the actual distribution of poorly behaved students across classrooms for grade 4. To construct this figure, we group schools in cells characterised by number of classrooms in the school and number of poorly behaved students in the school. Then, for each classroom in each cell, we compute the predicted distribution of successes (i.e., poorly-behaved students in a classroom) according to the hypergeometric distribution, using the specific parameters for each school in that cell. We then compare these to the actual distribution of successes within each cell. Each panel reports the number of schools in that cell at the bottom of the bar chart. We only report values for cells that have at least two schools.

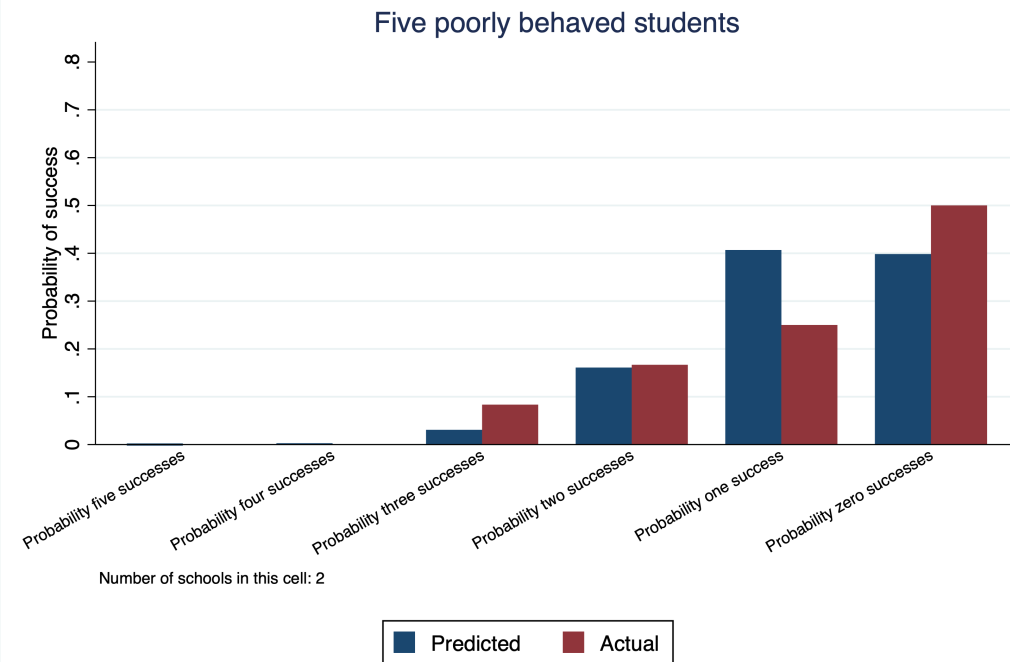
Figure C3: Predicted and actual distribution of poorly-behaved students across classrooms, grade 5



Schools with four classrooms



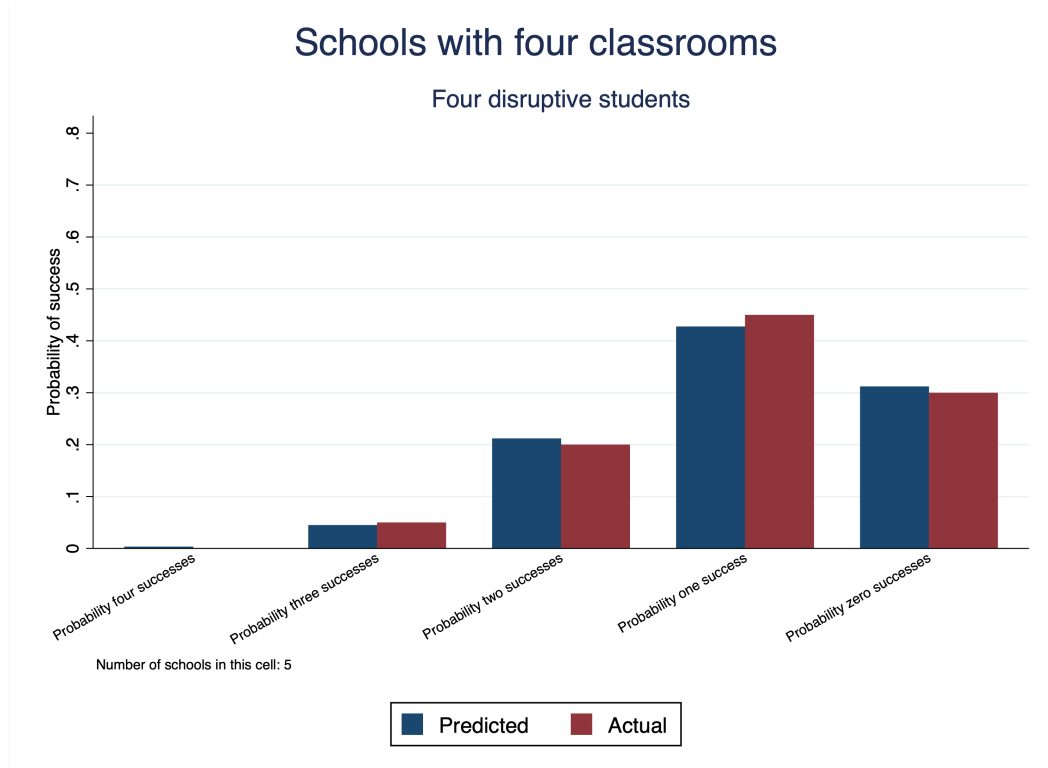
Schools with six classrooms



Notes: This figure shows the predicted distribution of poorly behaved student across classrooms according to the hypergeometric probability distribution, compared to the actual distribution of poorly behaved students across classrooms for grade 5. To construct this figure, we group schools in cells characterised by number of classrooms in the school and number of poorly behaved students in the school. Then, for each classroom in each cell, we compute the predicted distribution of successes (i.e., poorly-behaved students in a classroom) according to the hypergeometric distribution, using the specific parameters for each school in that cell. We then compare these to the actual distribution of successes within each cell. Each panel reports the number of schools in that cell at the bottom of the bar chart. We only report values for cells that have at least two schools.

Figure C4: Predicted and actual distribution of poorly-behaved students across classrooms, grade 6

