The effect of classroom rank on learning throughout elementary school: experimental evidence from Ecuador

Pedro Carneiro
University College London, IFS, CEMMAP, FAIR-NHH

Yyannu Cruz Aguayo
Inter-American Development Bank

Francesca Salvati
Bank of Italy, IFS, IZA

Norbert Schady
World Bank

March 7, 2022[1]

Abstract

We study the impact on learning of a child's rank in the classroom, using a unique experiment from Ecuador. Within each school, students were randomly assigned to classrooms in every grade between kindergarten and 6th grade. Therefore, two students with the same ability can have different classroom ranks because of the (random) peer composition of their classroom. In order to isolate the impact of rank from other peer influences we include flexible controls for average peer quality, as well as classroom fixed effects. We find that children with higher classroom rank at the beginning of the academic year have significantly higher math test scores at the end of that grade. Classroom rank in math, not language, drives our results. The impact of classroom math rank is larger for younger children, and grows substantially over time. Exogenous changes in classroom rank in math also improve executive function, child happiness, and teacher perceptions of students.

[1] We thank Carolina Echeverri for outstanding research assistance. We also thank Richard Murphy, and seminar participants at the 6th IZA Economics of Education workshop for valuable comments and feedback.

1. Introduction

There are many settings in economics where relative rank concerns are important. They emerge naturally in tournaments (Lazear and Rosen 1981), they affect job satisfaction (Card et al. 2012), and can interact in interesting ways with social preferences (Bandiera et al. 2005). Although the idea that rank is important in education dates back to Marsh (1987), recent years have seen a growing literature quantifying its importance (including Elsner and Isphording 2017; Elsner et al. 2018; Tincani 2018; Denning, Murphy, and Weinhardt 2018, Murphy and Weinhardt 2020; Delaney and Devereux 2021), and showing medium- and long-term impacts on college major choice and attendance, earnings, and the probability of engaging in risky behaviors.[2]

In this paper, we investigate the impact of classroom rank on learning and other outcomes throughout elementary school. We use a unique experiment in elementary schools in Ecuador where, at the start of every grade, a cohort of students was randomly assigned across classrooms within a given school. Compliance with the random assignment was almost perfect, 98.9 percent on average over the 7 years of the experiment. Variation in peer groups resulting from randomly assigning students to classrooms means that two students with the same underlying ability and in the same school and grade will have different classroom ranks.

This experimental design generates substantial variation in classroom rank achievement, which we measure in percentiles (so rank is normalized to be between 0 and 1). For example, for individuals with median ability in their school (i.e., with a rank close to 0.5 within their school), classroom rank can be as low as 0.364 or as high as 0.593 depending on the classroom they are assigned to.[3]

We find that increasing a child's classroom rank at the start of a given grade, keeping own ability constant, raises end-of-grade achievement. Rank, however, is just a particular form of peer effects, which could influence outcomes in various ways. The experimental design we explore generates variation in other classroom-level characteristics that may affect learning. In order to isolate the impact of classroom rank from other peer influences, we estimate models that control for average peer quality. We show that, because rank and peer quality are negatively related, estimates of the effect of classroom rank

[2] Existing papers show causal effects of rank on many domains. Murphy and Weinhardt (2020) and Cicala et al. (2018) show that rank can have positive effects on test scores in primary and secondary school. Elsner and Isphording (2017) and Elsner et al. (2018) document positive impacts of rank in high school on college enrollment and choice. Delaney and Devereux (2021) show that rank can partly explain the gender gap in the choice of STEM as a college major. In recent work, Denning et al. (2020) show that rank can also have long-term impacts on earnings later in life. Rank has also been shown to affect the likelihood of engaging in risky behaviors, as in Elsner and Isphording (2018).

[3] Even for very high and very low levels of student ability, for which variation in rank is bound to be more limited, there can be rank differences of 0.1 (10 percentile points) or more depending on the classroom one is assigned to. This alleviates the type of concerns raised by Angrist (2014), that studies exploring experimental assignment of children to peer groups may generate weak variation in peer characteristics, making it difficult to study peer effects in those settings.

increase when we control for peer quality using a standard linear-in-means model. Our results are robust to more flexible ways of incorporating peer quality and other classroom-level characteristics, such as by using classroom fixed effects.

We show that the effects of classroom rank are entirely driven by classroom rank in math, rather than language. Focusing on classroom rank in math, we document that rank effects are concentrated in the upper half of the rank distribution. Consistent with this result, we also find that rank effects are larger for children with higher lagged achievement, and for children who started kindergarten with higher levels of vocabulary. On the other hand, we see no evidence that the effect of classroom rank varies with the gender of the child.

A novelty of our paper is the focus on differences in the effects of classroom rank by grade, and how these evolve over time. For this purpose, we divide our sample into children in "early" grades (1st and 2nd grades), "middle" grades (3rd and 4th grades) and "late" grades (5th and 6th grades) in elementary school. We first show that classroom rank effects are largest in the early grades: moving a child from the 50th to the 75th percentile of classroom rank in 1st or 2nd grade increases her end-of-grade achievement by 1 percentile point of national achievement, while classroom rank in 5th and 6th grades has no effect on achievement. We reject the null that rank effects are the same in early, middle, and late grades (p-value<0.01).

Our analysis then turns to the evolution of classroom rank effects over time. We show, remarkably, that the effects of early classroom math rank increase substantially as children age. After 4 lags the effect of classroom rank in 1st or 2nd grade is almost twice as large as it was originally. This result stands in sharp contrast with the effects of many other determinants of achievement, which tend to fade out.[4]

The fact that the medium-term effects of early classroom rank in math are larger than the short-term effects could occur for a number of reasons. One possibility is that students randomly assigned a high classroom rank in 1st (or 2nd) grade benefit from a virtuous cycle, where a high rank in grade t leads to higher learning and higher rank in grade $t + 1$, which in turn lead to higher learning and higher ranks in the following grades. We find, however, that our estimates are more consistent with an alternative model where, in addition to its direct effect on achievement, early math rank affects an unobserved trait—for example, academic self-concept—which affects learning in later grades, and which does not depreciate over time.

[4] See, for example, Chetty, Friedman, and Rockoff (2014), and Jacob, Lefgren, and Sims (2010) for estimates of the fade-out of the effects of teacher quality, measured by teacher value added. For a review and discussion of fade-out in education interventions, with a particular focus on early childhood, see Bailey et al. (2020). For a recent example from a setting similar to ours, see Barrera-Osorio et al. (2020) on the fade-out of information on student performance provided to parents in Colombia.

With this insight, we turn to other child outcomes. Between 1st and 4th grades, we collected data on child executive function (EF). 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). Executive function in childhood has 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). We show that randomly assigning a child to a classroom where her rank is higher during the early grades improves her executive function in the medium term.

We also have data on a number of non-cognitive outcomes. At the end of 1st grade, we asked children whether they were happy in school, and at the end of 6th grade, we collected data on child depression, self-esteem, grit, and growth mindset. We show that children with higher math classroom rank (at the start of 1st grade) are more likely to say they are always happy (at the end of that grade). However, we do not find significant impacts of classroom rank on non-cognitive skills measured at the end of 6th grade.

Finally, we analyze whether classroom rank affects teacher perceptions of whether each student ranks at the top or bottom of the class. These perceptions are correlated with actual student ranks in test scores, but the correlation is not perfect, either because of measurement error, or because teacher perceptions reflect other student attributes beyond their performance on the test. Teacher perceptions of student ability or performance could be important if, for example, teachers pay more attention to children they consider higher-achieving—and if such teacher reactions, in turn, affect future achievement. We show that, controlling for own lagged achievement, students who have a higher classroom rank in one grade are more likely to be perceived to be at the top of the class (and less likely to be thought to be at the bottom of the class) by their teachers in subsequent grades.

In sum, children with higher classroom rank have higher end-of-grade achievement. The effects of early classroom rank in math in 1st and 2nd grade increase substantially over time. More highly-ranked children also have higher executive function, exhibit higher levels of happiness, and are thought to be higher-achieving by their future teachers.

Our paper extends the economics literature on the impacts of rank in important ways. First, ours is the first paper that analyzes rank effects in education using an experiment with multiple rounds of random assignment, with essentially perfect compliance. The fact that we have data on children randomly assigned to different classrooms in every grade in elementary school means that we can convincingly test whether classroom rank has larger effects in some grades than in others. Moreover, because we follow children over the entire elementary school cycle, we can credibly estimate how the effects of classroom rank evolve over time. As we show, both of these considerations—differences across grades and changes in the impact of classroom rank over time—are important, at least in the setting that we study.

In addition, because we have unusually rich data on teacher perceptions of student ability, executive function, happiness, depression, self-esteem, grit, growth mindset, we can study how different inputs affect the formation of multiple skills in elementary school.

The rest of the paper proceeds as follows. In section 2 we describe the setting and data, in section 3 we discuss our empirical strategy. Results are in section 4, and we conclude in section 5.

2. Setting and data

A. Experimental setting and descriptives

We study the acquisition of math, language, executive function, and non-cognitive skills in Ecuador, a middle-income country in South America. As is the case in most other Latin American countries, educational achievement of young children in Ecuador is low (Berlinski and Schady 2015).

The data we use comes from an experiment in 202 schools.[5] Schools have at least two classrooms per grade (most have exactly two). An incoming cohort of children was randomly assigned to kindergarten classrooms within schools in the 2012 school year. These children were reassigned to 1st grade classrooms in 2013, to 2nd grade classrooms in 2014, to 3rd grade classrooms in 2015, to 4th grade classrooms in 2016, to 5th grade classrooms in 2017, and to 6th grade classrooms in 2018. Compliance with the assignment rules was very high—98.9 percent on average.

Random assignment means that we can effectively deal with concerns about any purposeful matching of students with teachers and peers that often arise in non-experimental settings. 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. Table 1, Panel A, provides summary statistics for children in our sample. The table shows that children were 5 years of age on the first day of kindergarten, and half of them are girls. Mothers were in their early 30s and fathers in their mid-30s. Both parents had on average just under 9 years of schooling, which corresponds to completed middle school. 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 sample for the test.[6]

[5] 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.

[6] 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 was normed on samples of Mexican and Puerto Rican children. It has been used widely to measure development among Latin American children.

Table 1, Panel B, summarizes characteristics of classrooms and teachers. Average class size is 36. The average teacher in the sample has 18 years of experience. Eighty-two percent of teachers are women, and 82 percent are tenured. An important consideration in interpreting our results is that there is always one teacher per classroom, without a classroom aide, and that the same teacher is responsible for all academic subjects (all subjects other than physical education and, when they are available, art and music).

We collected data on math and language achievement 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.[7] 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.[8]

In every grade between kindergarten and 4th grade, we tested child executive function. EF includes a set of basic self-regulatory skills which involve various parts of the brain, but in particular the prefrontal cortex.[9] Executive function is generally thought of as having three domains: working memory, inhibitory control, and cognitive flexibility. It is an important determinant of how well young children adapt to and learn in school. Basic EF skills are needed to pay attention to a teacher; wait to take a turn or raise one's hand to ask a question; and remember steps in, and shift from one approach to another, when solving a math problem, among many other tasks that children are expected to learn and carry out in the classroom. Children with high EF levels are able to concentrate, stay on task, focus, be goal-directed, and make good use of learning opportunities. Low levels of EF are associated with low levels of self-control and "externalizing" behavior, including disruptive behavior, aggression, and inability to sit still and pay

---

See Paxson and Schady (2007) for a comparison of vocabulary scores between children in Ecuador and the U.S., and Schady et al. (2015) for evidence on levels and socioeconomic gradients in the TVIP in five Latin American countries, including Ecuador.

[7] The number line task works as follows. Children are shown a line with the two clearly marked endpoints—for example, in 1st grade, the left end of the line is marked with a 0, and the right end is marked with a 20. They are then asked to place various numbers on the line—for example, the number 2 or the number 18. The accuracy with which children place the numbers has been shown to predict general math achievement (see Siegler and Booth 2004).

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

[9] Volumetric measures of prefrontal cortex size predict executive function skills; children and adults experiencing traumatic damage to the prefrontal cortex sustain immediate (and frequently irreversible) deficits in EF (Nelson and Sheridan 2011, cited in Obradovic et al. 2012).

attention, which affects a child's own ability to learn, as well as that of her classmates (Séguin and Zelazo 2005).[10]

At the end of each grade, we asked teachers who were the 5 children with the highest achievement, and 5 with the lowest achievement.[11] In 1st grade, we asked children whether they were happy in school and in their classroom (two separate questions). In both cases, children had the option of answering "always", "sometimes", or "never". Most children in the sample answered "always" to both questions, so we use their responses to construct a single variable for children who were always happy, almost always happy, or mostly happy.

In 6th grade, we collected data on child depression, self-esteem, growth mindset, and grit. To measure child depression, we used the Patient-Reported Outcomes Measurement Information System (PROMIS) Depression Scale for children aged 11-17 years, developed by the American Psychiatric Association.[12] To measure self-esteem, we selected 5 questions from the National Longitudinal Study of Adolescent to Adult Health (Add Health).[13] To measure Growth Mindset, we selected 10 of the 20 questions on the Dweck "Mindset Quiz"; growth mindset refers to the belief that intelligence is malleable, rather than fixed, and can be increased with effort (Blackwell, Trzesniewski, and Dweck 2007; Dweck 2008). Finally, to measure grit, we adapted 4 questions from the 8-item Grit Scale for children (Duckworth and Quinn 2009); grit refers to the capacity of individuals to persevere at a given task. For each of these 6th grade outcomes, we aggregated responses by factor analysis. We also calculate an overall non-cognitive score that gives the same weight to each of the individual tests.

