NBER WORKING PAPER SERIES

CLASS RANK AND LONG-RUN OUTCOMES

Jeffrey T. Denning
Richard Murphy
Felix Weinhardt

Working Paper 27468
http://www.nber.org/papers/w27468

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2020, Revised April 2021

We thank Sandra Black and Brigham Frandsen as well as participants of the Society of Labor Economists Annual Meetings, the IZA Junior/Senior Symposium, the CESifo Area Conference on the Economics of Education, the University of Utah Department of Finance, and the STATA Texas Empirical Microeconomics Conference for feedback and comments. Felix Weinhardt gratefully acknowledges financial support by the German Science Foundation through CRC TRR 190 (Project number 280092119). All errors are our own. Disclaimer: The research presented here utilizes confidential data from the Texas Education Research Center (ERC) at the University of Texas at Austin. The views expressed are those of the authors and should not be attributed to the ERC or any of the funders or supporting organizations mentioned herein. Any errors are attributable to the authors alone. The conclusions of this research do not reflect the opinion or official position of the Texas Education Agency, Texas Higher Education Coordinating Board, the Texas Workforce Commission, the State of Texas, or the National Bureau of Economic Research.

Class Rank and Long-Run Outcomes
Jeffrey T. Denning, Richard Murphy, and Felix Weinhardt
NBER Working Paper No. 27468
July 2020, Revised April 2021
JEL No. I20,I23,I28

ABSTRACT

This paper considers an unavoidable feature of the school environment, class rank. What are the long-run effects of a student's ordinal rank in elementary school? Using administrative data on all public-school students in Texas, we show that students with a higher third-grade academic rank, conditional on achievement and classroom fixed effects, have higher subsequent test scores, are more likely to take AP classes, graduate from high school, enroll in and graduate from college, and ultimately have higher earnings 19 years later. We also discuss the necessary assumptions for the identification of rank effects and propose new solutions to identification challenges. The paper concludes by exploring the tradeoff between higher quality schools and higher rank in the presence of these rank-based peer effects.

Jeffrey T. Denning
Department of Economics
Brigham Young University
435N CTB
Provo, UT 84602
and NBER
jeffdenning@byu.edu

Richard Murphy
Department of Economics
University of Texas at Austin
2225 Speedway, C3100
Austin, TX 78712
and NBER
richard.murphy@austin.utexas.edu

Felix Weinhardt
DIW Berlin
Mohrenstraße 58
10117 Berlin
Germany
fweinhardt@diw.de

1. Introduction

There is a large literature examining peer effects in education. This literature typically focuses on either the benefits of high-performing peers (Sacerdote, 2001; Whitmore, 2005; Kremer and Levy, 2008; Carrell et al., 2009; Black et al., 2013; Booij et al., 2017), or the negative effects of having disruptive peers (Hoxby and Weingarth, 2006; Lavy et al., 2012; Carrell and Hoekstra, 2010; Carrell et al., 2018). However, there is another mechanism whereby having lower-performing peers could *improve* student outcomes—namely a student's rank**.** [1] We will explore a student's ordinal rank within their classroom and the persistence of this effect on their outcomes into adulthood. Rank is an appealing attribute to study because it occurs naturally in any group of people.

We consider a student's academic rank in third grade (8 to 9 years old), conditional on their achievement, on short- and long-run outcomes. We use the universe of public-school students in Texas from 1994 to 2006 and combine this with an identification strategy that leverages idiosyncratic variation in rank. We find that a student's rank in third grade impacts grade retention, test scores, AP course taking, high school graduation, college enrollment, and earnings up to 19 years later.

Academic achievement and rank are highly correlated. So, to isolate the effect of a student's rank, we extend the method used in Murphy and Weinhardt (2014/2020) and leverage the idiosyncratic variation in the distribution of test scores across schools, subjects, and cohorts. Consider the following hypothetical: two students who have the same level of math achievement (as measured by their place in the state-wide distribution) in successive cohorts of the same size and mean attainment, at the same school. Because school cohorts are small relative to the state cohort, there will be variation in the test score distribution such that one student may be the fifth best student in their class while the other may be the eighth.[2] This is the idiosyncratic variation we leverage to identify the effect of rank, which we argue originates from sampling variation in the distribution of human capital within subjects and

[1] This can occur through various channels. These channels can be categorized as internal (learning about ability, development of non-cognitive skills) or external (parental and school investments).
[2] We standardize rank to the percentile within classrooms throughout the paper, a procedure which we discuss in Section 3.

cohorts of each school. This variation exists because these groups sample relatively small numbers of students, so that distributional differences emerge by chance.[3]

While this source of variation may appear simple, it is not. Later, we set out a hypothetical scenario to illustrate the identification challenges of estimating rank effects. Some of these challenges extend to the estimation of non-rank types of peer effects. Importantly, rank may be correlated with other features of the classroom distribution, such as mean.[4] We discuss this hypothetical setting and its relationship to our empirical strategy in detail in Section 2, and the assumptions required. One contribution of this paper is to outline the challenges of identifying the effect of rank, many of which extend to estimating other peer-effects.