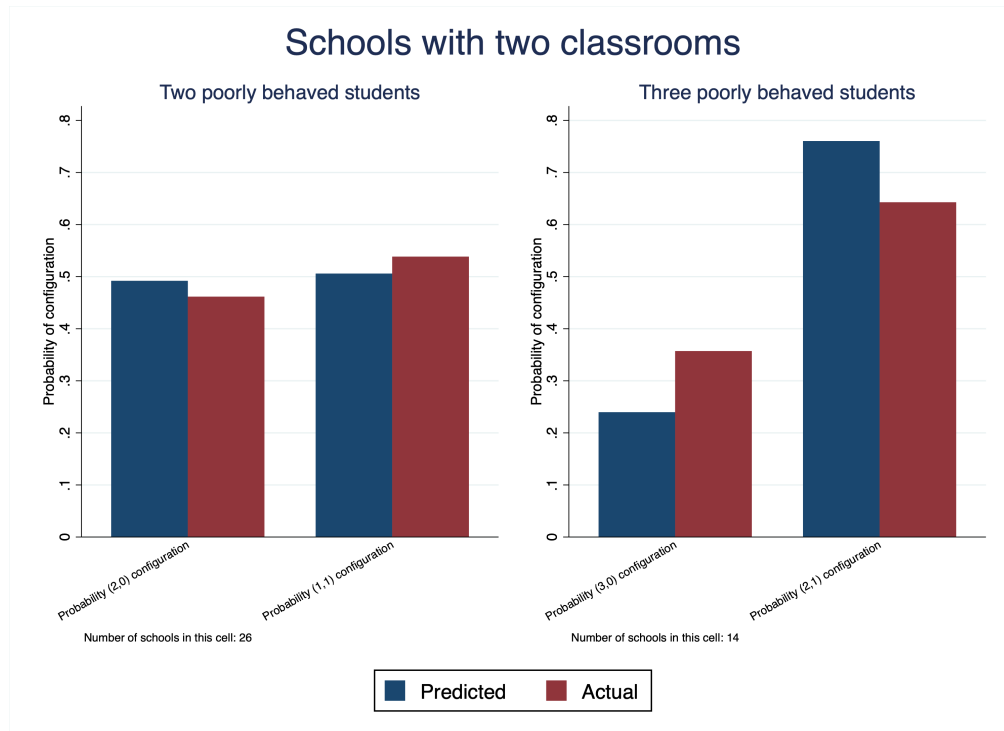




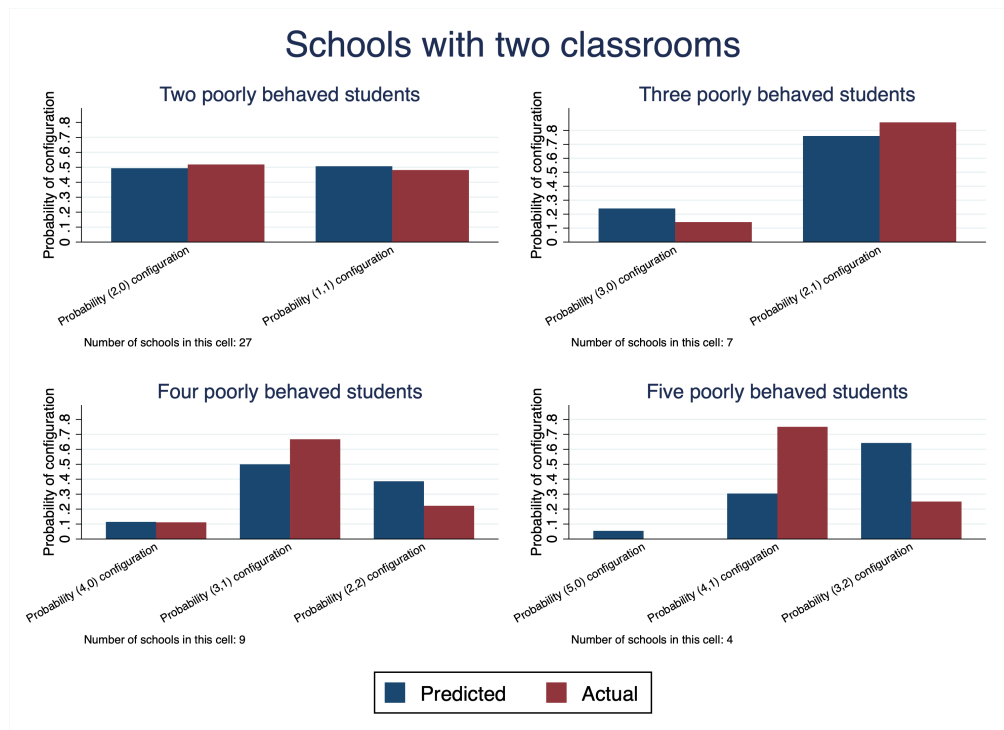
Notes: This figure shows the predicted distribution of poorly behaved student across classrooms according to the hypergeometric probability distribution, compared to the actual distribution of poorly behaved students across classrooms for grade 6. To construct this figure, we group schools in cells characterised by number of classrooms in the school and number of poorly behaved students in the school. Then, for each classroom in each cell, we compute the predicted distribution of successes (i.e., poorly-behaved students in a classroom) according to the hypergeometric distribution, using the specific parameters for each school in that cell. We then compare these to the actual distribution of successes within each cell. Each panel reports the number of schools in that cell at the bottom of the bar chart. We only report values for cells that have at least two schools.

Figure C5: Predicted and actual distribution of classroom configurations

Grade 3

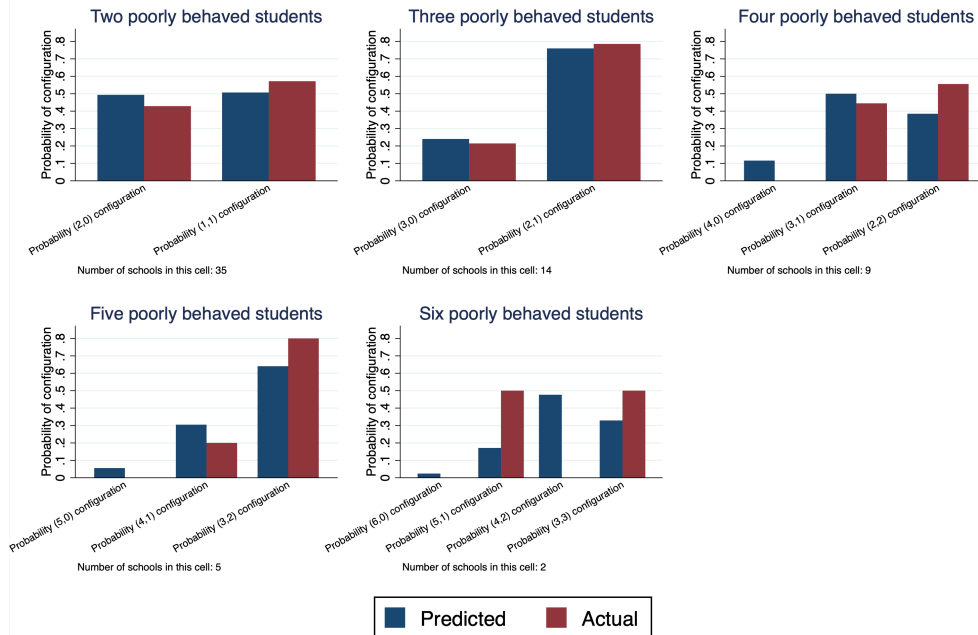


Grade 4



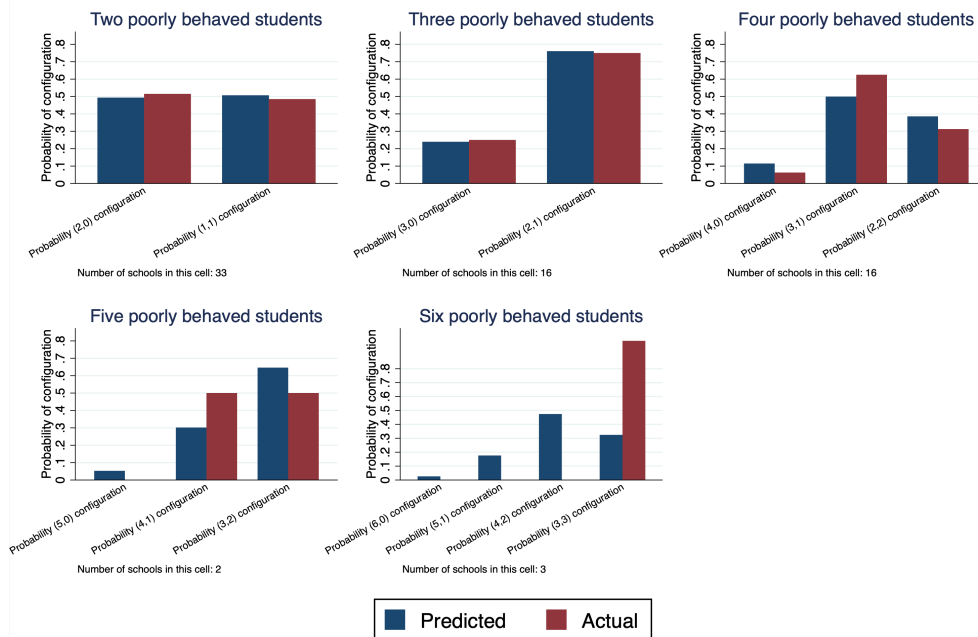
Grade 5

Schools with two classrooms



Grade 6

Schools with two classrooms



Notes: This figure shows the predicted distribution of poorly behaved student across classrooms, compared to the actual distribution of poorly behaved students across classrooms between grades 3 and 6, for schools with two classrooms. To construct this figure, we group schools in cells characterised by the number of classrooms in the school and the number of poorly behaved students in the

school. Then, for each cell, we compute the probability of each possible configuration of how poorly behaved students can be allocated across classrooms within a school, given school size, total number of poorly behaved students within the school, and the size of each classroom within the school. For example, in a school with two classrooms and two poorly behaved students, the possible configurations are either (2,0) (i.e. both poorly behaved students are allocated to the same classroom within the school) or (1,1) (i.e. each classroom within the school is allocated one poorly behaved student). We then compare these to the actual distribution of configurations within each cell. Each panel reports the number of schools in that cell at the bottom of the bar chart. We only report values for cells that have at least two schools, and with two classrooms (as those are the majority of schools in our sample).