Most of the tests were applied to children individually (as opposed to in a group setting) by specially trained enumerators.[14] All tests, other than the non-cognitive tests applied in 6th grade, were

[10] Working memory measures the ability to retain and manipulate information; for example, 2nd grade child were asked to remember (increasingly long) strings of numbers and repeat them in order and then backwards. Cognitive flexibility measures the ability to shift attention between tasks and adapt to different rules; for example, 1st grade children were shown picture cards that had trucks or stars, red or blue, and were asked to first sort cards by *shape* (trucks versus stars), and then by *color* (red versus blue). Inhibitory control refers to the capacity to suppress impulsive responses; for example, kindergarten children were quickly shown a series of 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. We calculate scores on each of the three domains in executive function, as well as a measure of overall EF scores that gives one-third of the weight to each individual domain. We do not have data on inhibitory control in 1st grade. In this grade, the overall measure of executive function includes only cognitive flexibility and working memory, with equal weight given to both.

[11] Importantly, these data are not disaggregated by subject—that is, we did not ask teachers who they thought were top and bottom performers in math and, separately, in language.

[12] Olino et al (2013), Klein et al. (2005), and Aylward et al. (2008) argue that the PROMIS depression scale has superior qualities (greater precision, more internal reliability, and more discriminant validity) than other commonly-used depression scales, including the Beck Depression Inventory (BDI), the Children's Depression Inventory (CDI), and the Center for Epidemiologic Studies-Depression (CES-D) scale.

[13] See Harris and Udry 2018 for a description of the Add Health data. The questions on self-esteem in Add Health build on the much-used Rosenberg Self-Esteem Scale (Rosenberg 1989).

[14] The only exception is some of the achievement tests in 4th through 6th grades, which were applied in a group setting.

applied in school. For all tests, to choose questions, we piloted the test; made changes as necessary; and selected questions that could be understood by children in our context, and which showed reasonable levels of variability in the pilot. Further details on child assessments are provided in Appendix B.

B. How much variation is there across classroom rank within school?

A potential concern is that, empirically, there is limited variation in classroom rank across students with the same ability who are assigned to different classrooms as a result of the randomization. As illustrated in Angrist (2014), experimental designs that rely on random assignment of individuals to groups to identify peer effects may suffer from modest variation in peer characteristics, resulting in identification issues that are akin to a weak instruments problem. This empirical concern potentially applies to our setting, as we rely on random assignment of students to classrooms within schools to identify the causal effect of classroom rank on learning.

We show however that, in practice, our empirical work does not suffer from this problem. This problem is more likely to emerge in linear in means models that regress student outcomes on average peer characteristics. It is less important for examining the role of rank, which depends on the whole distribution of student characteristics in the classroom, and not only on its mean. In our dataset, two students with the same ability rank who attend the same school have quite different classroom ranks.

We can graphically illustrate this point by plotting the distribution of classroom rank for students with the same level of ability. Figure 1, Panel A, relies on 1st grade data, and shows quantiles of the distribution of classroom rank (y-axis) for students with a given school rank (x-axis). In practice, to construct this figure, we plot different percentiles of classroom rank within each school rank ventile.

The figure shows that two 1st grade students with the same ability in the same school (and therefore, the same school rank), can have very different classroom ranks. For example, within the 10th ventile of school rank (with median value equal to 0.475), classroom rank varies between 0.364 (5th percentile) and 0.593 (95th percentile).

Another way to illustrate the same point is to plot percentiles of classroom rank against the residuals of a regression of national percentile rank (our measure of underlying ability) on school fixed effects. This is another way of normalizing ability within school (in Panel A, this was done by using school rank instead of actual test scores). The resulting figure is shown in Figure 1, Panel B, where we plot different percentiles of classroom rank within residualised national rank ventiles. Within the 10th ventile of residualised national rank, classroom rank varies between 0.324 (5th percentile) and 0.636 (95th percentile). A similar pattern is shown in Appendix Figures C1 and C2 for all other grades.

3. Empirical strategy

A. Main model

Our goal is to estimate the impact of child $i's$ classroom rank on her subsequent learning in elementary school. The dataset we use allows us to construct measures of classroom rank, lagged achievement, and achievement at the end of the current grade for children between 1st and 6th grades. With these data, we can investigate the impact of classroom rank in the short- and medium-run, starting as early as 1st grade.[15]

To begin our discussion, we focus on two questions. First, how should we measure rank, and which measure of rank is likely to be more relevant for future learning? Second, how do we identify the causal impact of rank on learning? The two issues are interlinked in our setting, so we discuss them together.

In the experiment we study, in each school, children were randomly assigned to classrooms at the start of every grade. As a result, every student has a randomly-assigned set of peers in each grade, so two students with the same underlying ability can nevertheless have different classroom ranks. Our study exploits this variation by estimating the impact of rank at the beginning of grade t on learning at the end of grade t, as well as on learning in subsequent grades (until the end of 6th grade). Achievement at the start of grade t is measured using tests administered at the end of grade $t-1$. Beginning-of-grade classroom rank for each student is based on her achievement at the end of $t-1$ and the achievement of her randomly-assigned classmates. From now on, we refer to this measure simply as classroom rank.[16]

Our approach assumes that students and teachers react to the beginning-of-grade student classroom rank. This makes most sense if they can perceive their rank, and act on it, fairly early during the school year. Furthermore, although school rank may also be important, we are not able to assess its impact as convincingly, since random assignment happens within schools.

Throughout the paper we denote $Y_{i,s,c,t}$ as student $i's$ performance (measured by an index of math and language scores), at the end of grade t, in school s, and classroom c. To be consistent with the literature, we define $Y_{i,s,c,t}$ in terms of percentiles of national rank.[17] $CR_{i,s,c,t}$ is student $i's$ classroom rank at the start of grade t, when the student is randomly assigned to classroom c. In our simplest model, we pool observations from all grades and estimate:

---

[15] We can construct measures of end-of-grade math and language achievement at the end, but not the beginning, of kindergarten. For this reason, our analysis focuses on 1st through 6th grades.

[16] One issue we face is that we only observe end of grade $t-1$ scores for students who were in a school in our sample in grade $t-1$, which means that we do not know what these test scores are for students who arrived at the school in grade t. Therefore, we cannot compute grade t ranks for these students, and our measures of ranks for all other children ignore the fact that new entrants are in their classroom. This will introduce random measurement error in rank.

[17] Percentile ranks normalize automatically by classroom (or school) size. The disadvantage of this is that being the first or last in a small or large classroom may mean something very different. In practice, however, we find no evidence that classroom rank measured in percentiles has different effects in smaller and larger classrooms.

$$Y_{i,s,c,t} = \beta CR_{i,s,c,t} + g_t(Y_{i,s,c,t-1}) + \delta_{st} + \varepsilon_{i,s,c,t} \quad (3.1)$$

where δ_{st} is a school (by grade) fixed effect and $\varepsilon_{i,s,c,t}$ 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. $g_t(Y_{i,s,c,t-1})$ is a third-order polynomial in $Y_{i,s,c,t-1}$.[18] We also estimate models in which, instead of pooling data for all six grades, we estimate separate coefficients on β for 1st and 2nd grades ("early" grades), 3rd and 4th grades ("middle" grades) and 5th and 6th grades ("late" grades).[19]

Next, we separately analyze the effects of classroom rank in math and language:

$$Y^k_{i,s,c,t} = \beta^k CR^k_{i,s,c,t} + g_t(Y^k_{i,s,c,t-1}) + \delta^k_{st} + \varepsilon^k_{i,s,c,t} \quad (3.2)$$

where the k superscript refers to a subject, math or language.

Up to this point, we have assumed that the effect of classroom rank on learning is linear in classroom rank. It is quite possible that this is not the case—it may be, for example, that rank has a different effect at the top and bottom of the (rank) distribution. Therefore, we also consider a version of equation (3.1) where the effect of classroom rank on learning is not constrained to be linear:

$$Y_{i,s,c,t} = \beta(CR_{i,s,c,t}) + g_t(Y_{i,s,c,t-1}) + \delta_{st} + \varepsilon_{i,s,c,t} \quad (3.3)$$

where $\beta(CR_{i,s,c,t})$ is now a flexible function of $CR_{i,s,c,t}$. In our preferred specification we discretize $CR_{i,s,c,t}$ into q values (forcing it to take $q = 10$ values, corresponding to 10 deciles of the distribution), so $\beta(CR_{i,s,c,t}) = \sum_{q=1}^{10} \beta_q CR_{i,s,c,t,q}$ (where $CR_{i,s,c,t,q}$ is an indicator variable that takes value 1 if $CR_{i,s,c,t}$ is in decile q).

B. Alternative models

Other papers in this literature use different measures of rank and different specifications. A leading example is Murphy and Weinhardt (2020), who study the impact of school rank at the end of elementary school on learning in secondary school. In contrast, we study the impact of classroom rank at the beginning of a grade on learning occurring in that grade. Because in Murphy and Weinhardt (2020) rank is measured at the end of elementary school, it is a result of a student's position relative to her peers, but also of the

[18] As Murphy and Weinhardt (2020) emphasize, it is important to use flexible specifications for this function, to avoid the potential problem that β does not capture a true rank effect, but is instead an artefact of the misspecification of this function. Our robustness checks show that considering polynomials in lagged scores of order higher than three does not lead to substantial changes in the results. For this reason, in our main empirical specification $g_t(Y_{i,s,c,t-1})$ is a cubic polynomial in its argument.

[19] In Appendix D we also present estimates where β varies by grade in an unrestricted way, and therefore it is also indexed by t. Those estimates are noisier than the ones we focus on in the paper.

student's reaction to her peers and any subsequent feedback, as well as other school shocks occurring between the beginning and the end of elementary school.[20] If we were to use the Murphy and Weinhardt (2020) specification instead of ours (and using classroom rank instead of school rank) we would estimate:

$$Y_{i,s,c,t} = \beta CR'_{i,s,c,t-1} + g_t\left(Y_{i,s,c,t-1}\right) + \delta_{st} + \varepsilon_{i,s,c,t} \quad (3.4)$$

where $CR'_{i,s,c,t-1}$ is the classroom rank at the end of grade $t-1$, computed using scores at the end of $t-1$ relative to peers in $t-1$.

The main reason why we use equation (3.1), rather than (3.4), as our preferred specification is that it follows directly from our experimental design. Classroom ranks at the start of a grade are randomly assigned and cannot be modified by student effort or other unobserved shocks, whereas classroom ranks at the end of a grade are both a result of random assignment of peers, student effort, peer effort, and potentially even the responses of teachers and parents. Also, our approach estimates the effects of classroom rank experienced in the same year as we measure learning, which is arguably more relevant in the short run for a "big fish little pond" mechanism (as in Marsh 1987).[21] That said, students may not know their rank at the start of the grade, and may take some time to learn about it. In contrast, in Murphy and Weinhardt (2020) students are more likely to have a reasonable perception of their rank since it is measured at the end of elementary school. Therefore, we present results from estimating (3.4) as a robustness check.

C. Peer effects

By construction, the quality of a student's peers is negatively correlated with her rank. Non-parametrically it is impossible to distinguish the impact of rank from that of other impacts of peers. This is a concern faced by every paper focused on rank effects, which are a particular form of peer effects.

However, there are some forms of peer effects that can be distinguished from rank effects. For example, in many empirical peer effects papers, a student's outcome depends on the average ability of her

[20] Murphy and Weinhardt (2020) document the impact of this measure on future learning, when a student moves to another school and experiences a different set of peers. In their model students are motivated to work hard in secondary school because of the rank they experienced and perceived in their past school, as opposed to their rank in the current school.

[21] That said, one advantage of using end-of-classroom rank in our setting is that we can construct rank using everyone in the classroom at the end of the previous grade, while with our preferred measure of beginning of grade rank excludes new school entrants for whom we do not have test scores at the end of the previous grade, introducing measurement error in our preferred measure of classroom rank (on average, 7.1 percent of students are new entrants in each classroom). However, as we show in the robustness checks below, our results are robust to a standard multiple imputation procedure for missing data.

peers (also known as the linear-in-means model).[22] We augment our specification to include average peer ability:

$$Y_{i,s,c,t} = \beta CR_{i,s,c,t} + g_t\big(Y_{i,s,c,t-1}\big) + \theta \bar{Y}_{i,s,c,t-1} + \delta_{st} + \varepsilon_{i,s,c,t} \tag{3.5}$$

where $\bar{Y}_{i,s,c,t-1}$ is average peer ability in the classroom at the beginning of grade t, based on end of grade $t-1$ test scores (using leave-one-out estimates, as is standard in this literature; see, for example, Duflo et al. 2011).

One can also allow for more general peer influences as long they have the same impact on all individuals in the same classroom, by including classroom fixed effects in the model:

$$Y_{i,s,c,t} = \beta CR_{i,s,c,t} + g_t\big(Y_{i,s,c,t-1}\big) + \delta_{sct} + \varepsilon_{i,s,c,t} \tag{3.6}$$

where δ_{sct} are classroom (by grade) fixed effects. This specification also allows us to address any concerns related to the fact that our experimental design generates variation in multiple factors (other than classroom rank) that could affect learning. This specification is analogous to the main approach in Murphy and Weinhardt (2020). As explained in their paper, there is variation in rank across individuals assigned to different classrooms that one can exploit even after accounting for classroom fixed effects, due to differences in the distribution of ability across classrooms.[23] Equations (3.5) and (3.6) are estimated below.