We perform a battery of further robustness checks to establish the validity of the underlying assumptions, including balancing of rank on individual-level covariates, alternate functional forms, and definitions of rank. We also present estimates for small schools where there is likely to be one classroom per grade. We address any further concerns about measurement error in student test scores resulting in rank being a proxy for student human capital: we show that individual non-systematic additive measurement error in student test scores, which would be non–rank preserving, would cause a small downward bias to our estimates. Our exercise shows that to the extent that test scores are a noisy measure of human capital and student rank, our estimates should be interpreted as a lower bound.

We test for heterogeneous rank effects by gender, parental income, and race. We find the impact of rank on male and female students to be very similar for most outcomes. In contrast, we find that disadvantaged students (non-white or Free and Reduced-Price Lunch) are significantly more affected by rank than their advantaged counterparts. This is seen throughout a student's life, affecting eighth-grade test scores, high school graduation, college enrollment, and earnings.

---

[3] Hoxby (2000) is an early example of a study using sampling variation in the context of peer effect estimation. Importantly, the variation in the treatment of interest in the present paper (rank) occurs within a class, which means we can condition on class effects.

[4] This makes the identification of any specific type of peer effect nontrivial even with randomization.

A natural question to ask is: Why would third-grade rank, conditional on achievement, impact these outcomes? One possibility is that humans think in terms of heuristics (Tversky and Kahneman, 1974) and therefore rely on ordinal rank position rather than the more precise cardinal position within a group. Alternatively, ordinal rank may be easier to observe than cardinal position. Regardless of the precise behavioral origins of the effect, its impact can be seen in the findings that an individual's ordinal position within a group predicts well-being (Luttmer, 2005; Brown et al., 2008) and job satisfaction (Card et al., 2012), conditional on cardinal measures of relative standing. Hence, rank may also impact investment decisions and subsequent productivity. In the education literature, this is known as the Big-Fish-Little-Pond effect, where individuals gain in confidence when they are highly ranked in their local peer group (for a review, see Marsh et al., 2008).[5]

In this paper, however, we are agnostic about what is driving the effects. We define the rank effect to include any reactions to the rank of a student by any individual—the student, parents, or teachers. For example, if teachers invest more effort in the worst (or best) students, or parents invest in their child less if they believe their child is performing better than their peers irrespective of their absolute performance, then this would be included into the rank effect (Kinsler and Pavan, 2021). In summary, anything that is a reaction to a student's rank is a potential mechanism including student effort, parental investment, and teacher investment. In contrast, any predetermined factor that covaries with rank conditional on achievement could generate a bias. Note that, while we are agnostic about the mechanisms that give rise to the rank effect, we do take great care in establishing that the rank effect is not a product of confounding factors.

Recent studies have documented that a student's relative rank affects short-run outcomes independent of achievement. Murphy and Weinhardt (2020) document the effect of primary school rank in England, conditional on achievement, on high school test scores and confidence. Two papers by Elsner and Isphording apply the same idea to the United States using data from the National Longitudinal Study of Adolescent to Adult Health (AddHealth) to study effects of contemporaneous high school rank on high school completion, college

[5] This has been referred to as the invidious comparison peer effect by Hoxby and Weingarth (2006).

going (2017a), and health outcomes (2017b).[6] Zax and Rees (2002) consider the effect of peers and other student characteristics, including IQ and rank within high school on earnings at ages 35 and 53, using a sample of approximately 3,000 male students from Wisconsin. [7]

We make two contributions to this literature. The first is conceptual—we discuss identification of rank effects and threats to identification in ways that have not been discussed to date. These identification challenges affect all rank studies, including those that use randomization to classrooms. We also propose tests for and solutions to identification challenges that are common to all papers that estimate rank effects. These issues are also important for the estimation of many other peer effects. The second contribution is empirical; we extend the literature by using administrative data on three million individuals to look at the long-term effects of a rank at a young age (in third grade) on adult outcomes.