Appendix D

Table D1 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 D2 shows the distribution of low-achieving students across classrooms.

Table D1: Balance test				
	Math+Language index	Age	Female	Family factor
Has poorly-behaved student	-.013 (.008)	-.033 (.065)	-.003* (.002)	-.025 (.015)

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 D2: Distribution of low-achieving students				
	(1) Average number of low-achieving students per school	(2) Average classroom size	(3) Average school size	(4) Proportion of schools in which only some classrooms have a low- achieving student
3 rd grade	1.42	37.0	89.7	.503
4 th grade	2.03	38.7	94.3	.466
5 th grade	2.58	36.8	88.3	.440
6 th grade	2.67	35.9	87.1	.464

Note: The table shows descriptive statistics about the distribution of low-achieving students across schools 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 worst achieving students in the classroom according to the teacher in all grades between $g-1$, $g-2$ and $g-3$. In column 1 we report the average number of low-achieving students within a school. Columns 2 and 3 report average classroom and school size. Column 4 reports the proportion of schools in which only some (but not all) classrooms have a low-achieving student. The total number of low-achieving students is 273 in grade 3, 396 in grade 4, 517 in grade 5, 530 in grade 6. The sample only considers schools in which there are at least two classrooms.

Appendix E

Tables E1 to E6 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 E1, E2 and E3 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 E1, E2 and E3 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 E4, E5 and E6 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. Tables E7 to E12 do the same robustness check but changing the definition of low-achieving students to explore different teacher rating cutoffs and different numbers of years considered.

Table E13 reports correlations between teachers' evaluations of students and student achievement in a given grade. Each coefficient corresponds to a separate regression of an indicator for being among

the 5 worst behaved students in a given grade according to the teacher (Panel A) or an indicator for being among the 5 worst achievers in a given grade according to the teacher (Panel B) on achievement in that grade, controlling for school fixed effects.

Table E14 reports estimates for our model in which we include indicators for having a poorly-behaved student in the classroom as well as having a low-achieving student in the classroom simultaneously, but changing the definition we use for low-achieving students to be based on student performance instead of teachers' assessment. In particular, in this table, a low-achieving student in grade g is defined as such if they were in the bottom 10% of the math and language index distribution in every grade $g-1$, $g-2$ and $g-3$.

Table E1: Effect of poorly-behaved students on student achievement, using grades g-1 and g-2 to define poorly-behaved students

	Rank 1					
	Pooled	2014	2015	2016	2017	2018
Has poorly-behaved student	-.003 (.006)	.002 (.018)	.002 (.014)	-.011 (.016)	.003 (.012)	-.012 (.011)
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 (.005)	-.008 (.013)	.003 (.012)	.006 (.011)	.004 (.009)	-.015* (.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	-.007 (.005)	-.016 (.013)	-.003 (.011)	.001 (.013)	.000 (.011)	-.014 (.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	-.016** (.006)	-.022 (.014)	-.009 (.012)	-.023 (.015)	-.009 (.011)	-.013 (.015)
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	-.012* (.007)	-.023 (.016)	.001 (.013)	-.021 (.015)	-.014 (.013)	.010 (.017)
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 g-1 and g-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 E2: Effect of poorly-behaved students on student achievement, using grades g-1, g-2 and g-3 to define poorly-behaved students

	Rank 1				
	Pooled	2015	2016	2017	2018
Has poorly-behaved student	-.029*** (.011)	-.022 (.022)	-.037 (.028)	-.021 (.020)	-.035* (.019)
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	-.012* (.007)	-.003 (.016)	-.016 (.017)	-.002 (.012)	-.023** (.011)
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	-.015*** (.005)	-.017 (.013)	-.024* (.014)	.001 (.010)	-.021** (.010)
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	-.023*** (.005)	-.034*** (.011)	-.027** (.014)	-.010 (.010)	-.021** (.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	-.021*** (.005)	-.034*** (.011)	-.030** (.013)	-.008 (.010)	-.012 (.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 g-1, g-2 and g-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 E3: Effect of poorly-behaved students on student achievement, using grades g-1, g-2, g-3 and g-4 to define poorly-behaved students

	Rank 1			
	Pooled	2016	2017	2018
Has poorly-behaved student	-.039*	-.070	-.034	-.012
	(.021)	(.047)	(.024)	(.035)
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	-.019	-.067**	.010	-.007
	(.012)	(.026)	(.017)	(.014)
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	-.018**	-.039**	.000	-.021*
	(.008)	(.019)	(.012)	(.011)
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	-.012*	-.007	-.018	-.011
	(.007)	(.015)	(.012)	(.010)
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	-.011*	-.009	-.011	-.012
	(.006)	(.015)	(.011)	(.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 g-1, g-2, g-3 and g-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 E4: 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	-.007 (.008)	.002 (.018)	-.016 (.016)	-.019 (.018)	-.003 (.012)	.002 (.014)
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	-.010* (.005)	-.008 (.013)	-.006 (.011)	-.025* (.014)	-.011 (.010)	.000 (.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	-.013** (.005)	-.016 (.013)	-.023** (.010)	-.016 (.014)	-.000 (.010)	-.007 (.010)
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	-.016*** (.005)	-.022 (.014)	-.035*** (.011)	-.020 (.013)	-.004 (.009)	-.003 (.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	-.015*** (.005)	-.023 (.016)	-.031*** (.011)	-.024* (.013)	-.007 (.009)	.003 (.009)
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 E5: Effect of poorly-behaved students on student achievement, using first three grades to define poorly-behaved students

	Rank 1				
	Pooled	2015	2016	2017	2018
Has poorly-behaved student	-.019 (.012)	-.022 (.022)	-.041 (.028)	.001 (.015)	-.008 (.023)
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	-.010 (.008)	-.003 (.016)	-.034* (.019)	-.008 (.014)	.007 (.015)
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	-.019*** (.006)	-.017 (.013)	-.026* (.015)	-.025** (.011)	-.007 (.012)
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	-.016*** (.006)	-.034*** (.011)	-.015 (.013)	-.023** (.010)	.009 (.010)
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	-.015*** (.005)	-.034*** (.011)	-.028** (.013)	-.020** (.009)	.020** (.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 E6: 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	-.034 (.025)	-.070 (.047)	-.028 (.020)	.008 (.033)
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	-.012 (.014)	-.067** (.026)	.015 (.020)	.030 (.018)
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	-.020** (.010)	-.039** (.019)	-.012 (.015)	-.006 (.016)
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 (.008)	-.007 (.015)	-.017 (.012)	.005 (.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 (.007)	-.009 (.015)	-.021* (.011)	.013 (.011)
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 E7: Effect of low achieving students on student achievement, using grades g-1 and g-2 to define low achieving students