D. Dynamics

To estimate the dynamics of rank effects, we begin with specifications of the following form:

$$Y_{i,s,c,t+l} = \beta_{t,l}\, CR_{i,s,c,t} + g_{t+l}\big(Y_{i,s,c,t-1}\big) + \delta_{st+l} + \varepsilon_{i,s,c,t+l} \tag{3.7}$$

We estimate these regressions separately for "early", "middle", and "late" grades, as discussed above. When $l = 0$, equation (3.7) is equivalent to (3.1), and provides estimates of the short-run impact of classroom rank at the start of grade t, $CR_{i,s,c,t}$, on learning at the end of that same grade, $Y_{i,s,c,t}$. We label this effect $\beta_{t,0}$. When $l > 0$, equation (3.7) provides estimates of the medium-term effect of classroom rank at various lags, which we label $\beta_{t,l}$.

[22] Some examples include Duflo et al. (2011), Booij et al. (2017), Feld and Zölitz (2017). See Epple and Romano (2011) and Ioannides (2011) for recent surveys of the peer effects literature.

[23] Note that this is not equivalent to exploring within-classroom variation. In each classroom, individual ability and classroom rank are perfectly correlated, so one would not be able to estimate this model classroom by classroom (allowing for classroom-specific parameters). The model is identified because it imposes that the impact of classroom rank is the same across classrooms. With this assumption, it would be identified even if we allowed for some restricted forms of interactive fixed effects, as in, for example, Bai (2009).

$\beta_{t,l}$ (medium-run impact) and $\beta_{t,0}$ (short-run impact) are related through three main channels: (i) class rank in grade t affects learning at the end of that grade, and therefore also affects student achievement in grade $t+1$ and in subsequent grades, through the function $g_{t+1}(Y_{i,s,c,t})$ in the $(t+1)$ version of equation (3.1); (ii) since learning at the end of grade t is affected, classroom rank in $t+1$ and subsequent grades is also affected, which can have a further impact on learning in those grades, captured by $\beta_{t+1,0}$ in equation (3.1); (iii) in addition, class rank in grade t may affect other skills not captured by our tests at the end of that grade (unobserved skills), but which nevertheless can affect learning in grades $t+1$ and beyond.

To quantify the importance of these channels, we first estimate an additional equation relating classroom rank at the beginning of grade $t+1$ with learning at the end of grade t, which we will assume can be approximated by the following linear relationship:

$$CR_{i,s,c,t+1} = \delta_{t+1} + \gamma_{t+1} Y_{i,s,c,t} + \tau_{st+1} + \eta_{i,s,c,t+1} \tag{3.8}$$

where τ_{st+1} is a school fixed effect, and $\eta_{i,s,c,t+1}$ is the variation coming from random assignment of students to different peer groups. In practice, as we show below, when $Y_{i,s,c,t}$ is the national percentile rank, $\gamma_{t+j} \approx 1$.

Suppose, for simplicity, that $g_t(Y_{i,s,c,t-1})$ is also linear: $g_t(Y_{i,s,c,t-1}) = \lambda_t Y_{i,s,c,t-1}$. Taking equations (3.1) and (3.8) together, and assuming that all medium-term impacts of rank on learning operate through observed tests scores and observed rank:

$$\frac{\partial Y_{i,s,c,t}}{\partial CR_{i,s,c,t}} = \beta_{t,0}$$

$$\frac{\partial Y_{i,s,c,t+1}}{\partial CR_{i,s,c,t}} = \left(\frac{\partial Y_{i,s,c,t+1}}{\partial CR_{i,s,c,t+1}} \frac{\partial CR_{i,s,c,t+1}}{\partial Y_{i,s,c,t}} + \frac{\partial Y_{i,s,c,t+1}}{\partial Y_{i,s,c,t}}\right) \frac{\partial Y_{i,s,c,t}}{\partial CR_{i,s,c,t}}$$

$$= (\beta_{t+1,0}\, \gamma_{t+1} + \lambda_{t+1}) \beta_{t,0}$$

Similarly:

$$\frac{\partial Y_{i,s,c,t+2}}{\partial CR_{i,s,c,t}} = (\beta_{t+2,0}\, \gamma_{t+2} + \lambda_{t+2})(\beta_{t+1,0}\, \gamma_{t+1} + \lambda_{t+1}) \beta_{t,0}$$

Substituting these expressions, in subsequent grades we get:

$$\frac{\partial Y_{i,s,c,t+l}}{\partial CR_{i,s,c,t}} = \beta_{t,0} \prod_{j=1}^{l} (\beta_{t+j,0}\, \gamma_{t+j} + \lambda_{t+j}) \tag{3.9}$$

Equation (3.9) tells us how the medium-term impact of rank in grade t on learning in grade $t + j$ depends on the short-term impact of rank at the beginning of each grade on learning at the end of that grade ($\beta_{t,0}$), the impact of learning in one grade on learning in the subsequent grade (λ_t), and the impact of learning in one grade on classroom rank in the subsequent grade (γ_t). We also note that, because (as we show below) $\gamma_{t+j} \approx 1$, equation (3.9) indicates that it is possible that we may see little or no decay of rank effects over time. This is because a high classroom rank early in elementary school can in principle lead to a self-fulfilling cycle, where a high rank produces high learning, which in turn leads to a high rank, which in turn leads to high learning. Observed differences between estimates of actual medium-term impacts of rank ($\beta_{t,l}$), and $\frac{\partial Y_{i,s,c,t+l}}{\partial CR_{i,s,c,t}}$ from equation (3.9) tell us about the importance of unobserved skills as mediators of the medium-term impacts of rank.

E. Executive function, non-cognitive skills, and teacher perceptions

To estimate effects of ability classroom rank on happiness in 1st grade and non-cognitive skills in 6th grade, we run regressions comparable to (3.1), replacing achievement in grade t with the relevant outcome.[24] To estimate rank effects on executive function, we also use the model in (3.1), but add to this model a third-order polynomial in lagged EF (in addition to the polynomial in lagged achievement). These regressions use information in 1st through 4th grades, where data on current and lagged EF are available. Finally, as discussed above, we have data on teacher perceptions of students—specifically, a list of the 5 students each teacher thought had the highest, and lowest, achievement in their classrooms. We generate indicator variables for children who are seen to be at the top and, separately, bottom of their classroom by their teachers, and use them as outcomes. To assess medium-term effects of classroom rank on non-cognitive skills, executive function and teacher perceptions we also run regressions analogous to (3.7), replacing achievement in grade $t + l$ with the relevant outcome.

4. Results

A. Graphical evidence

To motivate our analysis, we start with some simple figures. For this purpose, we first sort children into ventiles on the basis of their test scores in math and language at the end of grade $t - 1$ (say, end of

[24] Child happiness is only available in 1st grade, so we run regressions of child happiness in 1st grade on math classroom rank at the beginning of 1st grade and the polynomial on math achievement at the end of kindergarten. Non-cognitive skills are only available at the end of 6th grade, so it is not obvious whether we should regress these skills on rank in 1st grade, 6th grade, or any grade in between. Because (as we argue below) we are most interested in possible medium-term effects of early rank (rank in 1st and 2nd grade), and to be consistent with the results on child happiness, we report the results of 6th grade non-cognitive skills on 1st grade math classroom rank.

kindergarten). Then, within each ventile, we calculate average test scores at the end of grade t (end of 1st grade) for two groups of children: those who, relative to other children in that ventile, were randomly assigned to classrooms where their rank at the beginning of t was "high"—classroom rank in the top 25 percent for that ventile—and those in classrooms where their rank was "low"—in the bottom 25 percent for that ventile. If classroom rank has a positive effect on test scores, we would expect the line that corresponds to high-ranked children to be above that which corresponds to low-ranked children.

Results are shown in Figure 2. Panel A focuses on the short-term effects of classroom rank in 1st grade. The figure shows that children with high classroom ranks have higher achievement than those with low classroom ranks, but only above the 40th percentile of the distribution of national rank. Panel B compares these two groups of children at the end of 3rd grade. The figure shows that the vertical distance between the two lines is larger than in Panel A, indicating that the effect of 1st grade classroom rank increases over time.

Panel C focuses on the short-term effects of classroom rank in 4th grade. Children with higher classroom ranks appear to have higher achievement at the at the end of 4th grade, but the difference is quite small. Panel D focuses on these same children at the end of 6th grade. The lines in this panel are very similar to those in Panel C, suggesting that the (modest) effects of 4th grade classroom rank do not grow over time. In sum, Figure 2 suggests that: the effects of classroom rank are larger in 1st grade than in 4th grade; the 1st grade effects are larger in the upper half of the distribution; and these effects increase over time.

B. Static model

Table 2 reports estimates of the effect of classroom rank on learning, measured by an index of math and language, using equation (3.1) above. Column (1) shows that the coefficient on β is 0.014, with a standard error of 0.007.[25] This implies that moving a child from the 50th to the 60th percentile of classroom rank increases her end-of-grade achievement by 0.14 percentiles of the national distribution. To get a sense of magnitude, we take all children who are in the same school and grade, who have the same value of lagged achievement in $t-1$, but are assigned to different classrooms in grade t. The (absolute value) of the median difference in classroom rank between children in these pairs is 5.5 percentiles (that is, on average, in these pairs of identical children, one child has a classroom rank of 47.25 and the other has a rank of 52.75. At the 75th and 90th percentiles of the difference, the values are 9.8 and 14.7 percentiles, respectively). Column (2) replaces the measure of classroom rank with the average achievement of classroom peers—

[25] Standard errors for all models that pool data across grades are clustered at the student level.

the linear-in-means peer effects model. The coefficient on average peer quality is positive and significant (0.008, with a standard error of 0.004). In column (3), we include both classroom rank and the baseline achievement of peers, as in equation (3.5) above. Because rank and peer quality are negatively related, the coefficient on classroom rank increases to 0.027 (with a standard error of 0.007), and the measure of average peer quality also increases (coefficient of 0.017, with a standard error of 0.005). In column (4), finally, we include classroom-by-grade fixed effects. In this specification, the coefficient on β is 0.030, with a standard error of 0.007—more than twice as large in magnitude as that in the specification in column (1).

Table 3 is based on estimates of equation (3.2) above. Specifically, we report estimates of the effects of classroom rank in math on achievement in math and language (Panels A and B, respectively), as well as the effects of classroom rank in language on achievement in language and math (Panels C and D, respectively). The table shows that classroom rank in math affects math achievement (coefficient of 0.026, with a standard error of 0.007) and, to a lesser extent, language achievement; classroom rank in language does not affect either math or language achievement.[26]

Table 4 reports the results from a number of robustness checks. Panel A refers to specifications with school-by-grade fixed effects (corresponding to equation (3.1) above), while Panel B refers to specifications with classroom-by-grade fixed effects (corresponding to equation (3.6) above). To facilitate comparisons, column (1) in Panel A reproduces the coefficient and standard error from column (1) in Panel A of Table 3, while column (1) in Panel B corresponds to column (4) in Panel A of Table 3.

Columns (2) to (4) report estimates where $g_t(Y_{i,s,c,t-1})$ is specified as a polynomial of orders 1, 2, and 4, respectively (as opposed to our main estimates, in which we include a cubic in lagged achievement). The estimated classroom rank effects are larger when we include only linear or quadratic terms in lagged achievement. Reassuringly, however, the coefficient on classroom rank is essentially unchanged when we include a polynomial of order 4 (rather than order 3) in lagged achievement as a control. Column (5) shows that, as expected given the random assignment, including controls for child gender, as well as age and its square, does not affect our results.

There are between 14,322 (kindergarten) and 17,529 (5th grade) students per grade in our data. These are not always the same students. Typically, from one year to the next, between 7.5 and 10 percent

[26] We do not know why math rank, but not language rank, affects achievement. It is in principle possible that classroom rank in math is more visible to students and teachers than is the case with language rank. However, we do not find strong evidence that this is the case in our setting. As we discuss below, teachers appear to place similar weights on math and language achievement in determining which students in their class have the highest and lowest achievement. We note that it is not uncommon in the literature to find larger effects of school-based interventions on math than on language (see the discussion in Fryer 2017). In the U.S., teachers have larger effects on math than on language achievement (Hanushek and Rivkin 2010).

of students leave our sample of schools (the exception is the transition from kindergarten to 1st grade, where this number is 15 percent).[27] Selective attrition out of our sample of schools could generate a correlation between $CR_{i,s,c,t}$ and $\varepsilon_{i,s,c,t}$, which may bias estimates of the effect of classroom rank. To assess whether our estimates are likely to be affected by these considerations, we use a standard inverse probability weighting (IPW) correction, which gives greater weight to observations that had a higher probability of being lost to follow-up.[28] Column (6) in Table 4 indicates that our estimates are robust to this correction for missing data.

In column (7), finally, we present estimates of equation (3.4). As we discuss above, in this specification—which is similar in character to that used by Murphy and Weinhardt (2020)—classroom rank refers to rank at the end of $t-1$, rather than at the beginning of t. Table 4 shows that these estimates are substantially larger than those from our preferred specification.[29] The approach to classroom rank we take therefore yields conservative estimates of rank effects on learning.