For our empirical contribution, we consider a student's rank at ages younger than have previously been considered (age 8–9) on outcomes up to 19 years later.[8] Our large dataset allows us to examine nonlinear effects of rank and we find that rank has a nonlinear relationship with outcomes in several instances. We also consider the effects of rank in elementary as opposed to high school. Studying the effects of rank in high school is interesting but conceptually different. Elementary school students are arranged in classrooms where students share the same teacher, peers, and curriculum for the majority of the day, whereas high school students do not share the same teacher, peers, or curriculum. Hence, rank may

[6] The AddHealth home-survey contains a sample of only 34 students for each school cohort, data which are used to compute a measure of rank (Elsner and Isphording, 2017a, 2017b). More recently, there is a set of working papers that estimates the impact of rank within college on contemporaneous outcomes in various countries (Elsner et al., 2018; Payne and Smith, 2018; Ribas et al., 2018, Ribas et al. 2020, Delaney and Devereux, 2019).

[7] A distinct literature has studied the introduction of relative achievement feedback measures in education settings. Azmat and Iriberri (2010) find that providing information on relative performance feedback during high school increases productivity of all students when they are rewarded for absolute test scores. In contrast, Azmat et al. (2019) find relative feedback in college causes significant short-run decreases in student performance, but no long-run effects.

[8] Murphy and Weinhardt (*2020*) consider the effect of rank on test scores 3 and 5 years later. Elsner and Isphording (2017a) consider rank in high school on college enrollment and graduation—approximately 2 to 8 years later. Elsner and Ishpording (2017b) consider the effect of rank on risky behavior as reported 18 months later. Murphy and Weinhardt consider rank at age 10–11. Elsner and Isphording consider rank at age 14–18.

affect factors such as course taking or peers, which complicates the interpretation of rank in high school.

Our findings contribute to a growing literature that documents how childhood conditions affect adult outcomes. These conditions include a child's health (Oreopoulos et al., 2008), where a child lives (Chetty et al., 2016), the quality of a student's teacher (Chetty et al., 2014), the size of a student's classroom (Chetty et al., 2011), the age of a student when they start school (Black et al., 2011), and the presence of disruptive peers (Carrell et al., 2018; Bietenbeck, 2020), among others. We add to this list that a child's rank in their third-grade classroom, independent of their achievement, has meaningful effects on education and earnings in adulthood.

We find that rank effects are larger for historically disadvantaged groups such as non-white students or students eligible for Free or Reduced-Price Lunch (FRPL). One implication of this is that, unlike linear-in-means peer effects, where moving students between groups would have no net impact, rearranging students with rank in mind could improve overall outcomes. However, this sort of exercise merits caution because the changes in classroom distribution will have general equilibrium effects not accounted for in this paper (Carrell et al., 2013).

The more practical implication of this finding is that programs that move disadvantaged students into "high quality" schooling sometimes come at a cost that they will then be among the lowest-ranked students. The extensive literature on selective schools and school integration has shown mixed results from students attending selective or predominantly non-minority schools (Angrist and Lang, 2004; Clark, 2010; Cullen et al., 2006; Kling et al., 2007; Abdulkadiroğlu et al., 2014; Dobbie and Fryer, 2014; Bergman, 2018; Barrow et al., 2020). Our findings would speak to why the potential benefits of prestigious schools may be attenuated among these marginal or "bussed" students. We explore the magnitude of the tradeoffs between rank and school quality in Section 7.

To examine any such potential rank/school-quality tradeoffs, we address the question of which school parents should select for their children, given the existence of rank effects. We find that, if parents were to choose schools solely on the basis of mean peer achievement, rank effects would reduce 39 percent of the potential gains for median-performing students in the state. In contrast, when choosing based on value added, rank effects only reduce the gains from choosing a better school by 12 percent.

The rest of the paper is structured as follows. Section 2 briefly sets out the empirical design. Section 3 describes the data, while Section 4 determines the appropriate specification. Section 5 presents the results. Section 6 performs a series of robustness tests, and examines heterogeneity. Section 7 discusses implications for school choice. Lastly, Section 8 concludes.

2. Empirical Design

Figure 1 illustrates the spirit of the variation we will use. It shows the math test score distributions of seven hypothetical elementary school classes. The classes share a similar test score distribution, each has a student scoring 22 and another scoring 38, and all classes have a mean test score of 30. In four of the classes there is a student scoring exactly 35, however due to the idiosyncratic variation in the test score distributions, each of these students has a different third-grade math rank. To compare ranks across classrooms of different sizes, we normalize rank to range from zero to one (see Section 3.2 for details). In this hypothetical, the ranks of the students with a score of 35 range between 0.7 and 0.9.