	Rank 1					
	Pooled	2014	2015	2016	2017	2018
Has low achieving student	-.012*	-.018	-.014	.024	-.036***	-.017
	(.006)	(.018)	(.014)	(.018)	(.012)	(.011)
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	-.010	.017	-.019**	-.015*
	(.005)	(.014)	(.012)	(.013)	(.009)	(.009)
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*	-.006	-.017	.010	-.020**	-.016
	(.005)	(.013)	(.012)	(.014)	(.010)	(.010)
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	-.007	.027*	-.022	.003	-.014	-.037***
	(.006)	(.014)	(.014)	(.015)	(.013)	(.013)
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	.026	-.010	.012	-.002	-.032*
	(.008)	(.017)	(.016)	(.020)	(.015)	(.018)
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 g-1 and g-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 E8: Effect of low achieving students on student achievement, using grades g-1, g-2 and g-3 to define low achieving students

	Rank 1				
	Pooled	2015	2016	2017	2018
Has low achieving student	-.007 (.015)	-.026 (.022)	.051 (.035)	-.040* (.024)	-.024 (.022)
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	-.004 (.007)	.006 (.017)	.010 (.017)	-.019 (.012)	-.003 (.012)
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	-.013** (.006)	-.021 (.015)	-.022 (.014)	-.018* (.009)	.003 (.009)
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 (.006)	-.007 (.013)	-.014 (.014)	-.011 (.010)	-.005 (.009)
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 (.006)	-.008 (.012)	.003 (.015)	.001 (.010)	-.004 (.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 g-1, g-2 and g-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 E9: Effect of low achieving students on student achievement, using grades g-1, g-2, g-3 and g-4 to define low achieving students

	Rank 1			
	Pooled	2016	2017	2018
Has low achieving student	-.028 (.027)	.075 (.061)	-.066** (.032)	-.066** (.028)
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	.018 (.011)	.056* (.030)	.007 (.020)	.008 (.013)
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 (.008)	-.020 (.021)	-.019 (.012)	.027** (.012)
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 (.007)	.015 (.016)	-.010 (.011)	.018* (.010)
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 (.006)	.014 (.015)	-.004 (.009)	.016* (.009)
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 g-1, g-2, g-3 and g-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 E10: 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	-.024** (.010)	-.018 (.018)	-.046** (.021)	.018 (.020)	-.059** (.025)	-.018 (.018)
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 (.006)	.000 (.014)	.009 (.014)	-.001 (.014)	-.025** (.012)	.007 (.013)
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	-.007 (.006)	-.006 (.013)	-.007 (.013)	-.009 (.013)	-.028** (.011)	.015 (.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	.005 (.005)	.027* (.014)	-.001 (.011)	-.003 (.013)	-.027*** (.010)	.023** (.010)
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 (.006)	.026 (.017)	-.013 (.012)	.002 (.013)	-.022** (.010)	.017* (.010)
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 E11: Effect of low achieving students on student achievement, using first three grades to define low achieving students

	Rank 1				
	Pooled	2015	2016	2017	2018
Has low achieving student	-.021 (.016)	-.026 (.022)	-.002 (.038)	-.000 (.023)	-.061*** (.020)
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 (.009)	.006 (.017)	.015 (.019)	.009 (.017)	-.039** (.019)
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 (.008)	-.021 (.015)	-.009 (.015)	.011 (.015)	.009 (.014)
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)	-.007 (.013)	.007 (.013)	.006 (.012)	.021** (.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 (.006)	-.008 (.012)	.011 (.013)	-.015 (.010)	.029*** (.010)
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 E12: Effect of low achieving students on student achievement, using first four grades to define low achieving students

	Rank 1			
	Pooled	2016	2017	2018
Has low achieving student	.031 (.034)	.075 (.061)	.026 (.043)	-.039 (.034)
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 (.016)	.056* (.030)	-.007 (.024)	-.065*** (.022)
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	-.018* (.011)	-.020 (.021)	-.032* (.017)	-.002 (.016)
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 (.008)	.015 (.016)	-.012 (.014)	.014 (.012)
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 (.007)	.014 (.015)	-.019 (.012)	.033*** (.011)
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.

Table E13: Correlations between teacher evaluations and student achievement

Panel A	Child is among bottom 5 worst behaved according to teacher						
	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Math+language index decile (omitted: 10)							
1	0.0744*** (0.0127)	0.0717*** (0.0103)	0.173*** (0.0115)	0.147*** (0.0109)	0.179*** (0.0119)	0.207*** (0.0116)	0.171*** (0.0110)
2	0.0652*** (0.0125)	0.0694*** (0.0101)	0.128*** (0.0114)	0.110*** (0.0108)	0.156*** (0.0117)	0.157*** (0.0115)	0.150*** (0.0109)
3	0.0599*** (0.0125)	0.0459*** (0.0101)	0.101*** (0.0113)	0.0856*** (0.0107)	0.126*** (0.0117)	0.128*** (0.0114)	0.116*** (0.0108)
4	0.0657*** (0.0124)	0.0390*** (0.0101)	0.0606*** (0.0113)	0.0664*** (0.0106)	0.123*** (0.0116)	0.109*** (0.0114)	0.0860*** (0.0108)
5	0.0472*** (0.0124)	0.0432*** (0.0100)	0.0511*** (0.0113)	0.0553*** (0.0106)	0.0941*** (0.0116)	0.0928*** (0.0113)	0.0789*** (0.0108)
6	0.0231* (0.0124)	0.0104 (0.0100)	0.0507*** (0.0112)	0.0487*** (0.0106)	0.0816*** (0.0115)	0.0735*** (0.0114)	0.0621*** (0.0108)
7	0.0297** (0.0123)	0.0113 (0.00999)	0.0241** (0.0112)	0.0373*** (0.0105)	0.0472*** (0.0116)	0.0617*** (0.0113)	0.0666*** (0.0107)
8	0.0193 (0.0123)	0.0204** (0.00998)	0.0320*** (0.0112)	0.0276*** (0.0105)	0.0347*** (0.0115)	0.0406*** (0.0113)	0.0314*** (0.0107)
9	0.00408 (0.0122)	0.00984 (0.00995)	0.0298*** (0.0112)	0.00658 (0.0105)	0.0146 (0.0115)	0.0191* (0.0112)	0.0135 (0.0107)
Panel B	Child is among bottom 5 worst achievers according to teacher						
	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Math+language index decile (omitted: 10)							
1	0.416*** (0.0124)	0.472*** (0.0106)	0.630*** (0.0104)	0.572*** (0.0106)	0.604*** (0.0106)	0.541*** (0.0108)	0.525*** (0.0111)
2	0.299*** (0.0122)	0.326*** (0.0104)	0.359*** (0.0103)	0.302*** (0.0104)	0.335*** (0.0105)	0.320*** (0.0107)	0.311*** (0.0110)
3	0.204*** (0.0122)	0.208*** (0.0104)	0.210*** (0.0102)	0.191*** (0.0103)	0.205*** (0.0104)	0.222*** (0.0107)	0.216*** (0.0109)
4	0.155*** (0.0121)	0.138*** (0.0103)	0.129*** (0.0102)	0.125*** (0.0103)	0.149*** (0.0104)	0.147*** (0.0106)	0.155*** (0.0109)
5	0.120*** (0.0121)	0.0864*** (0.0103)	0.0926*** (0.0102)	0.0925*** (0.0103)	0.107*** (0.0103)	0.109*** (0.0106)	0.103*** (0.0109)
6	0.0927*** (0.0120)	0.0573*** (0.0103)	0.0574*** (0.0101)	0.0567*** (0.0102)	0.0752*** (0.0103)	0.0746*** (0.0106)	0.0857*** (0.0109)
7	0.0612*** (0.0120)	0.0348*** (0.0102)	0.0347*** (0.0101)	0.0516*** (0.0102)	0.0530*** (0.0103)	0.0487*** (0.0106)	0.0635*** (0.0108)
8	0.0343*** (0.0119)	0.0221** (0.0102)	0.0338*** (0.0101)	0.0382*** (0.0102)	0.0372*** (0.0103)	0.0340*** (0.0106)	0.0448*** (0.0108)
9	0.0177 (0.0119)	0.0137 (0.0102)	0.00793 (0.0101)	0.0235** (0.0102)	0.0194* (0.0102)	0.0138 (0.0105)	0.0139 (0.0108)

Notes: This table reports correlations between teacher evaluations of students and student achievement. Each coefficient corresponds to a separate regression in which we regress an indicator for being among the 5 worst behaved students in a given grade according to the teacher (Panel A) or an indicator for being among the 5 worst achievers in a given grade according to the teacher (Panel B) on achievement in that grade, controlling for school fixed effects.