C. Heterogeneity

We begin our analysis of heterogeneity by estimating effects at different points in the distribution of classroom rank, again focusing on the effects of math classroom rank on math achievement. In Figure 2 we graph coefficients and confidence intervals on deciles 1 through 4, and 7 through 10 from equation (3.3) above, with deciles 5 and 6 as the omitted category. The figure shows that there is essentially no impact (or a negative impact) of classroom rank in the bottom half of the distribution. For example, we cannot reject the null that being in the lowest decile of classroom math rank has the same effect on achievement as being in the middle of the distribution of rank. The coefficients that correspond to deciles 7 through 10, on the other hand, are all positive, and are generally larger in the higher deciles (so that the coefficient for the 10th decile is larger in magnitude than that for the 7th decile). In this case, we can reject the null that being in the highest decile of classroom math rank has the same effect on achievement as being in the middle of the distribution.

In Table 5, we analyze other possible sources of heterogeneity in the effect of classroom rank. We first focus on gender. In theory, if girls have different levels of self-confidence than boys (as in Bordalo et al. 2019), or react differently to competition than boys (as in Niederle and Vesterlund 2011), it is possible

[27] Similarly, in any given grade, between 7 and 13 percent of students are new entrants to the sample (with the exception of 1st grade, where this value is 24 percent). New students are randomly assigned to classrooms just like any other students.
[28] We regress attrition on gender, age and its square, a third-order polynomial in lagged test scores, and school-by-grade fixed effects (in Panel A) or classroom-by-grade fixed effects (in Panel B).
[29] It is interesting that classroom rank effects estimated by (3.5) are larger than those estimated by (3.1). As mentioned above, this could happen because children are more aware of end-of-grade rank than beginning-of-grade rank and therefore react more to it, or because end-of-grade rank captures other aspects of the school experience besides rank.

that they react differently to rank. In Table 5, we present estimates of a specification where all the coefficients in equation (3.2) for math are interacted with gender. Girls have significantly lower math scores than boys, but the impact of classroom rank on learning is the same for girls and boys.

Finally, we interact classroom rank with vocabulary at the beginning of kindergarten, or with lagged achievement. These results show that classroom rank effects are substantially and significantly larger for children with higher baseline vocabulary levels, as well as for those with higher lagged achievement. In sum, and consistent with the results in Figure 2, Table 5 shows that classroom rank seems to have larger impacts for higher-achieving children.

D. Dynamics

We begin our analysis of dynamics by estimating effects of classroom rank separately for children in the "early" grades (1st and 2nd grade), "middle" grades (3rd and 4th grade), and "late" grades, both contemporaneously (without lags, as in equation (3.1) above) and at various lags (as in equation (3.6)). These results are in Table 6.[30]

Column (1) shows that the short-term effect of math classroom rank in the early and middle grades are both positive and of a comparable magnitude. On the other hand, classroom rank in 5th and 6th grades has no effect on achievement. Chi-squared tests reject the null that the coefficients on the early, middle, and late grades are the same (p-value: 0.004).

We next turn to the evolution of rank effects over time. Specifically, in columns (2) through (5), we report estimates of the effect of classroom math rank on math achievement after (up to) 1, 2, 3, and 4 lags, respectively. In the first row, corresponding to rank in the early grades, the coefficients increase monotonically over time—the impact after 4 lags is 0.077 (with a standard error of 0.018), almost twice as large as the short-term effect. We reject the null that the coefficients for lag 0 and lag 4 are the same (p-value: 0.044). On the other hand, the coefficients in the second row of the table show that the classroom rank effects in the middle grades decline, although we cannot reject the null that the coefficients for lag=0 and lag=2 are the same (p-value: 0.343).[31]

The results in Table 6 show that the effects of classroom rank in the early grades increase substantially and significantly over time. Given these results, we now ask the following question: can we

[30] In Appendix D we report results from estimating rank effects by grade, rather than by aggregating estimates into "early", "middle", and "late" grades.

[31] We do not know why the effects of rank in the early grades increase, while those in the middle grades do not. We note, however, that some theories of human capital argue that earlier investments tend to have the largest effects (as in Cunha and Heckman 2007), in part because earlier investments allow children to better take advantage of later investments. It is possible that early classroom rank affects achievement in this way.

account for the increase in the effect of early classroom rank using estimates of $\beta_{t,0}$ (short-term impact of classroom rank), γ_t (impact of learning on future rank), and λ_t (impact of lagged skills on current skills)?

Estimates of $\beta_{t,0}$, γ_t, and λ_t for each grade t show that $\gamma_t \approx 1$ but $\beta_t + \lambda_t$, <1 for every grade, which means that, if rank operated primarily through short-term learning gains and the resulting improvement in subsequent rank, we should observe fade-out in the impacts of rank on learning. As we have seen in Table 6, this is not the case. Rather, our results suggest that classroom rank operates at least in part by producing sustainable changes in another unobserved skill, which has independent effects on learning.

A useful way to make this point is in Figure 4, which plots the implied change in learning in grades $t + l$, as a response to an exogenous change in early (1st or 2nd grade) achievement ($t = 0$) percentile rank by 10 points, under two scenarios: (i) using the estimates of β_{t+l} (for l =0, 1…4) from equation (3.7), labeled *reduced form* in the figure; and (ii) using the estimates of γ_t, β_t and λ_t, (for several values of t) from equations (3.1) and (3.8), and then simulating the response to a particular change in rank using the equations (3.9), labeled *structural* in the figure. The figure shows that, under scenario (i), the effect of rank grows over time, while in scenario (ii) it does not.[32]

In sum, Figure 4 suggests that early classroom rank affects future achievement through channels that are not modeled explicitly in our equations, such as through its impact on other unobservable skills. We now turn directly to this question.

E. Executive function, non-cognitive skills, and teacher perceptions

As discussed above, there is a large literature in child psychology that argues that executive function is a key determinant of learning (Anderson 2002; Espy 2004; Senn et al. 2004). In our data, too, EF in a given grade predicts future achievement.[33] It is therefore of interest to analyze whether classroom rank improves executive function. Table 7 shows estimated effects of classroom rank in math on executive function (standardized to have mean zero and unit standard deviation) separately for children in the "early" grades (1st and 2nd grade) and "middle" grades (3rd and 4th grade), both contemporaneously and at various lags.[34] Below the standard errors in parentheses we report p-values computed based on the Romano-Wolf

[32] For the purpose of illustration, we first normalize the estimate of $\beta_{t,0}$ to be the same across the two scenarios.

[33] We do not have the data to identify the causal effect of executive function on learning. Nevertheless, in a regression of achievement in grade t on executive function in grade $t-1$, including school fixed effects, the coefficient on EF is 0.538 (with a standard error of 0.005). If, in addition, we control for achievement at the end of $t-1$, the coefficient on EF is 0.096 (with a standard error of 0.002).

[34] We did not apply executive function tests past 4th grade.

stepdown procedure using 5,000 bootstrap replications (see Romano and Wolf, 2005 and Clarke, Romano and Wolf, 2020).

The first column of Table 7 shows that the coefficients on contemporaneous classroom math rank are positive for both early and middle grades, albeit not significant. Looking across rows, we find that early classroom rank has a positive and significant effect on executive function after one lag. The estimated effect implies that an increase in early classroom rank from the 50th to the 60th percentile of the distribution improves executive function by 2.02 percent of a standard deviation. However, the effect of early classroom rank on executive function fades out after two lags (the estimated coefficient is 0.021 with standard error 0.067).

Table 8 reports the marginal effects from ordered probit regressions of child happiness on math classroom rank in 1st grade. These results show that moving a child from the 50th to the 60th percentile of the distribution of classroom rank increases the probability that a child says she is always happy in school by 1 percentage point.

Table 9 reports the results from regressions of 6th grade non-cognitive skills on math classroom rank in "early" (1st and 2nd), "middle" (3rd and 4th) and "late" (5th and 6th) grades. While all estimated coefficients are positive, they are not significant (with the exception of the effect of late classroom rank on self-esteem, which is estimated to be 0.185 with standard error 0.110).

Finally, we turn to teacher perceptions. Teacher ratings are informative about student performance. Although the top (bottom) 5 students reported by the teacher are not always the 5 highest (lowest) performing students in the tests we administer, there is some overlap.[35] The fact that these two measures are only imperfectly correlated could in part be a result of measurement error in either one of them. It is also possible, however, that in assessing "achievement" teachers are in fact taking account of a broader or somewhat different construct, i.e., teacher perceptions do not measure exactly the same thing as the tests. Furthermore, if teachers perceive highly-ranked children to be particularly high achieving—even conditional on their actual ability—they may reinforce academic self-concept of highly-ranked children, and thus contribute to the impact of rank on learning outcomes we observe in both the short- and medium-run.

In Table 10 we report the results of regressions of teacher perceptions on (randomly assigned) classroom rank in math, conditional on achievement in math. We pool across "early" (1st and 2nd), "middle"

[35] There is no evidence that teachers make more use of math or language achievement in assessing who are top and bottom students: The correlations between being in the top 5 by measured math achievement and language achievement, on the one hand, and having a teacher report a student as being in the top 5 are 0.38 and 0.33, respectively, while the correlations between being in the bottom 5 by measured math achievement and language achievement, on the one hand, and having a teacher report a student as being in the bottom 5 are 0.38 and 0.41, respectively.

(3^{rd} and 4^{th}), and "late" (5^{th} and 6^{th}) grades. Here too we report p-values computed based on the Romano-Wolf stepdown procedure using 5,000 bootstrap replications. In column 1 of Table 10, we show that, in a regression that controls for achievement at the end of $t-1$, higher classroom rank at the start of grade t increases the probability that a child is seen as a top student by her teacher in $t+1$ for early, middle and late grades. Moreover, in column (5) we show that higher classroom rank at the start of grade t in early grades reduces the probability that the student is seen as a bottom student by her teacher in $t+1$. Looking across rows, we find that the effect of early and middle classroom rank on the probability of being seen as a top student by a teacher is positive and significant after several lags, but tends to fade out for early rank. Interestingly, a higher classroom rank in early grades significantly reduces the probability of being seen as a bottom student after 4 lags.

In sum, we show that children who, as a result of random assignment, have higher math classroom rank have higher levels of executive function, are more likely to be happy, and are also perceived to be higher-achieving by their future teachers.

5. Conclusion

This paper analyzes the impact of classroom ability rank measured at the start of the academic year on learning during that year, and learning in subsequent years. In our data, which comes from a longitudinal study of students in elementary schools in Ecuador, two students with the same underlying ability and attending the same school can have different classroom ranks because they are randomly assigned to different classrooms, with slightly different peers.

We measure classroom rank and learning in math and language. Beginning-of-grade classroom rank and end-of-grade achievement are available for all grades from 1^{st} to 6^{th} grade. We also observe executive function in kindergarten through 4^{th} grade, self-reported child happiness in 1^{st} grade, and non-cognitive skills at the end of 6^{th} grade. In addition, we have data on teacher perceptions of student ability in every grade between kindergarten and 6^{th} grade.

We show that classroom rank has modest short-term effects on achievement. Estimated effects of classroom rank can be confounded by the effects of peer quality. Students randomly assigned to classrooms with better peers will in general have lower classroom ability rank, but potentially benefit from better peers. Our estimates of the effect of classroom rank on learning are not affected when we control for average peer quality, or when we include classroom fixed effects. The conflation of rank and peer quality effects is a feature of any study where both change at the same time, such as studies of the impact of selective schools, affirmative action, or neighborhood effects.

We also show that classroom rank in math, but not language, affects achievement. The impact of classroom rank in math is larger for younger children and grows substantially over time. Moving a child from the 50^{th} to the 60^{th} percentiles of classroom rank in 1^{st} (2^{nd}) grade increases her achievement in 5^{th} (6^{th}) grade by 1 percentile of the national distribution. The increase in the magnitude of rank effects is remarkable given the evidence that impacts of many other interventions in elementary school fade out over time. Exogenous changes in classroom math rank also improve executive function, happiness, and teacher perceptions of students. Changes in these skills, or others that we do not observe, are likely to be important in explaining how classroom rank raises child learning.

References

Angrist, J. 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.

Aylward, G. P., and T. Stancin. 2008. "Measurement and Psychometric Considerations." In Wolraich, M. L., D. D. Drotar, P. H. Dworkin, and E. C. Perrin, Eds., *Developmental-Behavioral Pediatrics*. Philadelphia: Mosby.

Bai, J. 2009. "Panel Data Models with Interactive Fixed Effects." *Econometrica* 77(4): 1229–79.

Bailey, D., G. J. Duncan, F. Cunha, B. Foorman, and D. Yeager. 2020. "Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions." *Psychological Science in the Public Interest* 21(2): 55-97.

Bandiera, O., I. Barankay, and I. Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." *Quarterly Journal of Economics* 120(3): 917–62.

Barrera-Osorio, F., K. Gonzalez, F. Lagos, and D. Deming. 2020. "Providing Performance Information in Education: An Experimental Evaluation in Colombia." *Journal of Public Economics* 186 (June).

Berlinski, S., and N. Schady. 2015. *The Early Years: Child Well-Being and the Role of Public Policy*. New York: Palgrave Macmillan.