In this section we discuss how we use such variation to identify the effect of class rank. We first discuss an idealized hypothetical setting that illustrates the pitfalls of identifying rank effects, along with our solutions. Many of these pitfalls have not been well understood in the rank literature to date, and some of them apply to the peer effect literature more generally.[9] We will then discuss how our application departs from this hypothetical setting and how that motivates our choice of specification.

2.1. Challenges in identification

The first challenge is that rank is cross-sectionally correlated with other features of classrooms. For example, a student assigned to a low-performing class will have a comparatively high rank, given their ability. Consider the following hypothetical: students

[9] Even in this hypothetical we would consider all reactions to the rank of a student as part of the rank effect. This includes reactions of the students themselves, teachers, parents, or other students. Understanding what is driving the rank effect is interesting but not the primary focus of this paper.

with the same ability are randomly assigned to different classrooms. Sampling variation would lead to naturally occurring variation in rank for these students.

In the hypothetical setting with randomization to classrooms, our identical students would be assigned to classrooms that are balanced on all characteristics, *in expectation*. However, once students are assigned to classrooms, the classroom distribution of ability will be correlated with rank; for example, classroom mean ability will be negatively correlated with a student's rank. As a result, random classroom allocation is not sufficient for the identification of rank effects. By the same argument, other types of peer effects such as linear-in-means peer effects are also non-separable from rank effects using only randomization to groups.[10] Our strategy, therefore, needs to account for factors correlated with rank that also impact student outcomes (e.g. classroom mean ability). To do so, we compare outcomes of students with the same predetermined human capital but differences in rank due to sampling variation, while holding classroom characteristics (e.g. mean and variance) constant.

To be precise about the factors that can and cannot be accounted for, as well as the assumptions, we now set this hypothetical out more formally. We would want to estimate the following equation among students with the same human capital. Let Y_{ijsc} be the outcome of student i who attended elementary school j in subject s from cohort c. Let R be a student's rank and θ_{jsc} be an indicator for classrooms defined by a school-subject-cohort (SSC) grouping. Consider the following equation:

$$Y_{ijsc} = f\left(R_{ijsc}, \theta_{jsc}\right) + \epsilon_{ijsc} \tag{1}$$

Any impact of rank cannot be identified without additional structure about the relationship between classroom characteristics and rank. This type of assumption is required in all work on peer effects—features of the classroom distribution cannot arbitrarily interact with all other features. For instance, the classic linear-in-means approach assumes that mean peer quality affects all students equally and is additively separable. Similarly, estimating the effect of the presence of a disruptive peer generally also assumes that the effect of a disruptive

[10] By definition, there will be a negative relationship between peer ability and rank for a given student. The omission of a rank parameter would therefore attenuate any linear-in-means parameter (Bertoni and Nisticò, 2019).

peer is linear and the same for all students. Hence, we make the assumption that rank and classroom effects are additively separable, as in Equation 2 below.

$$Y_{ijsc} = f(R_{ijsc}) + \theta_{jsc} + \epsilon_{ijsc} \quad (2)$$

In order to identify the effect of rank in Equation 2, we need a conditional independence assumption (CIA), $E[\epsilon_{ijsc}|f(R_{ijsc}), \theta_{jsc}] = 0$. Stated differently, we require rank to be uncorrelated with unobserved features of the classroom or student—that is, $R_{ijsc} \perp U_{ijsc}$, where U_{ijsc} is an observed or unobserved determinant of future outcomes. How likely is this CIA to be met? An obvious violation would be if a student's rank were systematically correlated with classroom mean or variance. This example does not violate the CIA because of the inclusion of the classroom fixed effect, θ_{jsc}. Equation 2 ensures that a student's rank will be uncorrelated with all features of the classroom distribution that affect all students equally. For example, θ_{jsc} controls for commonly studied linear-in-means peer effects, the presence of a disruptive peer, and the variance of the classroom, among many other things.[11]

To this point, this model is similar to the equation that has been estimated in the rank literature. However, that equation does not account for the possibility of heterogeneous effects of the classroom distribution by prior ability (Booij et al., 2017). If there are heterogeneous effects of the classroom distribution by ability, the inclusion of class fixed effects would not account for them. This is because they only account for classroom features that impact all students equally, such as linear-in-means peer effects. We now propose and discuss our approach to dealing with heterogeneous effects of the classroom distribution by ability.