Table E14: Effect of poorly-behaved students and low-achievers on student achievement

	Pooled	3 rd grade	4 th grade	5 th grade	6 th grade	Joint test 1	Joint test 2
Has poorly-behaved student	-.018*** (.005)	-0.0335*** (0.0106)	-0.0213* (0.0124)	-0.0133 (0.00979)	-0.00360 (0.00907)	.196	.082
Has low achieving student	-.008 (.006)	0.00451 (0.0133)	-0.0109 (0.0136)	-0.0225* (0.0122)	-0.00411 (0.0103)	.495	.427
Joint test 3	.216	.026	.570	.544	.974		

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 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 in the bottom 10% of the math and language index distribution in all grades between $g-1$, $g-2$ and $g-3$. Column 1 pools information across grades 3-6; and columns 2-5 report estimates from regressions by grade. 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. Joint test 1 shows the p-value for a joint test of equality of estimated coefficients for contemporaneous effects across grades 3 to 6. Joint test 2 shows the p-value for a joint test of equality of average contemporaneous effects in grades 3 and 4 and in grades 5 and 6. Joint test 3 shows the p-value for a test of equality of estimated coefficients for exposure to a poorly-behaved peer and a low-achieving peer. Standard errors are clustered at the classroom and student level when pooling data, and at the classroom level in the regressions by grade.

Appendix F

In this appendix, we show grade-by-grade estimates of equation (3.2) in the main body of the paper. The peer variables we consider in these regressions are the presence of a poorly behaved student, the presence of a low achieving student, the mean lagged student achievement in math and language, the standard deviation of lagged achievement in math and language (or different quantiles of the distribution of this variable), the mean lagged achievement in EF, the standard deviation of lagged achievement in EF (or different quantiles of the distribution of this variable), the mean student age, and the proportion of female students. In addition, we also consider the following variables, constructed for students who were present in the sample at baseline (i.e., at the start of kindergarten): the mean family wealth at baseline, the standard deviation of family wealth at baseline, the proportion of students whose mother's education is below secondary, the proportion of students whose father's education is below secondary, the mean vocabulary score at baseline, the standard deviation of the vocabulary score at baseline, and the proportion of students who were in preschool.

Tables F1 and F2 present the results. In the first column we include one variable at a time, while in the second we include all peer characteristics simultaneously. Coefficients and standard errors are shown, as well as p-values for the test that each coefficient is equal to zero after accounting for multiple hypothesis testing, using the procedure in Romano and Wolf (2005). The tables report the noise-corrected R-square to account for measurement error in the demeaned predicted classroom effect.

Table F1: Effect of peer- and non-peer characteristics on classroom value added

	3rd grade		4th grade		5th grade		6th grade	
	Separate regressions	Joint regression	Separate regressions	Joint regression	Separate regressions	Joint regression	Separate regressions	Joint regression
Has poorly-behaved student	-0.0336** (0.0155) [.360]	-0.0370** (0.0161) [.320]	-0.0330* (0.0173) [.558]	-0.0436** (0.0183) [.288]	-0.00669 (0.0133) [.997]	-0.0102 (0.0134) [.995]	-0.0161 (0.0119) [.871]	-0.0111 (0.0129) [.992]
Has low-achieving student	-0.0131 (0.0162) [.992]	-0.0171 (0.0171) [.978]	-0.00647 (0.0193) [.999]	-0.00367 (0.0208) [1.00]	-0.00351 (0.0132) [.997]	-0.00276 (0.0132) [1.00]	-0.0106 (0.0116) [.987]	-0.00636 (0.0129) [.996]
Mean lagged math+language index	0.00185 (0.0479) [1.00]	-0.0521 (0.0610) [.986]	-0.0724 (0.0565) [.935]	-0.110 (0.0763) [.854]	0.0509 (0.0366) [.868]	0.00812 (0.0508) [1.00]	0.0290 (0.0310) [.987]	0.0306 (0.0328) [.991]
SD lagged math+language index	0.0369 (0.0573) [.998]	0.00783 (0.0738) [.998]	-0.0277 (0.0624) [.999]	-0.127 (0.0785) [.772]	-0.0163 (0.0522) [.997]	-0.00514 (0.0723) [1.00]	-0.00866 (0.0477) [.997]	0.00986 (0.0524) [.996]
Mean lagged executive function	0.0586 (0.0501) [.953]	0.109* (0.0633) [.714]	-0.0526 (0.0506) [.966]	-0.00389 (0.0616) [1.00]	0.0574 (0.0377) [.804]	0.0396 (0.0471) [.993]		
SD lagged executive function	0.0839 (0.0543) [.794]	0.106* (0.0633) [.721]	0.0297 (0.0606) [.999]	0.01000 (0.0709) [1.00]	0.0199 (0.0491) [.997]	0.0292 (0.0662) [.999]		
Mean age	-0.00267 (0.00596) [.998]	-0.00486 (0.00597) [.986]	-3.94e-05 (0.00567) [1.00]	3.85e-05 (0.00591) [1.00]	-0.00190 (0.00447) [.997]	-0.00276 (0.00449) [.998]	-0.00195 (0.00370) [.995]	-0.00164 (0.00380) [.996]
Proportion female	-0.107 (0.198) [.998]	-0.0847 (0.217) [.998]	-0.276 (0.286) [.971]	-0.185 (0.299) [.998]	0.154 (0.191) [.989]	0.258 (0.198) [.900]	-0.391* (0.203) [.486]	-0.386* (0.214) [.617]
Mean wealth at baseline	0.00282 (0.0392) [1.00]	0.00443 (0.0529) [.998]	0.00650 (0.0520) [1.00]	0.0334 (0.0584) [.998]	-0.0381 (0.0307) [.904]	-0.0636 (0.0408) [.783]	-0.0199 (0.0255) [.990]	-0.0269 (0.0357) [.994]
SD wealth at baseline	0.0474 (0.0537) [.989]	0.0403 (0.0580) [.988]	-0.000837 (0.0660) [1.00]	-0.00946 (0.0653) [1.00]	-0.0125 (0.0420) [.997]	-0.0148 (0.0449) [1.00]	-0.0190 (0.0450) [.997]	-0.0290 (0.0465) [.996]
Proportion mother education less than secondary	0.00994 (0.0692) [1.00]	0.0731 (0.0829) [.986]	0.00791 (0.0772) [1.00]	0.0340 (0.0788) [.999]	0.0359 (0.0529) [.990]	0.0208 (0.0626) [1.00]	-0.00226 (0.0490) [.997]	-0.0358 (0.0643) [.996]
Proportion father education less than secondary	-0.0620 (0.0603) [.974]	-0.0780 (0.0700) [.956]	-0.0306 (0.0752) [.999]	-0.0894 (0.0826) [.946]	-0.0639 (0.0487) [.894]	-0.0858* (0.0511) [.721]	-0.00415 (0.0433) [.997]	-0.00665 (0.0494) [.996]
Mean vocabulary at baseline (TVIP)	0.0135	0.0316	-0.0473	0.0573	0.0519*	0.0548	-0.0119	-0.00869