Blackwell, L. S., K. H. Trzesniewski, and C. S. Dweck. 2007. "Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention." *Child Development* 78(1): 246–63.

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.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer. 2019. "Beliefs about Gender." *American Economic Review* 109(3): 739–73.

Card, D., A. Mas, E. Moretti, and E. Saez. 2012. "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction. *American Economic Review* 102(6): 2981–3003.

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

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

Cicala, S., R. G. Fryer, and J. L. Spenkuch. 2018. "Self-Selection and Comparative Advantage in Social Interactions." *Journal of the European Economic Association* 16(4): 983–1020.

Clarke, D., J. P. Romano, and M. Wolf. (2020). "The Romano–Wolf multiple-hypothesis correction in Stata." *The Stata Journal*, 20(4), 812-843.

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

Delaney, J., and P. Devereux. 2021. "High school rank in math and English and the gender gap in STEM." *Labour Economics* 69.

Denning, J. T., R. Murphy, and F. Weinhardt. 2020. "Class Rank and Long-Run Outcomes." Working Paper 27468, National Bureau of Economic Research.

Duckworth, Angela, and Martin Seligman, 2005. "Self-Discipline Outdoes IQ in Predicting Academic Performance in Adolescence." *Psychological Science* 16(12): 939-44.

Duckworth, A., and P. D. Quinn. 2009. "Development and Validation of the Short Grit Scale (GritS)." *Journal of Personality Assessment* 91(2): 166-174.

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.

Dunn, L. D. Lugo, E. Padilla, and L. Dunn. 1986. *Test de Vocabulario en Imágenes Peabody.* (Circle Pines, MN: American Guidance Service).

Dweck, C. 2008. *Mindset: The New Psychology of Success.* (Ballantine Books: New York).

Elsner, B. and I. E. Isphording. 2017. "A Big Fish in a Small Pond: Ability Rank and Human Capital Investment." *Journal of Labor Economics* 35(3): 787–828.

Elsner, B. and I. E. Isphording. 2018. "Rank, Sex, Drugs, and Crime." *Journal of Human Resources* 53(2): 356–81.

Elsner, B., I. E. Isphording, and U. Zölitz. 2018. "Achievement Rank Affects Performance and Major Choices in College." University of Zurich, Department of Economics, Working Paper (300).

Epple, D. and R. E. Romano. 2011. "Peer Effects in Education: A Survey of the Theory and Evidence." In *Handbook of Social Economics*, Volume 1, pp. 1053–1163. Elsevier.

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.

Fryer, R. G. 2017. "The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments." In: *Handbook of Field Experiments*, Vol. 2, E. Duflo and A. Banerjee, eds. (Amsterdam: North Holland).

Goodman, A., and J. Smith. 2011. "The Long Shadow Cast by Childhood Physical and Mental Problems on Adult Life." *Proceedings of the National Academy of Sciences* 108: 6032-37.

Harris, K. M., and R. J. Udry. 2018. *National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2008* [Public Use]. Carolina Population Center, University of North Carolina-Chapel Hill, Inter-University Consortium for Political and Social Research.

Heckman, J. 2013. *Giving Kids a Fair Chance.* (MIT Press: Cambridge, MA).

Ioannides, Y. M. 2011. "Neighborhood Effects and Housing." In *Handbook of Social Economics*, Volume 1, pp. 1281–1340. Elsevier.

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

Jochmans, K. 2020. "Testing Random Assignment to Peer Groups." Cambridge Working Papers in Economics 2024, Faculty of Economics, University of Cambridge.

Klein, D. N., L. R. Dougherty, and T. M. Olino. 2005. "Toward Guidelines for Evidence-Based Assessment of Depression in Children and Adolescents." *Journal of Clinical Child and Adolescent Psychology* 34(3): 412-32.

Lazear, E. P. and S. Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy* 89(5): 841–64.

Little, R. J. and D. B. Rubin. 2019. *Statistical Analysis with Missing Data*, Volume 793. John Wiley & Sons.

Marsh, H. W. 1987. "The Big-Fish-Little-Pond Effect on Academic Self-Concept." *Journal of Educational Psychology* 79(3): 280-95.

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.

Murphy, R. and F. Weinhardt. 2020. "Top of the Class: The Importance of Ordinal Rank." *The Review of Economic Studies,* 87(6): 2777–2826.

Niederle, M. and L. Vesterlund. 2011. "Gender and Competition." *Annual Review of Economics* 3(1): 601–30.

Olino, T. M., L. Yu, D. L. McMakin, E. E. Forbes, J. R. Seeley, P. M. Lewinsohn, and P. A. Pilkonis. 2013. "Comparisons Across Depression Assessment Instruments in Adolescence and Young Adulthood: An Item Response Theory Study Using Two Linking Methods." *Journal of Abnormal Child Psychology* 41(8): 1267–77.

Paxson, C., and N. Schady. 2007. "Cognitive Development among Young Children in Ecuador: The Roles of Wealth, Health, and Parenting." *Journal of Human Resources* 42(1): 49-84.

Romano, J. P., and M. Wolf. 2005. "Stepwise multiple testing as formalized data snooping." *Econometrica*, 73(4), 1237-1282.

Rosenberg, M. 1989. *Society and the Adolescent Self-Image* (Rev. ed.). Middletown, CT: Wesleyan University Press.

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

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.

Segal, C. 2013. "Misbehavior, Education, and Labor Market Outcomes." *Journal of the European Economic Association* 11(4): 743-79.

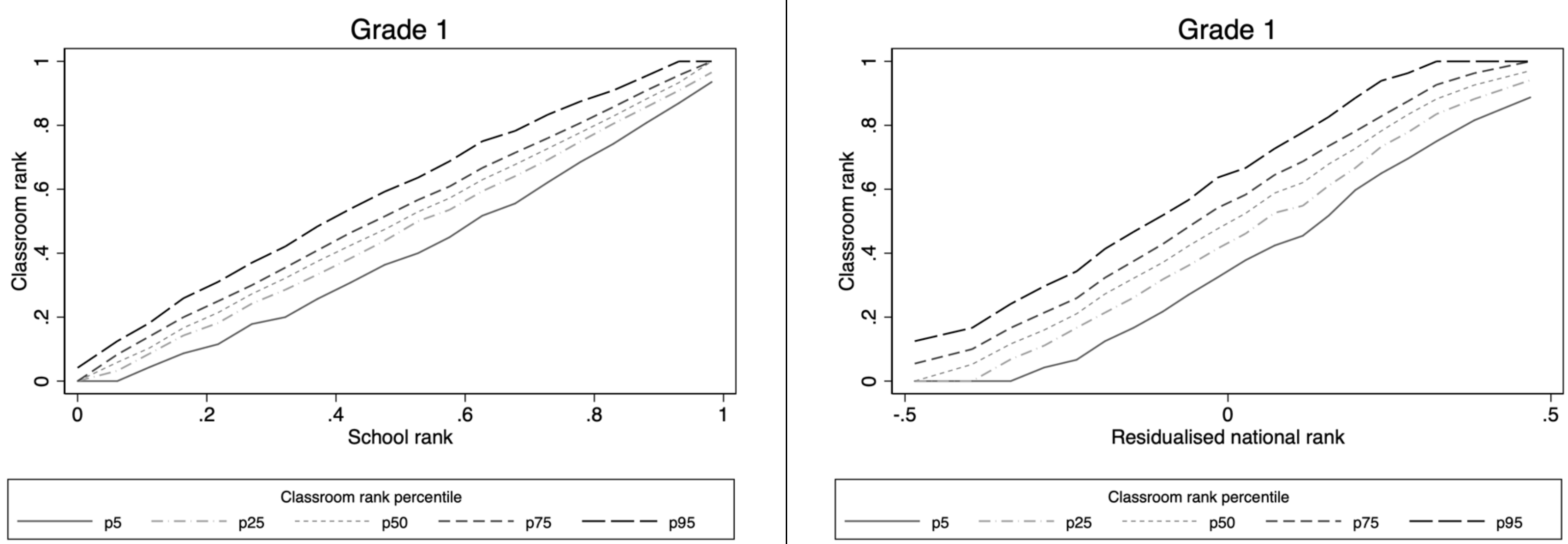
Siegler, R. S., and J. L. Booth. 2004. "Development of Numerical Estimation in Young Children." *Child Development* 75(2): 428-44.

Tincani, M. M. 2018. "Heterogeneous Peer Effects in the Classroom." Unpublished paper, University College London.

Yeager, D. S., P. Hanselman, G. M. Walton, J. S. Murray, R. Crosnoe, C. Muller, E. Tipton, B. Schneider, C. S. Hulleman, C. P. Hinojosa, D. Paunesku, C. Romero, K. Flint, A. Roberts, J. Trott, R. Iachan, J. Buontempo, S. Man Yang, C. M. Carvalho, P. R. Hahn, M. Gopalan, P. Mhatre, R. Ferguson, A. L. Duckworth, and C. S. Dweck. 2019. "A National Experiment Reveals Where a Growth Mindset Improves Achievement." *Science* 573(7774): 364-69.

Figure 1: Variation in classroom rank within school

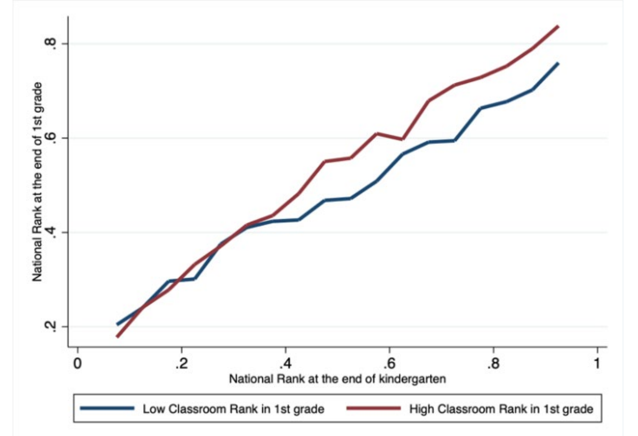
Panel A: Percentiles of classroom rank by school rank	Panel B: Percentiles of classroom rank by residualised national rank



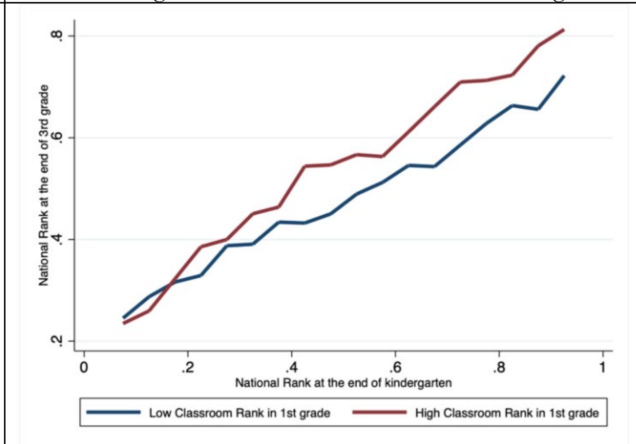
Notes: To construct the figure in Panel A we plot the 5th, 25th, 50th, 75th and 95th percentile of classroom rank within each ventile of school rank, against the median school rank in that ventile. To construct the figure in Panel B, we first regress national ability rank on school fixed effects and derive residuals. Then we plot percentiles of classroom rank within each ventile of residualised national ability rank, against the median residualised national rank in that ventile.

Figure 2: Visual evidence of classroom rank effects on achievement

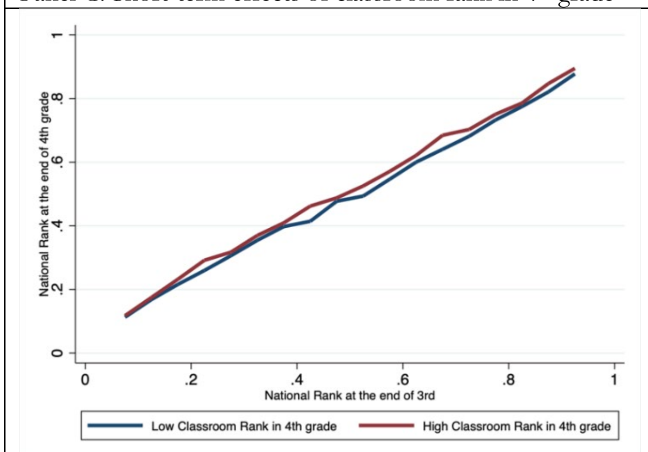
Panel A: Short-term effects of classroom rank in 1st grade



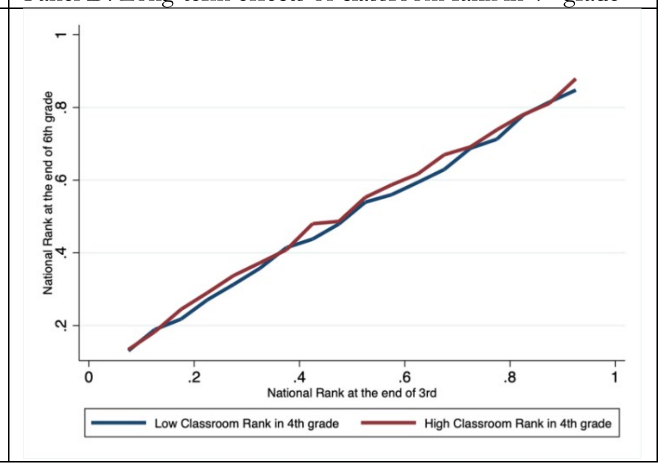
Panel B: Long-term effects of classroom rank in 1st grade