If there are heterogeneous effects of the classroom by ability that are correlated with rank, they need to be accounted for. Failing to do so could cause an omitted variable bias. To illustrate, we will show how rank can be correlated with the classroom distribution for a student of a given ability. In Panel A of Figure 3 there are two approximately normally distributed classrooms with different variances. Students at the red line have the same absolute level of achievement but have different rank. Generally, above (below) average

[11] The inclusion of θ_{jsc} reduces the unexplained variation in rank conditional on achievement. Appendix Figure 1 presents the extent of this variation by plotting math achievement de-meaned by school and cohort against class rank. Even when accounting for subject-school-cohort variation there is a large amount of variation in rank for a given test score. This variation exists throughout the achievement distribution.

students in a higher variance class would have a lower (higher) rank. Hence, a student's rank is correlated with the interaction of classroom variance and that student's ability. If the interaction of student ability and classroom variance has an effect on outcomes, this would not be accounted for in Equation 2, and there would be an omitted variable bias in the rank estimate. More generally, heterogenous effects of the classroom distribution by student ability that are not accounted for lead to an omitted variable bias.[12]

To combat this, the ideal comparison would be among classrooms with similar distributions, which would limit the potential for rank to be correlated with heterogeneous effects of the classroom distribution. However, if two classrooms were identical in every way, there would be no variation in rank and therefore no way to identify the rank effect. On the other hand, as soon as the two classroom distributions are different from each other, these differences may be correlated with rank, which could result in omitted variables bias.

This tension will motivate our choice of specification—we want to compare students in similar but not identical classroom distributions. Comparing students in similar but not identical distributions would ensure that variation in rank is idiosyncratic, rather than driven by specific features of the distribution. Comparisons among similar classrooms could be accomplished either by estimating Equation 2 separately on groups of distributions and aggregating the coefficients, or in a way that we will discuss later. Moreover, coefficient stability across such groups can indicate the empirical relevance of this form of omitted variable bias.

If other measures of the classroom ability distribution are systematically correlated with rank, even after making comparisons across similar classroom distributions, we think of them as part of what defines rank. This is because some features may be inherently linked with rank and cannot be separated. To repurpose a proverb, you cannot make an omelet without breaking eggs: some factors are always linked, and we call these the effect of rank.

[12] Booij et al. (2017) find evidence that in university study groups, high-prior-achieving students are not significantly impacted by the distribution of the classroom while low-prior-achieving students are positively affected by the mean and negatively affected by the variance. This would downward-bias our estimates. As a robustness test we estimate our effects on high-prior-achievement students.

Another way to look at this is that our approach requires something akin to an "exclusion restriction." Namely, after accounting for common impacts of the classroom, θ_{jsc}, and making comparisons among classrooms with similar distributions, any remaining effects of the classroom for a student (with a given ability) operate through rank. This may seem like a strong assumption, but we believe it is satisfied because we can account for many factors, including commonly studied peer effects and heterogeneous effects of the classroom on students. While we note the possibility of other classroom features correlated with rank that we do not account for, we are unaware of any candidate features from models of peer effects.

Similar exclusion assumptions are implicit in all empirical work that estimates peer effects from variation in the student achievement distribution, of both experimental and non-experimental nature. For example, when estimating the impact of mean peer attainment, a student in a class with higher peer attainment would also have a lower rank, even if classes are randomly assigned. Papers estimating this type of peer effect must make a similar exclusion restriction assumption—that other features of the distribution tied to the mean are not impacting the outcomes. However, they typically do not account for a student's rank in their estimation. Given that we go on to establish that rank does have an impact, and would be negatively correlated with mean achievement, there will be a downward bias on linear-in-means peer-effects estimates.

Finally, we could extend this hypothetical to students with different levels of human capital by estimating the following, where T is a measure of human capital.

$$Y_{ijsc} = f(R_{ijsc}) + g(T_{ijsc}) + \theta_{jsc} + \epsilon_{ijsc} \tag{3}$$

This estimating equation further assumes that rank, human capital, and classroom effects are additively separable. If this functional form is misspecified, it may cause rank to be correlated with omitted factors, leading to bias. As a result, we relax this additive separability assumption by allowing for interactions of classroom characteristics and human capital as well as classroom characteristics and rank.

Note that with inclusion of class effects, the measurement of individuals' human capital will capture any cardinal measures of individual-level position in the classroom ability distribution, such as distance to the mean, distance to the top/bottom of the distribution, and so on. Therefore, the rank parameter picks up information only about ordinal position, and not about cardinal position relative to the class mean. In other words, if cardinal measures of relative achievement were sufficient to explain student outcomes, the rank parameter would be insignificant. A significant rank parameter implies that humans additionally use ordinal rank information to make decisions.