	(0.0383)	(0.0496)	(0.0373)	(0.0503)	(0.0301)	(0.0382)	(0.0277)	(0.0359)
	[.999]	[.988]	[.935]	[.945]	[.669]	[.845]	[.997]	[.996]
SD vocabulary at baseline (IVIP)	0.0141	0.0212	0.132***	0.212***	-0.0415	0.00307	0.0194	0.0128
	(0.0490)	(0.0584)	(0.0434)	(0.0640)	(0.0381)	(0.0426)	(0.0310)	(0.0381)
	[.999]	[.998]	[.049]	[.037]	[.942]	[1.00]	[.994]	[.996]
Proportion preschool	-0.00582	-0.0140	-0.0812	-0.104	-0.0365	-0.0270	-0.0476	-0.0519
	(0.0778)	(0.0794)	(0.0815)	(0.0783)	(0.0506)	(0.0485)	(0.0660)	(0.0664)
	[1.00]	[.998]	[.971]	[.896]	[.990]	[.998]	[.993]	[.994]
CLASS score	0.104***	0.117***	0.0657	0.0648				
	(0.0329)	(0.0344)	(0.0591)	(0.0595)				
	[.036]	[.029]	[.966]	[.946]				
Observations	454		461		462		470	
SD of classroom value added	.094		.102		.076		.074	
Corrected SD of classroom value added	.076		.089		.057		.055	
P-value joint significance all peer characteristics		.255		.087		.399		.855
P-value joint significance peer characteristics excluding poorly-behaved student dummy		.675		.081		.362		.923
Noise-corrected R-squared		.142		.122		.095		.065
Contribution of poorly-behaved student dummy to noise-corrected R-squared		.030		.028		.004		.005

Notes: This table shows regressions of classroom value added on peer and non-peer classroom characteristics pooling across grades 3 to 6. To derive classroom value added and peer characteristics at the classroom level, we exclude poorly-behaved students from the sample. Column 1 reports estimated coefficients from regressions of classroom value added on each characteristic separately (hence, each coefficient corresponds to a separate regression). Column 2 reports estimated coefficients from a regression of classroom value added on all characteristics included jointly. All regressions control for school-by-grade fixed effects. Column 3 reports the noise-corrected R-squared of the regression to account for measurement error in classroom value added, as well as the proportion of the noise-corrected R-squared explained by the poorly-behaved student dummy alone. In square brackets, we report p-values from the Romano and Wolf (2005) stepdown procedure based on 5000 bootstrap replications.

Table F2: Effect of peer- and non-peer characteristics on classroom value added

	3rd grade		4th grade		5th grade		6th grade		Pooled	
	Separate regressions	Joint regression	Separate regressions	Joint regression	Separate regressions	Joint regression	Separate regressions	Joint regression	Separate regressions	Joint regression
Has poorly-behaved student	-0.0336** (0.0155) [.548]	-0.0442*** (0.0163) [.348]	-0.0330* (0.0173) [.714]	-0.0419** (0.0181) [.604]	-0.00669 (0.0133) [1.00]	-0.00800 (0.0142) [1.00]	-0.0161 (0.0119) [.970]	-0.00752 (0.0130) [1.00]	-0.0222*** (0.00836) [.116]	-0.0218** (0.00872) [.200]
Has low-achieving student	-0.0131 (0.0162) [1.00]	-0.0152 (0.0177) [1.00]	-0.00647 (0.0193) [1.00]	-0.00199 (0.0218) [1.00]	-0.00351 (0.0132) [1.00]	-0.00254 (0.0142) [1.00]	-0.0106 (0.0116) [.999]	-0.00421 (0.0141) [1.00]	-0.00855 (0.00989) [.997]	-0.00775 (0.0106) [1.00]
Math+language index: p10	-0.0144 (0.0214) [1.00]	-0.00447 (0.0368) [1.00]	-0.00170 (0.0227) [1.00]	-0.00199 (0.0365) [1.00]	0.0146 (0.0188) [1.00]	-0.0145 (0.0313) [1.00]	0.0214 (0.0166) [.974]	0.0481 (0.0312) [.940]	0.00249 (0.0130) [1.00]	0.00510 (0.0210) [1.00]
Math+language index: p25	0.0110 (0.0259) [1.00]	0.0188 (0.0439) [1.00]	-0.0436 (0.0377) [.994]	-0.0237 (0.0610) [1.00]	0.00661 (0.0235) [1.00]	-0.0630* (0.0366) [.918]	0.0135 (0.0215) [1.00]	0.00338 (0.0338) [1.00]	0.000950 (0.0181) [1.00]	-0.00819 (0.0255) [1.00]
Math+language index: p50	0.0239 (0.0414) [1.00]	-0.00375 (0.0585) [1.00]	-0.0929** (0.0460) [.645]	-0.111 (0.0772) [.987]	0.0492* (0.0291) [.864]	0.0435 (0.0481) [1.00]	0.0167 (0.0270) [1.00]	0.00258 (0.0442) [1.00]	0.00758 (0.0224) [1.00]	0.0116 (0.0320) [1.00]
Math+language index: p75	0.00790 (0.0446) [1.00]	0.0393 (0.0597) [1.00]	-0.0691 (0.0451) [.930]	-0.00153 (0.0700) [1.00]	0.0376 (0.0303) [.984]	0.0476 (0.0523) [1.00]	0.00966 (0.0260) [1.00]	-0.0176 (0.0479) [1.00]	-0.00423 (0.0232) [1.00]	-0.0117 (0.0337) [1.00]
Math+language index: p90	-0.0185 (0.0408) [1.00]	-0.0822 (0.0548) [.977]	-0.0169 (0.0410) [1.00]	0.0130 (0.0592) [1.00]	0.0264 (0.0236) [.995]	0.0213 (0.0421) [1.00]	0.0101 (0.0254) [1.00]	-0.0118 (0.0409) [1.00]	-0.00119 (0.0187) [1.00]	-0.00544 (0.0271) [1.00]
SD lagged math+language index	0.0369 (0.0573) [1.00]	0.0333 (0.118) [1.00]	-0.0277 (0.0624) [1.00]	-0.130 (0.127) [1.00]	-0.0163 (0.0522) [1.00]	-0.117 (0.120) [1.00]	-0.00866 (0.0477) [1.00]	0.123 (0.0915) [.981]	-0.00372 (0.0352) [1.00]	0.00547 (0.0636) [1.00]
Lagged executive function: p10	-0.0265 (0.0262) [1.00]	-0.0251 (0.0464) [1.00]	-0.0498* (0.0254) [.689]	-0.0577 (0.0443) [.995]	0.0195 (0.0202) [.998]	0.0260 (0.0355) [1.00]				
Lagged executive function: p25	0.00484 (0.0352)	-0.0128 (0.0557)	-0.0289 (0.0363)	0.0493 (0.0548)	0.0304 (0.0277)	0.00691 (0.0492)				