Panel C: Short-term effects of classroom rank in 4th grade

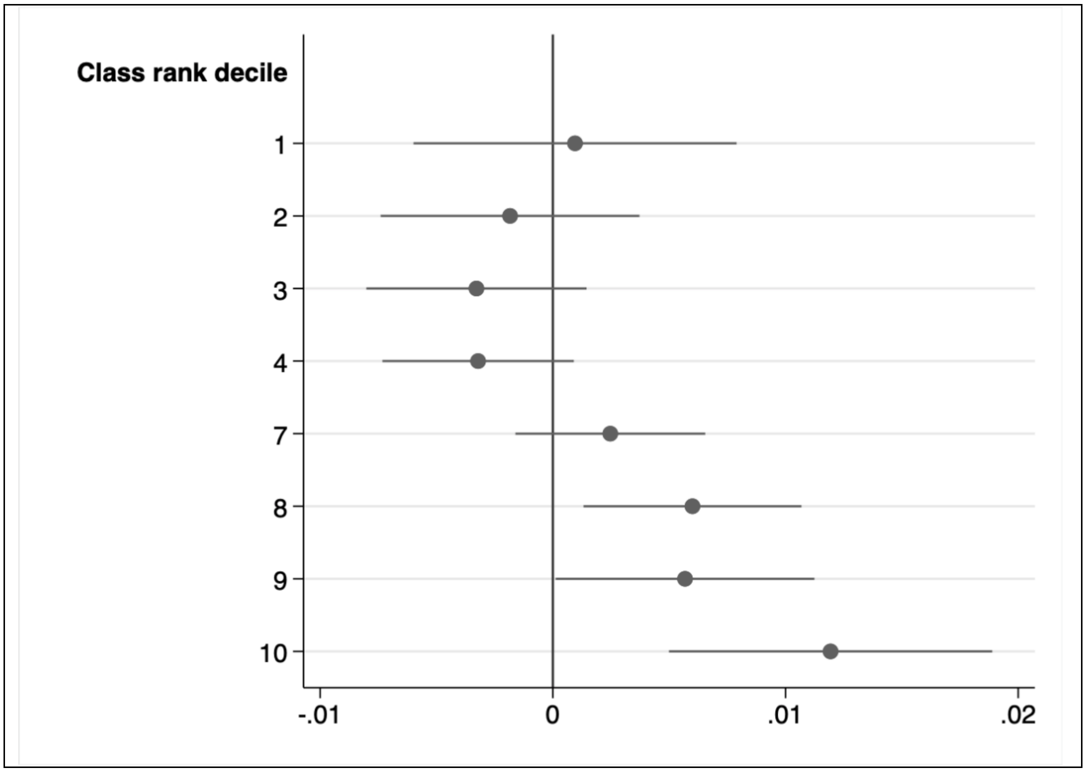


Panel D: Long-term effects of classroom rank in 4th grade



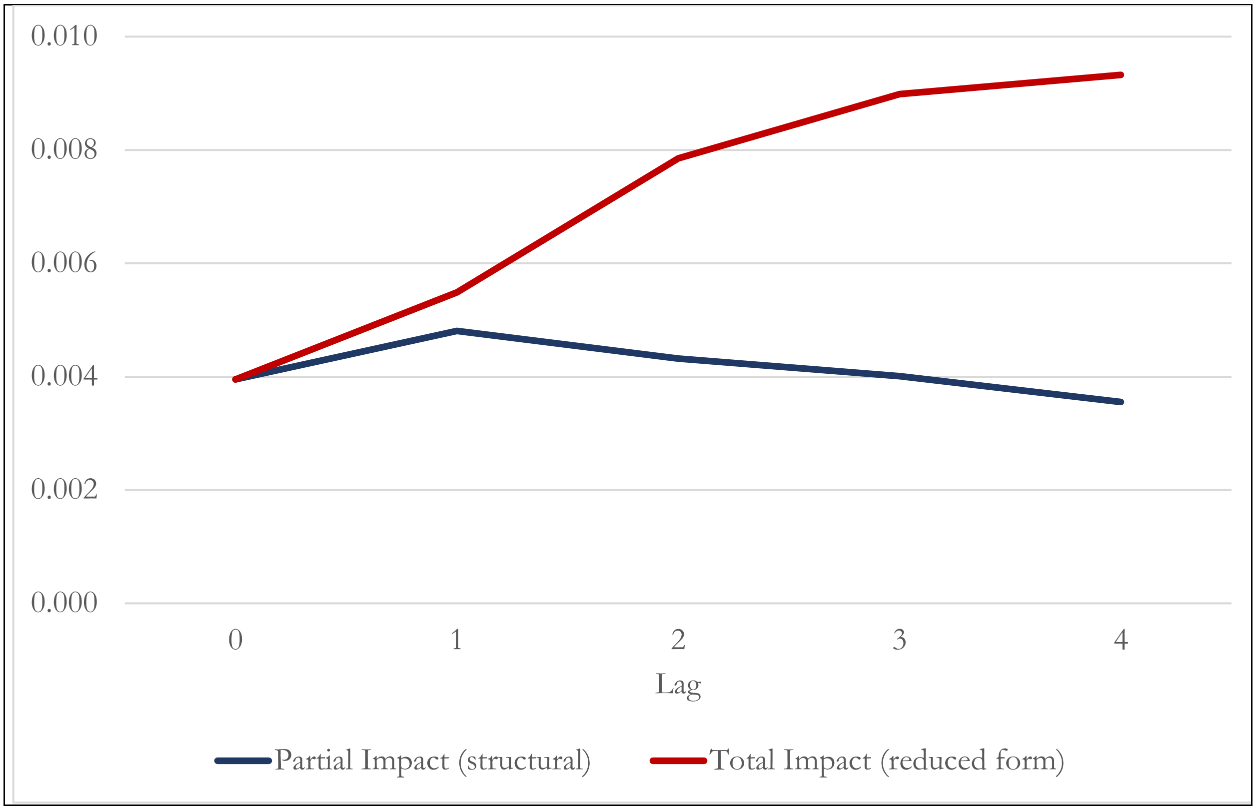
Notes: To generate this figure, we first sort children into ventiles on the basis of their test scores at the end of grade $t-1$. Then, for each ventile, we calculate average test scores at the end of grade t for two groups of children: those who, relative to other children in that ventile, were randomly assigned to classrooms where their rank at the beginning of t was "high"—classroom rank in the top 25 percent for that ventile—and those in classrooms where their rank was "low"—in the bottom 25 percent for that ventile. Panel A focuses on the short-term effects of classroom rank in 1st grade, and Panel B compares these two groups of children at the end of 3rd grade. Panel C focuses on the short-term effects of classroom rank in 4th grade, and Panel D focuses on these same children at the end of 6th grade.

Figure 3: Classroom rank effects at different deciles of the distribution of achievement



Notes: To produce this figure we discretize classroom rank in math into 10 deciles, and run regressions of math achievement on classroom rank deciles. We graph coefficients and 90 percent confidence intervals on deciles 1 through 4, and 7 through 10, with deciles 5 and 6 as the omitted category. All regressions include a third order polynomial in lagged national rank and school-by-grade fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout.

Figure 4: Total and partial effects of rank on achievement



Notes: The figure plots the implied change in learning in grades $t + l$, as a response to an exogenous change in early (1st or 2nd grade) achievement ($t = 0$) percentile rank by 10 points, under two scenarios: (i) using the estimates of β_{t+l} (for l=0, 1…4) from equation (3.7) in the main text "total impact"); and (ii) using the estimates of γ_t, β_t and λ_t, (for several values of t) from equations (3.1) and (3.8) in the main text, and then simulating the response to a particular change in rank using the equations (3.9) "partial impact". We normalize the estimate of $\beta_{t,0}$ to be the same across the two scenarios.

Table 1: Child, teacher, and classroom characteristics

	Mean	Standard deviation	N
Child and household characteristics			
Age of child (months)	60.3	4.9	13,858
Gender of child	0.49	0.50	14,477
Receptive vocabulary score (TVIP)	83.3	16.9	13,733
Mother's years of completed schooling	8.8	3.8	13,627
Father's years of completed schooling	8.5	3.8	10,594
Mother's age	30.2	6.6	13,637
Father's age	34.6	7.9	10,620
Proportion who attended preschool	0.61	0.49	14,472
Household has piped water in home	0.83	0.38	14,407
Household has flush toilet in home	0.46	0.50	14,407
Teacher and classroom characteristics			
Proportion female	0.82	0.38	2830
Proportion tenured	0.82	0.38	2818
Years of experience	18.1	10.5	2820
Class size	36.2	6.4	2838

Notes: Table reports means and standard deviations of the characteristics of children entering kindergarten in 2012, measured at the beginning of the school year, and of the teachers they had between kindergarten and 6th grade. The TVIP is the *Test de Vocabulario en Imágenes Peabody*, the Spanish version of the Peabody Picture Vocabulary Test (PPVT). The test is standardized using the tables provided by the test developers which set the mean at 100 and the standard deviation at 15 at each age.

Table 2: Effect of classroom rank and peer quality on achievement

	(1)	(2)	(3)	(4)
Classroom rank	0.014**		0.027***	0.030***
	(0.007)		(0.007)	(0.007)
Mean of classroom peers		0.008**	0.017***	
		(0.004)	(0.005)	
School-by-grade fixed effects	X	X	X	
Classroom-by-grade fixed effects				X

Notes: The table reports estimates from regressions of achievement national rank on classroom rank and the leave-one-out mean of classroom peer achievement, pooling observations across grades. Column (1) shows our main model results. We regress national rank on classroom rank at the beginning of the school year, including a third-order polynomial in lagged national rank. Column (2) regresses national achievement rank on the leave-one-out mean achievement of classroom peers. Column (3) combines classroom rank and the leave-one-out mean of classroom peers. Columns (1)-(3) include school-by-grade fixed effects. Column (4) estimates the main model using classroom-by-grade fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. Sample size is 87,706 observations in all columns. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 3: Classroom rank effects, separating math and language

	(1)	(2)	(3)	(4)
	Panel A: Effect of rank in math on math achievement			
Classroom rank	0.026***		0.037***	0.040***
	(0.007)		(0.009)	(0.009)
Mean of classroom peers		0.001	0.012***	
		(0.004)	(0.005)	
Observations	87,713	87,713	87,713	87,713
	Panel B: Effect of rank in math on language achievement			
Classroom rank	0.016		0.028**	0.030**
	(0.010)		(0.012)	(0.012)
Mean of classroom peers		0.005	0.014**	
		(0.005)	(0.006)	
Observations	87,760	87,760	87,760	87,760
	Panel C: Effect of rank in language on math achievement			
Classroom rank	0.001		0.015	0.017
	(0.010)		(0.012)	(0.012)
Mean of classroom peers		0.014**	0.018***	
		(0.006)	(0.007)	
Observations	87,740	87,740	87,740	87,740
	Panel D: Effect of rank in language on language achievement			
Classroom rank	-0.008		0.000	0.003
	(0.008)		(0.009)	(0.009)
Mean of classroom peers		0.011**	0.011**	
		(0.005)	(0.005)	
Observations	87,789	87,789	87,789	87,789
School-by-grade fixed effects	X	X	X	
Classroom-by-grade fixed effects				X

Notes: The table reports estimates from regressions of achievement national rank on classroom rank and the leave-one-out mean of classroom peer achievement, pooling observations across grades. For the regressions in Panels A and B, classroom rank is calculated on the basis of math test scores, and a third-order polynomial in lagged math achievement is included as a control, while for the regressions in Panels C and D, classroom rank is calculated on the basis of language test scores, and a third-order polynomial in lagged language achievement is included as a control. Column (1) shows our main model results. Column (2) regresses national achievement rank on the leave-one-out mean achievement of classroom peers. Column (3) combines classroom rank and the leave-one-out mean of classroom peers. Columns (1)-(3) include school-by-grade fixed effects. Column (4) estimates the main model using classroom-by-grade fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 4: Robustness checks, effects of math classroom rank on math achievement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Estimations with school-by-grade fixed effects						
Classroom rank	0.026***	0.042***	0.042***	0.026***	0.025***	0.032***	0.109***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)
School-by-grade fixed effects	X	X	X	X	X	X	X
Age and gender					X		
	Panel B: Estimations with classroom-by-grade fixed effects						
Classroom rank	0.040***	0.063***	0.063***	0.040***	0.039***	0.040***	0.107***
	(0.009)	(0.008)	(0.008)	(0.009)	(0.008)	(0.012)	(0.007)
Classroom-by-grade fixed effects	X	X	X	X	X	X	X
Age and gender					X		

Notes: The table reports estimates from regressions of national rank in math on classroom math rank. Observations are pooled across grades. All estimates include school-by-grade fixed effects. Column (1) reproduces the result from column (1) of Panel A in Table 4. In columns (2), (3), and (4), lagged achievement in math enters the regression as a linear term, quadratic term, or fourth-order polynomial (as opposed to a cubic term). Column (5) is comparable to column (1) but adds controls for child gender, age, and its square. In column (6) we use multiple imputation to correct for missing data, using the approach in Little and Rubin (2019). Column (7) corresponds to estimates of equation (3.3) in the main text. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 5: Heterogeneity of math classroom rank effects, by gender and ability