To summarize, even in the ideal setting, there are significant challenges to identifying the effect of rank. These challenges are not overcome with randomization to classrooms.

Reality differs from our hypothetical scenario in two ways. We do not have random allocation to schools and classes, and, we do not have direct measures of predetermined human capital. We now set out the specifics of how these departures could generate a spurious correlation between rank and future outcomes to illustrate how our specification addresses them.

2.2. Non-random sorting

The first deviation from the idealized experiment is that there is substantial sorting of students to classrooms in the United States. We categorize sorting into two types: active sorting and passive sorting. We define active sorting when parents (or students) choose their classroom on the basis of exact rank. Passive sorting occurs when students are sorted to schools for reasons unrelated to their rank but that may generate a spurious correlation between rank and other factors. If students are actively or passively sorted to schools, then there is the risk of omitted variables producing a spurious rank effect.

Active sorting is only possible if parents can predict their child's rank at a school. However, in practice this would be difficult. Two students with the same achievement, at the same school, in different cohorts will often have meaningfully different ranks. Figure 2 shows the distribution of class rank that a student with median statewide achievement would receive in each of the 13 cohorts from 1995 through 2007, by each school-subject group. Each school-subject is one column on the horizontal axis, which we have scaled from 1 to 100 for ease of interpretation.[13] The school-subject groups are sorted such that median class rank of the state

[13] We collapse schools-subject values into 100 percentiles, but the conclusions are unchanged when using finer data.

median student is increasing, e.g. each school-subject group has up to 13 observations, and the medians of these groups are used to sort school-subject groups.

We plot the class rank of the state median student at given percentiles (10th, 25th, 50th, 75th, 90th) within the school-subject over the 13 cohorts. There is considerable variation within a school-subject group for students at the statewide median level of achievement. For example, median state students enrolled in a school at the 50th elementary school-subject group could have ranks as low as 0.4 (10th percentile) or as high as 0.6 (90th percentile) depending upon which cohort they enrolled in at the school. This variation in rank would make it very difficult to sort actively on the basis of rank, even if parents had such preferences, full information about the school environment, and exact measures of student human capital before choosing a school. Hence, we think that active sorting is unlikely to be a problem in our setting and in similar settings in the United States. [14]

Even though it is difficult for a parent to actively sort their child into a school so as to guarantee the child's precise rank, the gradient in Figure 2 shows that in some schools a student at the statewide median will consistently have a high class rank (right hand side), while in others they would have a low rank. If students with certain characteristics systematically attend schools with a certain test score distribution, then this is what we refer to as passive sorting.

Passive sorting could also generate a spurious rank effect. To see this, let us return to panel A of Figure 3, which shows two types of schools, both with achievement approximately normally distributed with the same mean, but such that Type A has high variance in student achievement while Type B has low variance. Further consider two types of students whose type impacts their future outcomes independent of rank, such as disadvantaged and non-disadvantaged. The red line shows two students whose achievement is the same distance above their classroom mean in their respective distributions. Students at this point in the high-variance class will have systematically higher rankings than students in the low-variance class. If students are randomly assigned to classes, then rank will be uncorrelated with student characteristics such as disadvantage. However, if non-disadvantaged students more often sort

[14] If motivated parents sorted on the basis of the mean attainment of peers, then this would be negatively correlated with student rank and would downward-bias the effect.

into distribution Type A and disadvantaged students more often sort into distribution Type B, then rank becomes correlated with student characteristics (in this case disadvantage) conditional on test scores and class fixed effects. This would generate a spurious rank effects due to the passive rank-sorting into certain types of schools. In this hypothetical example, student characteristics are correlated with rank and future outcomes. This highlights that distributions with different variance can generate a spurious relationship between rank and student characteristics through passive sorting.

This is a simple example where only the variance of the class differs. However, the principal of passive sorting can be applied to schools varying in any higher moment of the test score distribution. All that needs to occur is for students of a certain type and achievement to systematically attend schools with distinct distributions from other types of students. If this is the case, then rank can become correlated with student characteristics conditional on achievement. Figure 3, Panel B provides another illustration of this, where two students with the same achievement relative to the classroom mean will have different ranks due to the test score distribution. As in Panel A, students in school type B will have a higher rank for a given distance to the class mean. This is a problem if students of a certain type systematically attend a school with this type of distribution. One way to address this is to control for predetermined characteristics such as gender, race, and income to alleviate some of these concerns, as shown in Equation 4.

$$Y_{ijsc} = f(R_{ijsc}) + g(T_{ijsc}) + \theta_{jsc} + X_i\beta + \epsilon_{ijsc} \tag{4}$$

However, it is unlikely that researchers can control for all relevant student characteristics because many are likely to be unobserved. We believe it is nevertheless informative to study coefficient movements across specifications with/without student controls to quantify potential biases through the passive sorting mechanism. But ultimately, studying coefficient movements based on observable controls alone is not satisfactory.