Lagged executive function: p50	[1.00]	[1.00]	[1.00]	[1.00]	[.995]	[1.00]				
	0.0516	0.0563	-0.0201	0.0331	0.0149	-0.0338				
	(0.0397)	(0.0574)	(0.0398)	(0.0576)	(0.0290)	(0.0468)				
Lagged executive function: p75	[.984]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
	0.0416	-0.0444	-0.0389	-0.00179	0.0145	-0.0242				
	(0.0399)	(0.0572)	(0.0359)	(0.0610)	(0.0300)	(0.0431)				
Lagged executive function: p90	[.999]	[1.00]	[.995]	[1.00]	[1.00]	[1.00]				
	0.0523*	0.0610	-0.0203	-0.0437	0.0497**	0.0391				
	(0.0297)	(0.0438)	(0.0277)	(0.0495)	(0.0222)	(0.0373)				
SD lagged executive function	[.832]	[.988]	[1.00]	[1.00]	[.458]	[.999]				
	0.0839	0.0377	0.0297	0.0541	0.0199	0.0407				
	(0.0543)	(0.109)	(0.0606)	(0.135)	(0.0491)	(0.110)				
Mean age	[.930]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]				
	-0.00267	-0.00381	-3.94e-05	-0.00264	-0.00190	-0.00443	-0.00195	-0.00213	-0.00155	-0.00184
	(0.00596)	(0.00653)	(0.00567)	(0.00623)	(0.00447)	(0.00482)	(0.00370)	(0.00391)	(0.00331)	(0.00334)
Proportion female	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	-0.107	-0.0677	-0.276	-0.186	0.154	0.292	-0.391*	-0.407*	-0.153	-0.0750
	(0.198)	(0.217)	(0.286)	(0.324)	(0.191)	(0.205)	(0.203)	(0.223)	(0.146)	(0.151)
Wealth baseline: p10	[1.00]	[1.00]	[.998]	[1.00]	[.999]	[.985]	[.679]	[.828]	[.984]	[1.00]
	-0.0126	0.0117	-0.00480	-0.0110	-0.00723	0.0132	-0.0111	-0.0192	-0.00815	-0.00222
	(0.0318)	(0.0531)	(0.0316)	(0.0438)	(0.0214)	(0.0319)	(0.0224)	(0.0303)	(0.0162)	(0.0229)
Wealth baseline: p25	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	-0.00236	0.00439	0.0108	0.0249	-0.00921	0.00989	0.000894	0.0120	-0.00154	0.00255
	(0.0328)	(0.0560)	(0.0428)	(0.0638)	(0.0246)	(0.0402)	(0.0202)	(0.0394)	(0.0183)	(0.0318)
Wealth baseline: p50	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	-0.00668	-0.00397	0.0438	0.0684	-0.0116	0.0103	-0.0143	-0.0365	-0.00315	0.0121
	(0.0320)	(0.0561)	(0.0387)	(0.0612)	(0.0252)	(0.0463)	(0.0209)	(0.0371)	(0.0186)	(0.0320)
Wealth baseline: p75	[1.00]	[1.00]	[.995]	[.999]	[1.00]	[1.00]	[1.00]	[.999]	[1.00]	[1.00]
	-0.00277	-0.0431	0.00133	-0.0116	-0.0306	-0.0617*	-0.00743	0.0175	-0.0176	-0.0387
	(0.0242)	(0.0416)	(0.0316)	(0.0454)	(0.0202)	(0.0359)	(0.0201)	(0.0355)	(0.0140)	(0.0243)
Wealth baseline: p90	[1.00]	[.999]	[1.00]	[1.00]	[.933]	[.918]	[1.00]	[1.00]	[.975]	[.737]
	0.0371*	0.0698*	-0.00564	-0.0222	-0.0151	-0.0140	-0.00803	0.00861	0.00183	0.0117

	(0.0210)	(0.0379)	(0.0264)	(0.0413)	(0.0153)	(0.0311)	(0.0148)	(0.0293)	(0.0126)	(0.0225)
SD wealth at baseline	[.832]	[.880]	[1.00]	[1.00]	[.998]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	0.0474	-0.0346	-0.000837	0.0526	-0.0125	0.0563	-0.0190	-0.0623	0.00391	0.00914
	(0.0537)	(0.120)	(0.0660)	(0.127)	(0.0420)	(0.0909)	(0.0450)	(0.0893)	(0.0350)	(0.0658)
Proportion mother education less than secondary	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	0.00994	0.0552	0.00791	0.0233	0.0359	0.0451	-0.00226	-0.0388	0.0132	0.00761
	(0.0692)	(0.0854)	(0.0772)	(0.0849)	(0.0529)	(0.0636)	(0.0490)	(0.0662)	(0.0384)	(0.0427)
Proportion father education less than secondary	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	-0.0620	-0.0782	-0.0306	-0.0720	-0.0639	-0.0776	-0.00415	0.000406	-0.0390	-0.0486
	(0.0603)	(0.0693)	(0.0752)	(0.0853)	(0.0487)	(0.0553)	(0.0433)	(0.0515)	(0.0345)	(0.0376)
Vocabulary baseline: p10	[.999]	[.999]	[1.00]	[1.00]	[.978]	[.985]	[1.00]	[1.00]	[.984]	[.950]
	0.0126	0.0385	-0.0350**	0.00634	0.0119	-0.0100	-0.00766	-0.00575	-0.00619	0.00680
	(0.0143)	(0.0329)	(0.0141)	(0.0371)	(0.0130)	(0.0248)	(0.0106)	(0.0241)	(0.00920)	(0.0172)
Vocabulary baseline: p25	[1.00]	[.998]	[.323]	[1.00]	[.998]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	0.00223	0.0204	-0.0296	0.0144	0.0287	0.0149	-0.00382	-0.00460	-0.00454	0.00250
	(0.0202)	(0.0296)	(0.0193)	(0.0369)	(0.0175)	(0.0276)	(0.0136)	(0.0212)	(0.0110)	(0.0167)
Vocabulary baseline: p50	[1.00]	[1.00]	[.930]	[1.00]	[.883]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	0.0186	-0.00375	-0.0107	-0.00646	0.0336	0.00793	0.00100	0.0202	0.00632	0.00551
	(0.0299)	(0.0472)	(0.0289)	(0.0410)	(0.0231)	(0.0341)	(0.0223)	(0.0359)	(0.0174)	(0.0252)
Vocabulary baseline: p75	[1.00]	[1.00]	[1.00]	[1.00]	[.948]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	0.0152	-0.00351	0.0221	0.0378	0.0268	0.0232	-0.0103	-0.0387	0.0114	0.0145
	(0.0395)	(0.0645)	(0.0437)	(0.0602)	(0.0282)	(0.0449)	(0.0262)	(0.0416)	(0.0216)	(0.0330)
Vocabulary baseline: p90	[1.00]	[1.00]	[1.00]	[1.00]	[.998]	[1.00]	[1.00]	[.999]	[1.00]	[1.00]
	-0.0110	-0.0507	0.0284	0.00538	0.0144	0.00648	0.0105	0.0400	0.00360	-0.0254
	(0.0526)	(0.0741)	(0.0573)	(0.0779)	(0.0396)	(0.0529)	(0.0394)	(0.0560)	(0.0315)	(0.0390)
SD vocabulary at baseline (TVIP)	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
	0.0141	0.119	0.132***	0.204*	-0.0415	-0.0398	0.0194	0.00204	0.0343	0.0584
	(0.0490)	(0.110)	(0.0434)	(0.123)	(0.0381)	(0.0738)	(0.0310)	(0.0813)	(0.0250)	(0.0513)
Proportion preschool	[1.00]	[.999]	[.091]	[.949]	[.995]	[1.00]	[1.00]	[1.00]	[.928]	[.991]
	-0.00582	0.0265	-0.0812	-0.102	-0.0365	-0.0362	-0.0476	-0.0604	-0.0434	-0.0401
	(0.0778)	(0.0800)	(0.0815)	(0.0798)	(0.0506)	(0.0531)	(0.0660)	(0.0689)	(0.0425)	(0.0421)
	[1.00]	[1.00]	[.998]	[.996]	[1.00]	[1.00]	[1.00]	[1.00]	[.984]	[.996]

CLASS score	0.104***	0.111***	0.0657	0.0524						
	(0.0329)	(0.0378)	(0.0591)	(0.0603)						
Observations	[.065] 454	[.237] 454	[.995] 461	[1.00] 461	462	462	470	470	1,847	1,847
Variance of classroom value added										
Corrected variance of classroom value added										
P-value joint significance all peer characteristics	.730	.201	.417	.355	.241	.271	.977	.983	.760	.907
P-value joint significance other peer characteristics		.598		.326		.241		.990		
Noise-corrected R-squared	.160	.200	.142	.167	.156	.158	.086	.088	.028	0.991 .040
Contribution of poorly-behaved student dummy to noise-corrected R-squared		.040		.024		.002		.002		.012

Notes: This table shows regressions of classroom value added on peer and non-peer classroom characteristics pooling across grades 3 to 6. To derive classroom value added and peer characteristics at the classroom level, we exclude poorly-behaved students from the sample. Column 1 reports estimated coefficients from regressions of classroom value added on each characteristic separately (hence, each coefficient corresponds to a separate regression). Column 2 reports estimated coefficients from a regression of classroom value added on all characteristics included jointly. All regressions control for school-by-grade fixed effects. Column 2 reports the noise-corrected R-squared of the regression to account for measurement error in classroom value added, as well as the proportion of the noise-corrected R-squared explained by the poorly-behaved student dummy alone. In square brackets, we report p-values from the Romano and Wolf (2005) stepdown procedure based on 5000 bootstrap replications.