	(1) Gender	(2) Baseline vocabulary	(3) Lagged national rank
Classroom rank	0.025***	0.018*	-0.009
	(0.008)	(0.009)	(0.014)
Main covariate effect	-0.006***	0.020***	1.056***
	(0.002)	(0.001)	(0.060)
Interaction (rank*covariate)	0.001	0.007**	0.069***
	(0.004)	(0.003)	(0.023)

Notes: The table reports estimates from regressions of national rank in math on classroom math rank and interactions. Observations are pooled across grades. Column (1) shows the results from a regression of national rank on classroom rank interacted with an indicator variable for girls. Column (2) shows the results from a regression of national rank on classroom rank interacted with baseline vocabulary. Column (3) shows the results from a regression of national rank on classroom rank interacted with lagged national rank. All regressions are limited to schools in which there are at least two classrooms. All regressions include a third order polynomial in lagged national rank in math and school-by-grade fixed effects. Standard errors are clustered at the student level. N is 87,700 for column (1), 61,178 for column (2), and 87,713 for column (3). N is lower for column (2) because we collected vocabulary at the beginning of kindergarten, so it is not available for children who joined schools in our sample after the beginning of kindergarten. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 6: Effects of math classroom rank on achievement, by grade and lag

	Lags						
	0	1	2	3	4	F-test 1	F-test 2
Classroom rank "early" (1st & 2nd grades)	0.042***	0.052***	0.057***	0.072***	0.077***	0.293	0.044
	(0.014)	(0.016)	(0.017)	(0.018)	(0.018)		
Classroom rank "middle" (3rd & 4th grades)	0.040***	0.032**	0.028*			0.626	0.343
	(0.012)	(0.014)	(0.015)				
Classroom rank "late" (5th & 6th grades)	-0.005						
	(0.010)						
F-test 3	0.004	0.329	0.162				
F-test 4	0.008						

Notes: The table reports estimates from regressions of national rank in math on classroom math rank for different lags of classroom rank, separately for children in the "early" (1^{st} and 2^{nd}), "middle" (3^{rd} and 4^{th}), and "late" grades (5^{th} and 6^{th}) grades. All regressions are limited to schools in which there are at least two classes. All regressions include a third order polynomial in lagged national rank in math and school-by-grade fixed effects. Standard errors are clustered at the student level. F-tests are calculated after running estimates by Seemingly Unrelated Regressions. F-test 1 is a test that the coefficient for all ranks is the same, and F-test 2 is a test that the coefficient on lag=0 is the same as that on lag=4 (for the "early" grades) or lag=2 (for the "middle" grades). F-test 3 is a test that the coefficients on "early", "middle", and "late" grades are the same, and F-test 4 is a test that the "early" and "late" effects are the same. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 7: Math classroom rank effects on executive function

	Lags		
	0	1	2
Classroom rank "early" (1st & 2nd grades)	0.071	0.202***	0.021
	(0.061)	(0.064)	(0.067)
	0.2787	0.0002	0.6597
Classroom rank "middle" (3rd & 4th grades)	0.038		
	(0.061)		
	0.6169		

Notes: The table reports estimates from regressions of executive function, in SDs, on classroom achievement rank in math, for different lags of classroom rank, separately for children in the "early" (1st and 2nd) and "middle" (3rd and 4th) grades. All regressions include third order polynomials in lagged national rank in math, a third-order polynomial in lagged executive function, and school-by-grade fixed effects. Below the standard errors in 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. All regressions are limited to schools in which there are at least two classrooms per grade. We cannot assess the impact of classroom rank past 4th grade, as we did not apply executive function tests after that grade. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 8: Math classroom rank effects on happiness

	Child happiness (1st grade)		
	Mostly happy	Almost always happy	Always happy
Classroom rank	-0.055**	-0.042**	0.098**
	(0.023)	(0.018)	(0.040)

Notes: The table reports estimates from a regressions of child happiness in 1st grade on classroom achievement rank in math in 1st grade, estimated by ordered probit. All regressions include third order polynomials in kindergarten national rank in math and school fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. N is 12,062. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 9: Math classroom rank effect on non-cognitive skills

	(1)	(2)	(3)	(4)	(5)
	Non-cognitive aggregate	Depression	Self-esteem	Growth mindset	Grit
Classroom rank "early" (1st & 2nd grades)	0.124	0.112	0.106	0.117	0.036
	(0.096)	(0.095)	(0.098)	(0.095)	(0.099)
Classroom rank "middle" (3rd & 4th grades)	0.120	0.099	0.127	0.040	0.091
	(0.106)	(0.105)	(0.109)	(0.105)	(0.110)
Classroom rank "late" (5th & 6th grades)	0.171	0.124	0.185*	0.036	0.154
	(0.106)	(0.105)	(0.110)	(0.105)	(0.111)

Notes: The table reports estimates from regressions of a given 6th grade non-cognitive skill, or the non-cognitive aggregate of all four skills, on classroom achievement rank in math in "early" (1st and 2nd), "middle" (3rd and 4th), and "late" grades (5th and 6th). All regressions include third order polynomials in kindergarten national rank in math and school-by-grade fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. N is 15,578. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 10: Math classroom rank effects on teacher perceptions

	Top 5 student				Bottom 5 student			
	Lags				Lags			
	0	1	2	3	0	1	2	3
Rank "early" (1st & 2nd grade)	0.071***	0.077***	0.048**	0.032	-0.046**	-0.023	-0.009	-0.043**
	(0.024)	(0.023)	(0.024)	(0.023)	(0.022)	(0.021)	(0.021)	(0.021)
	0.0008	0.0002	0.0452	0.2801	0.0416	0.5069	0.8914	0.0372
Rank "middle" (3rd & 4th grade)	0.051**	0.057**			0.011	-0.002		
	(0.025)	(0.025)			(0.025)	(0.025)		
	0.0452	0.0154			0.8914	0.9156		
Rank "late" (5th & 6th grade)	0.061*				-0.029			
	(0.035)				(0.034)			
	0.1062				0.6559			

Notes: The table reports the results from regressions of a child being reported to be among the top 5 (bottom 5) by achievement by her teachers in grade *t+1* on classroom rank in grade *t*, controlling for a third-order polynomial in national achievement in math in grade *t-1*, and school-by-grade fixed effects, pooling across "early" (1st and 2nd), "middle" (3rd and 4th) and "late" (5th and 6th) grades. Below the standard errors in 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. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. *Significant at 10%, **significant at 5%, ***significant at 1%.

Appendix A

An important assumption underlying our empirical strategy is that children's classroom rank at the beginning of a given grade is random, due to random assignment of children to classrooms within schools in every year.[36] Random assignment is closely monitored, and compliance is very high, 98.9 percent on average. In this appendix, we present tests of random assignment using a methodology developed in Jochmans (2020).

First, we briefly discuss the procedure outlined in Jochmans (2020). 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 (2020) shows that a fixed-effects regression of $x_{s,i}$ on $\bar{x}_{s,i}$ will yield biased estimates due to inconsistency of the within-group estimator. The proposed corrected estimator is given by

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

where $\tilde{x}_{s,i}$ is the deviation of $x_{s,i}$ from its within-school mean. The null hypothesis is thus absence of correlation between $i's$ characteristics and those of her peers. To test the random assignment in our setting, we implement this procedure by testing for the presence of correlation between child $i's$ scores measured at the end of grade $t-1$ and the average end-of-grade scores in $t-1$ of the classroom peers assigned to her in a given grade t. We do so for each grade. We implement the test for all children in the sample, and restricting the sample to those children who have both end of grade $t-1$ scores as well as end of grade t scores (as these will be the children that end up being included in the estimation of our models). The results are shown in tables A1 and A2, respectively. 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. This result is true for all grades and both samples. Hence, we conclude that random assignment was successful in our setting.

[36] We use the word "random" as shorthand but, as discussed at length in Araujo et al. (2016) and Campos et al. (2020), 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.

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

	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Test statistic	1.359	-0.383	0.905	0.300	-0.445	-0.222	0.980
P-value	0.174	0.702	0.366	0.764	0.657	0.825	0.327

Notes: In this table, we report results for tests of random assignment of children to classrooms within schools using a methodology proposed by Jochmans (2020). The null hypothesis is absence of correlation between a child's ability measured at the end of the previous grade and the average ability of classroom peers assigned to her at the beginning of a given grade, conditional on school. The sample includes all children.

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

	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Test statistic	1.392	-0.005	1.425	0.413	-0.043	0.001	1.037
P-value	0.164	0.996	0.154	0.680	0.966	0.999	0.300

Notes: In this table, we report results for tests of random assignment of children to classrooms within schools using a method- ology proposed by Jochmans (2020). The null hypothesis is absence of correlation between a child's ability measured at the end of the previous grade and the average ability of classroom peers assigned to her at the beginning of a given grade, conditional on school. The sample is restricted to children who have available both beginning- and end-of-grade scores for a given grade.

Appendix B

This appendix presents additional information on test scores, executive function, and non-cognitive skills. 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.

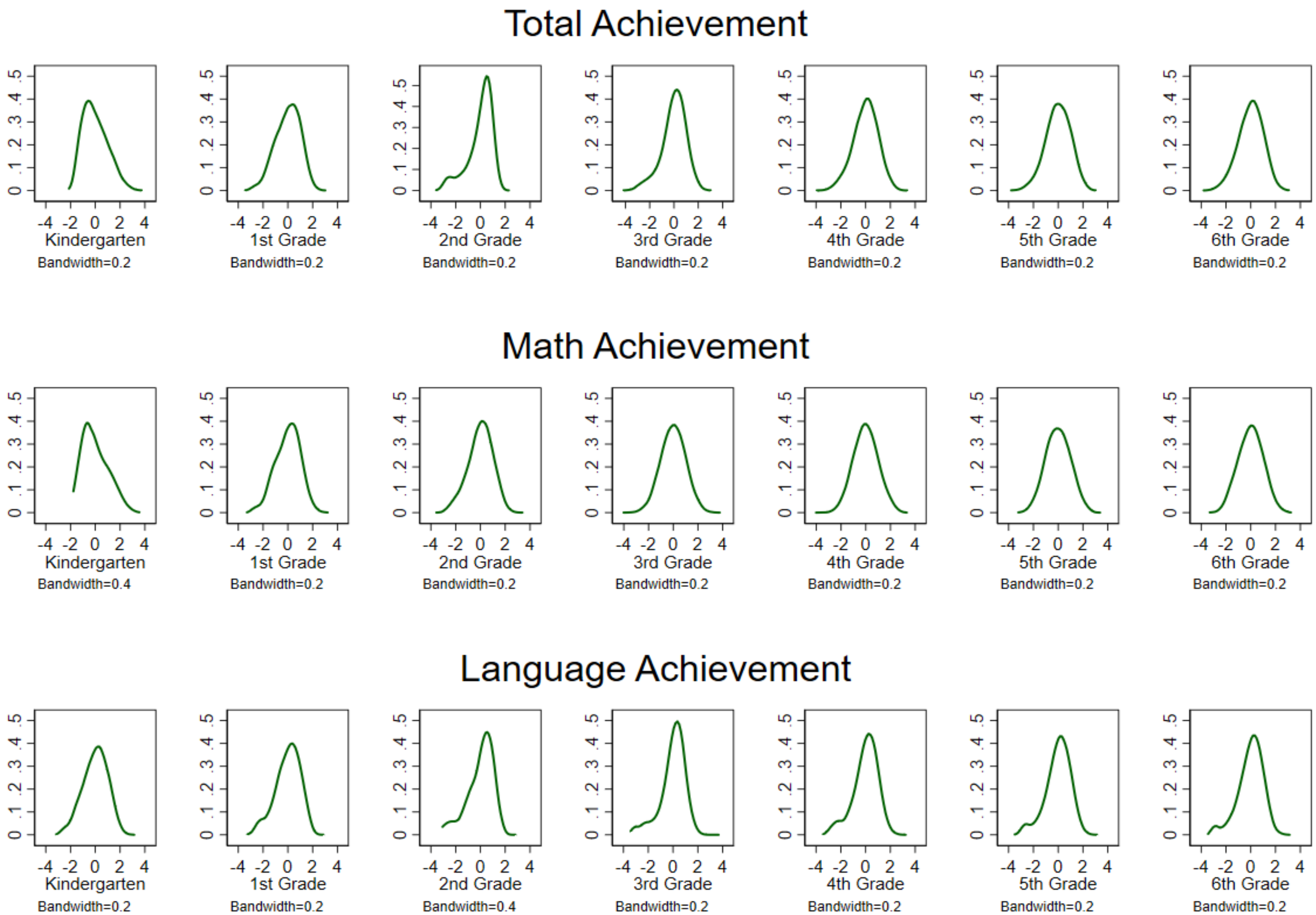
The working memory test had two parts. In the first part, children were given 2 minutes to find as many sequences of dog, house, and ball, in that order, on a sheet that has rows of dogs, houses, and balls in various possible sequences. The score on this part of the test is the number of correct sequences 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.[37] When the scores across the three dimensions are averaged, the distributions of the total executive function score are generally smooth and symmetric.

Figure B3, finally, 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).