Note that we have discussed passive sorting with regards to student characteristics; however, *any* factor that is systematically related to the test score distribution of the school may cause a spurious relationship between rank and the outcomes. For example, there is substantial segregation by income in Texas schools and so a low-income student and a high-income student with the same human capital are likely to have different school environments and distributions. We document this fact in Appendix Figure 9. In this figure, we show average classroom distribution characteristics for students from various demographic backgrounds (FRPL, race, and ethnicity). We calculate this by computing the average classroom median achievement, 90th percentile achievement, and so on for students of different characteristics. Notably, the average class median is lower for lower-income students, and also for black and Hispanic students. In general, Appendix Figure 9 shows that the average classroom that minority and low-income students experience has lower achievement.

As illustrated above, the inclusion of class fixed effects would not account for this passive sorting based on higher moments of the distribution. The intuition is that classroom fixed effects account for the effects of a higher-order moment (such as variance) that affect all students equally. However, the relationship between variance and rank depends on a student's achievement. For students at the mean level of achievement, higher variance does not affect rank, whereas for students at the 75th percentile of statewide achievement, variance is negatively correlated with rank. This is fine if there is random assignment of students to classrooms. However, if certain types of students are assigned to high-variance classrooms, this will generate a relationship between rank and student type. Hence, controlling for variance in an additively separable way does not address this problem.

Fortunately, there is a solution to passive sorting—researchers should make comparisons among classrooms of a similar type, that is with similar test score distributions. If distributions are similar, then there is less scope for features of the distribution to be systematically correlated with rank, leaving variation in rank to be more likely due to idiosyncratic variation.[15] To accomplish this, we modify our estimating equation by allowing $g(T_{ijsc})$ to vary by characteristics of the test score distribution, indicated by $g_d(T_{ijsc})$. [16]

$$Y_{ijsc} = f(R_{ijsc}) + g_d(T_{ijsc}) + \theta_{jsc} + X_i\beta + \epsilon_{ijsc} \tag{5}$$

[15] Note that if schools within a type all had exactly identical distributions, there would be no variation in rank for a given test score, hence we still require sampling variation to exist within types.
[16] Alternatively, Equation 4 could be estimated within groups of distributions and then aggregated across groups.

In doing so, this addresses passive sorting through making comparisons among similar distributions. If comparisons are among very similar distributions, conditional rank is unlikely to be correlated with student characteristics because sorting into classrooms of similar distributions is likely to be much smaller than sorting to classrooms of different distributions. This equation is closer to our hypothetical where students only experience a different rank due to sampling variation in the test score distribution, rather than systematic differences, because in the thought experiment, students are randomly assigned to classes. One way to achieve comparisons among similar distributions is to classify distributions of student achievement into groups based on characteristics of the distribution (for example mean and variance) and interact $g()$ with indicators for these groups of distributions.[17] Essentially this relaxes the assumption that the relationship between achievement and classrooms is additively separable.

Interacting achievement with features of the class distribution is our preferred specification. The identifying assumption is a conditional independence assumption which is that $\epsilon_{ijsc} \perp R_{ijsc}$ where R_{ijsc} is realized rank conditional on $g_d(T_{ijsc}), \theta_{jsc}, X_i\beta$.

2.3. Specification error

The conditional independence assumption can be restated slightly. Consider the following equation where η_{jsc} is specification error.

$$Y_{ijsc} = f(R_{ijsc}) + g_d(T_{ijsc}) + \theta_{jsc} + X_i\beta + \eta_{ijsc} + \epsilon_{ijsc} \tag{6}$$

If there is any specification error, we would need to assume that it is unrelated to rank, $\eta_{ijsc} \perp R_{ijsc}$.[18] One example of a violation of this assumption is if different schools map current achievement to future outcomes differently. As an example, in a high-achieving classroom high test scores would lead to more four-year college enrollment and less two-year college

---

[17] Another way would be to estimate Equation 5 separately by groups of distributions and aggregate the resulting coefficients. Because student achievement is measured in percentiles and is uniformly distributed, school mean achievement will be correlated to the skewness of the school distribution. Low (high) mean achievement classes will typically have positive (negative) skewness. The skewness of a distribution can be important for passive sorting; therefore, allowing for the impact of achievement to vary by mean achievement could remain important even with SSC effects due to the correlation of the mean and skewness.