Table F3: Within-school variation of classroom peer characteristics

	Grade			
	3	4	5	6
Has poorly-behaved student	0.549	0.511	0.544	0.546
Has low-achieving student	0.497	0.451	0.453	0.492
Mean lagged math+language index	0.124	0.099	0.122	0.147
SD lagged math+language index	0.293	0.333	0.374	0.422
Mean lagged executive function	0.172	0.218	0.190	
SD lagged executive function	0.447	0.520	0.466	
Mean age	0.198	0.235	0.272	0.304
Proportion female	0.124	0.086	0.086	0.100
Mean wealth at baseline	0.067	0.060	0.078	0.089
SD wealth at baseline	0.329	0.315	0.318	0.33
Proportion mother education less than secondary	0.196	0.208	0.228	0.217
Proportion father education less than secondary	0.262	0.236	0.295	0.305
Mean vocabulary at baseline (TVIP)	0.167	0.193	0.182	0.189
SD vocabulary at baseline (TVIP)	0.340	0.364	0.325	0.372
Proportion preschool	0.202	0.213	0.215	0.218
CLASS score	0.394	0.434		

Notes: This table shows 1-R2 from regressions of classroom-level peer characteristics on school fixed effects.

Table F4: Within school correlations of classroom peer and non-peer characteristics

	Has poorly-behaved student	Has low-achieving student	Mean math+language index	SD math+language index	Mean EF	SD EF	Mean age	Proportion female	Mean wealth at baseline	SD wealth at baseline	Proportion mother education less than secondary	Proportion father education less than secondary	Mean vocabulary at baseline	SD vocabulary at baseline	Proportion preschool	CLASS score
Has low-achieving student	0.063*	1														
Mean math+language index	0.001	-0.185*	1													
SD math+language index	0.002	0.158*	-0.300*	1												
Mean EF	0.019	-0.129*	0.603*	-0.197*	1											
SD EF	0.049	0.066*	-0.124*	0.446*	-0.116*	1										
Mean age	-0.002	-0.023	-0.104*	0.114*	-0.035	0.108*	1									
Proportion female	0.143*	-0.024	0.038	-0.01	0.043	0.023	0.027	1								
Mean wealth at baseline	0.004	-0.022	0.170*	-0.047*	0.124*	-0.006	-0.022	0.035	1							
SD wealth at baseline	-0.016	-0.041	0.019	0.043	0.012	0.025	-0.013	-0.023	0.190*	1						
Proportion mother education less than secondary	0.04	0.001	-0.140*	0.021	-0.083*	0.011	0.041	0.015	-0.498*	-0.104*	1					
Proportion father education less than secondary	0.022	0.003	-0.100*	-0.033	-0.059*	-0.057*	-0.018	0.033	-0.324*	-0.018	0.289*	1				
Mean vocabulary at baseline	-0.055*	-0.017	0.270*	-0.072*	0.221*	-0.060*	0.116*	0.015	0.212*	-0.007	-0.136*	-0.130*	1			
SD vocabulary at baseline	0.050*	0.047*	-0.123*	0.115*	-0.121*	0.083*	-0.071*	-0.052*	-0.142*	-0.019	0.052*	0.093*	-0.567*	1		
Proportion preschool	-0.006	-0.026	0.056*	-0.028	0.041	-0.064*	-0.017	-0.023	0.018	-0.036	-0.102*	0.025	0.067*	-0.021	1	
CLASS score	0.026	0.05	0.009	0.017	0.001	0.071*	-0.053	0.031	-0.023	0.003	-0.008	0.095*	-0.063	0.065*	0.005	1

Notes: This table shows within-school correlations between peer and non-peer characteristics at the classroom level, pooling information across grades 3 to 6. First, we regress each classroom-level variable on school (by grade) fixed effects. For each regression we derive the residual. We show correlations among the residuals of these regressions.

Appendix G

Table G1: Effect of poorly-behaved students on student achievement

Panel A	Lag				
	0	1	2	3	Joint test 1
Has poorly-behaved student	-0.0218*** (0.00504)	-0.0162*** (0.00506)	-0.0206*** (0.00681)	0.000964 (0.0107)	.086
Panel B	Lag				
	0	1	2	3	Joint test 2
Has poorly-behaved student 3rd grade	-0.0374*** (0.0114)	-0.0191* (0.0100)	-0.0134 (0.0106)	0.000863 (0.0107)	.013
Has poorly-behaved student 4th grade	-0.0297** (0.0133)	-0.0196** (0.00884)	-0.0278*** (0.00909)		.467
Has poorly-behaved student 5th grade	-0.00711 (0.00999)	-0.0105 (0.00833)			.690
Has poorly-behaved student 6th grade	-0.0137 (0.00936)				
Joint test 3	.191				
Joint test 4	.043				

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 a classroom with a poorly-behaved student excluding poorly-behaved students from the sample, pooling information across grades. 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, child age and gender, and school by grade fixed effects. Joint test 1 shows the p-value for a joint test of equality of estimated coefficients for different lags of exposure to a poorly-behaved student. Standard errors are clustered at the classroom and student level. 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 for various lags of year of assignment to a classroom with a poorly-behaved student excluding poorly-behaved students from the sample, for various grades. 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, child age and gender, and school fixed effects. Joint test 2 shows the p-value for a joint test of equality of estimated coefficients for different lags of exposure to a poorly-behaved student. Joint test 3 shows the p-value for a joint test of equality of estimated coefficients for contemporaneous effects across grades 3 to 6. Joint test 4 shows the p-value for a joint test of equality of average contemporaneous effects in grades 3 and 4 and in grades 5 and 6. Standard errors are clustered at the classroom level.

Table G2: Dynamic effects of poorly-behaved students on student achievement

Panel A	(1)	(2)		
Has poorly behaved student in g	-0.0224*** (0.00504)	-0.0190** (0.00911)		
Has poorly behaved student in $g-1$		-0.0192** (0.00900)		
Has poorly behaved student in g * Has poorly behaved student in $g-1$		0.00825 (0.0120)		
Has poorly behaved student in g * Lagged ability	0.00764* (0.00444)			
Panel B	Once	Twice	Three times	Four times
Has poorly behaved student	-0.0245 (0.0215)	-0.0478** (0.0210)	-0.0624*** (0.0233)	-0.0458* (0.0268)

Notes: This table reports estimates of dynamic effects of poorly-behaved students on their peer's math and language achievement excluding poorly-behaved students from the sample. 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 an interaction between the indicator and lagged ability. The regression controls for a fourth-order polynomial in lagged achievement, child age and gender, and school-by-grade fixed effects. Standard errors are clustered at the student and classroom level. Panel A, 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 $g-1$, and an interaction between the two indicators, pooling information across grades. The regression controls for a fourth-order polynomial in twice lagged achievement, 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 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, excluding poorly-behaved students from the sample. The regression controls for a fourth-order polynomial in achievement at the end of second grade, child age and gender. Standard errors are clustered at the classroom level.

Table G3: Dynamic effects of poorly-behaved students on student achievement

	(1)	(2)	(3)
Has poorly behaved student in g	-.014* (.008)	-.015** (.007)	-.018** (.008)
Has poorly behaved student in $g-1$	-.014** (.006)	.000 (.005)	-0.003 (.007)
Has poorly behaved student in g * Has poorly behaved student in $g-1$			0.005 (.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 $g-1$, 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 $g-1$, 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 $g-1$, 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 .