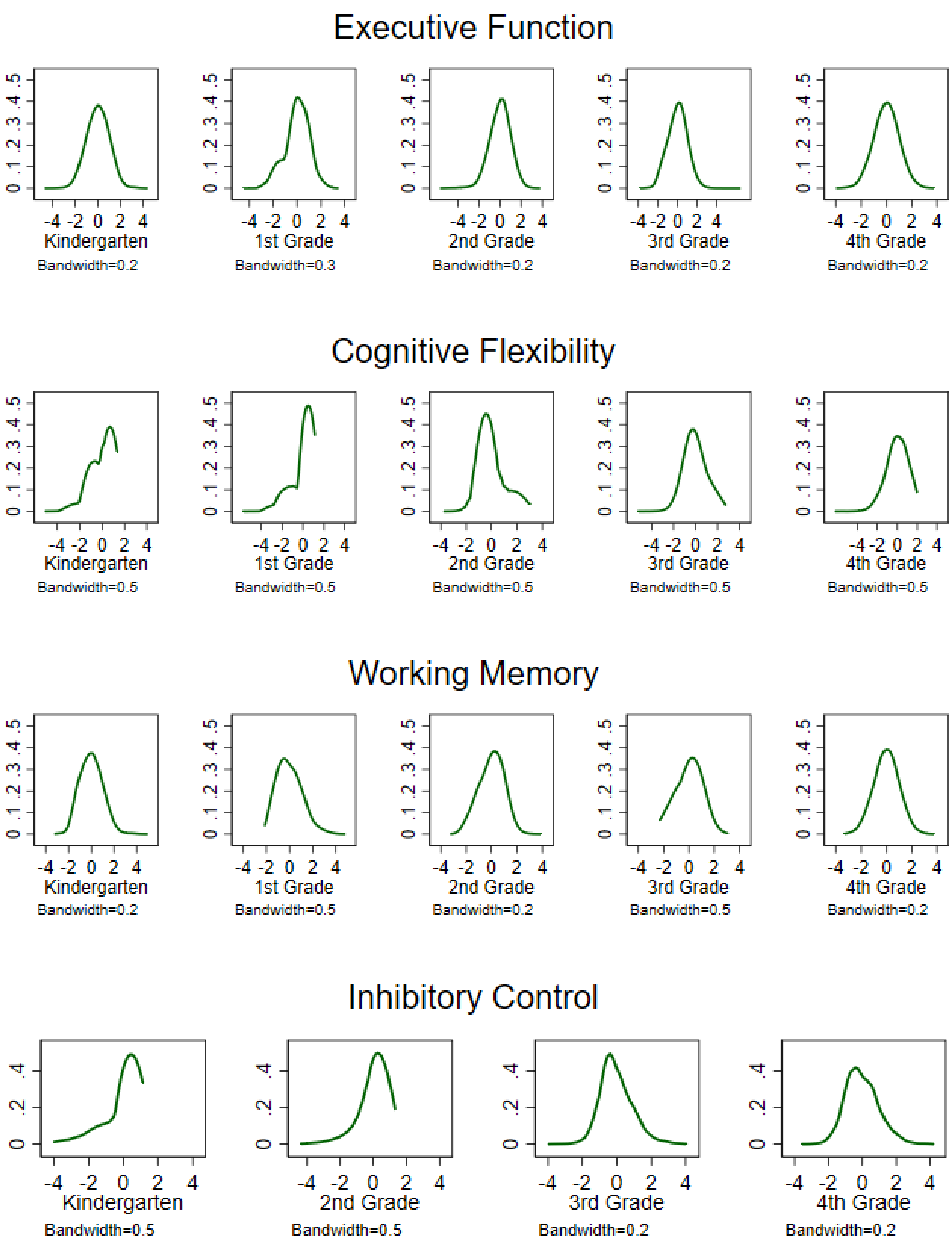
[37] The fact that these correlations are very low is likely to be a result of both measurement error and differences across the constructs that each domain measures.

Figure B1: Distributions of achievement, by grade



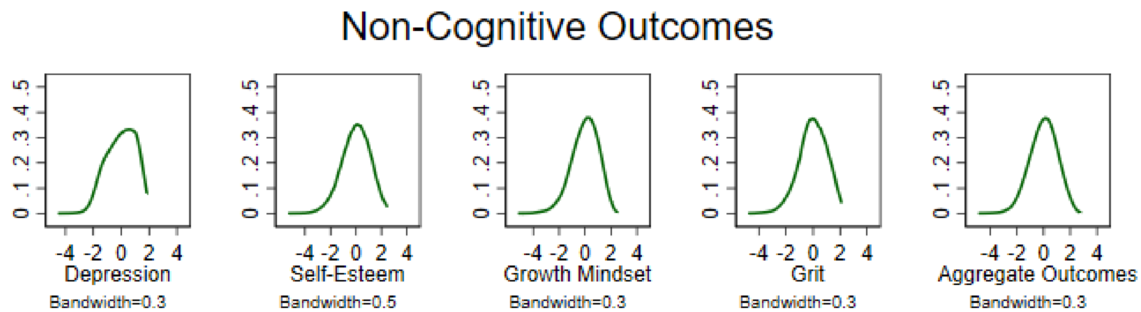
Note: The figure shows univariate densities of achievement, in z-scores, by grade.

Figure B2: Distributions of executive function, by grade



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

Figure B3: Distributions of non-cognitive outcomes, by grade



Note: The figure shows univariate densities of non-cognitive outcomes, in z-scores, by grade.

Table B1: Correlations across dimensions in executive function

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

Note: 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

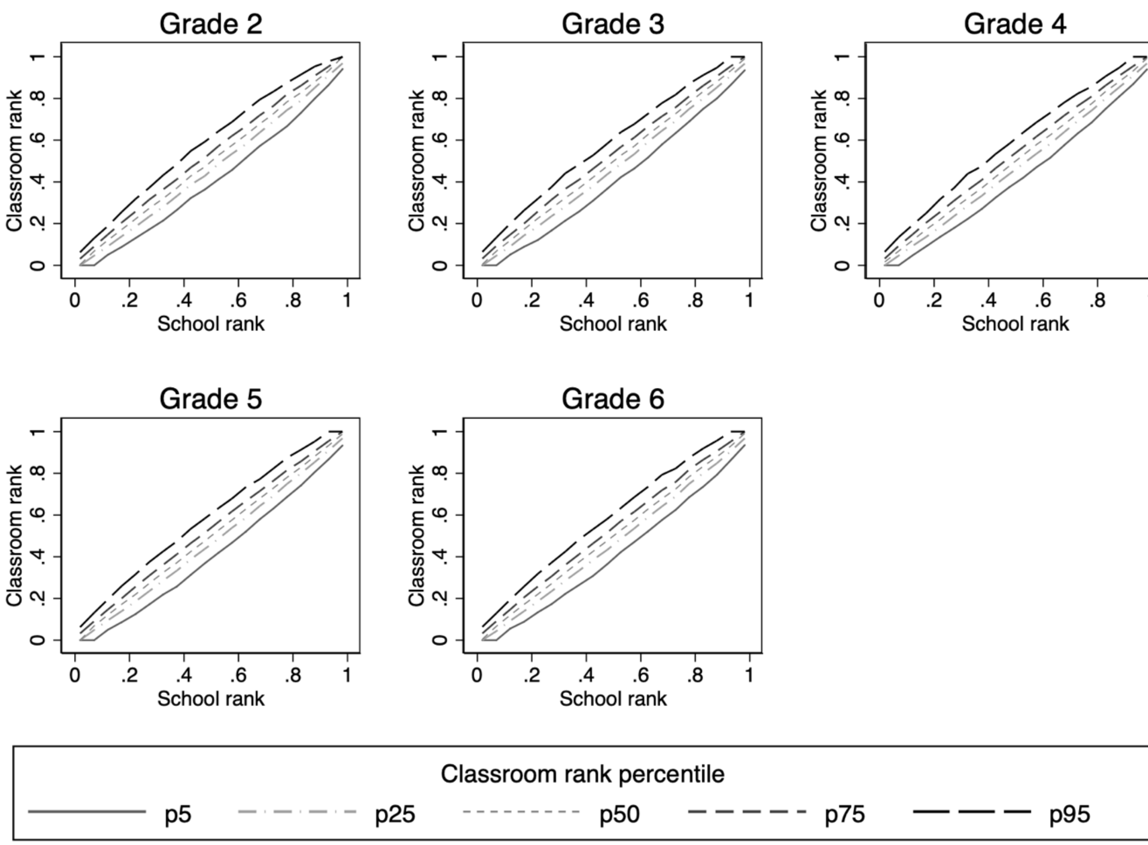
	Depression	**Self- Esteem**	**Growth Mindset**
Self- Esteem	0.24		
Growth Mindset	0.26	0.49	
Grit	0.20	0.45	0.38

Note: 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

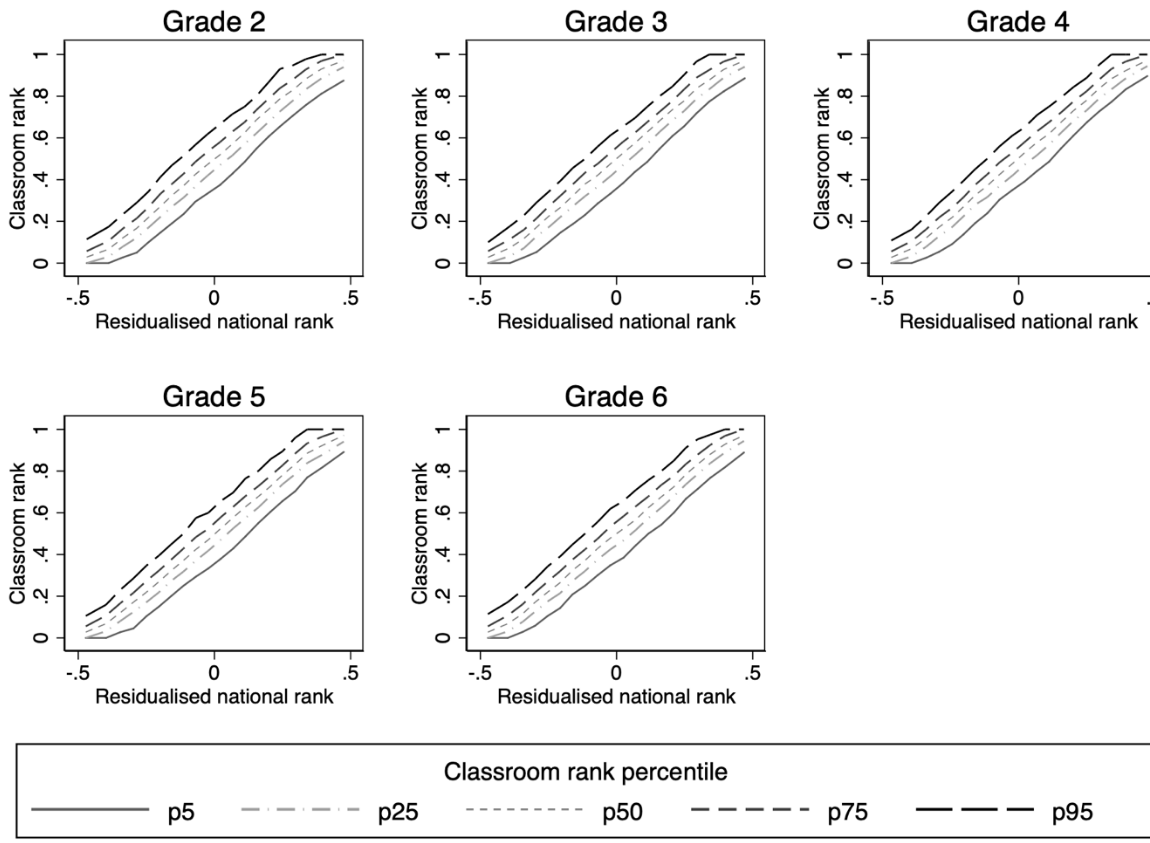
In this Appendix, we provide evidence of the variation in classroom achievement rank within school for Grades 2 to 6.

Figure C1: Percentiles of classroom rank by school rank, Grades 2 to 6



Notes: To construct this figure we plot the 5th, 25th, 50th, 75th and 95th percentile of classroom rank within each ventile of school rank, against the median school rank in that ventile, for each grade.

Figure C2: Percentiles of classroom rank by residualised national rank, Grades 2 to 6



Notes: To construct the figure, we first regress national ability rank on school fixed effects and derive residuals. Then we plot percentiles of classroom rank within each ventile of residualised national ability rank, against the median residualised national rank in that ventile, for each grade.

Appendix D

In this Appendix, we report classroom rank effects by grade (for grades 1, 2 … 6).

Grade-specific estimates of effects of classroom rank: Table d1 presents estimates of math classroom rank effects, by grade (rather than when grades are aggregated into "early", "middle" and "late" periods. These results are consistent with those in the first column of Table 6 in the main body of the paper, although they are noisier.

Table D1: Grade-specific estimates of effects of classroom rank

	Grade					
	1	2	3	4	5	6
Classroom rank	0.058***	0.026	0.011	0.069***	0.011	-0.020
	(0.022)	(0.019)	(0.018)	(0.016)	(0.015)	(0.015)
N	12,161	14,534	14,823	15,255	15,350	15,590

Notes: The table reports estimates from regressions of national rank in math on classroom math rank, separately by grade. All regressions are limited to schools in which there are at least two classes. All regressions include a third order polynomial in lagged national rank in math and school-by-grade fixed effects. Standard errors are clustered at the student level. *Significant at 10%, **significant at 5%, ***significant at 1%.