[18] This assumption is necessary even when students of different abilities are randomly allocated to classrooms.

enrollment. In low-achieving classrooms, high test scores might lead to more two-year college enrollment. If human capital has a different relationship to future outcomes in one school versus another but is modelled as having the same relationship, this would introduce error into the model. If this error is correlated with rank, then our estimate of the rank effect would be biased.

The solution to specification error is to allow the mapping of current achievement to future outcomes to be very flexible. This can be thought of as a partial relaxation of the assumption of additive separability of the classroom effects and of past achievement. Allowing the way current achievement is mapped to future outcomes to be flexible removes specification error from the relationship between rank and future outcomes. Consequently, papers using this strategy should include robustness exercises where results are shown for different flexible specifications of achievement.[19] Stability of estimates for different flexibly specifications of achievement is consistent with $\eta_{jsc} \perp R_{jsc}$.

2.4. Measurement error in human capital

Last, we do not have direct measures of human capital, instead using student achievement in externally graded statewide examinations. We use third-grade test scores because it is the earliest achievement measure available. Ideally, we would rank students based on a before-school measure of achievement, so that it would be unaffected by their ranking in the class. This would be the most similar to the human capital measurement in our thought experiment. However, there is no statewide data on student ability available before third grade.[20] We choose the earliest measure to study how childhood conditions affect a range of (long-run) outcomes. It is possible to estimate the impact of later grade ranks on outcomes. However, as we find that future test scores are linearly increasing with previous rank, it means that later rank itself is an outcome and that rank is self-perpetuating. Hence, it is difficult to determine when a child's ranking has the largest impact.

[19] Again, this applies to the estimation of rank and of other types of peer effects, and includes settings that use classroom randomization.

[20] As we find the impact of rank is linear on future achievement, the effects of rank are rank preserving. Therefore using measures of human capital at the beginning or end of a grouping are equivalent. Rank within a group would impact the prior achievement parameters, but would not impact the rank parameters. We choose the earliest measure to minimize any concerns about rank effects contaminating the achievement metric.

One concern is that despite having the same absolute achievement measures (e.g. the same scores on the same test), these test scores are a noisy measure of true human capital. Measurement error could be at the individual level or at the class level. Individual-level measurement error would attenuate any rank effects, as is common in many settings.[21] In contrast, class-level measurement error in test scores could potentially generate a spurious relationship between rank and achievement. An example of a class level measurement error would be a disturbance on test day. This class shock would affect all students' test scores, creating measurement error in our measure of human capital, but would also be rank preserving. Such class-level measurement error that cannot correlate with rank is not a concern because of the inclusion of the classroom fixed effects which would account for such shocks. [22]

In summary, our preferred specification will address the problems in the estimation of the effect of rank. First, we demonstrate that active sorting on the basis of rank would be impractical, and if existent would likely downward-bias the rank effect. Second, we show that passive sorting can be addressed by comparing similar distributions. Third, we show that misspecification can be addressed by flexibly controlling for the mapping of current achievement to future outcomes. While we describe each of these issues separately, in reality they are likely to be interrelated. For example, students could be passively sorted to schools which have different mapping functions to future outcomes. Notably, addressing passive sorting will not necessarily solve misspecification. We return to determining the exact specification of Equation 6 in Section 4.

[21] Despite the measurement error in treatment being nonstandard, as it will be correlated nonlinearly with achievement, in Section 6.2 (and Appendix Figure 6) we show that adding measurement error to achievement and then recalculating ranks on this measure will lead to a downward bias.

[22] Class-level measurement error in test scores, which is rank preserving, would make rank a proxy for underlying ability, because the same absolute achievement measures in different schools will reflect different levels of human capital. In this case, larger measurement error would cause test scores to become less reliable, but would leave rank unaffected, which would lead to rank acting as a proxy for underlying student human capital and generate a spurious "effect" of rank. Since researchers always have imperfect measures of human capital, this provides an additional motivation for including class fixed effect.

3. Data

This study uses de-identified data from the Texas Education Research Center (ERC), which contains information from many state-level institutions.[23] Data concerning students' experience during their school years cover the period 1994–2012, although the primary estimating sample will focus on the 13 cohorts of 1995–2007. These data contain demographic and academic performance information for all students in public K–12 schools in Texas provided by the Texas Education Agency (TEA). These records are linked to individual-level enrollment and graduation from all public institutions of higher education in the state of Texas using data provided by the Texas Higher Education Coordinating Board (THECB).[24] Ultimately, these records are linked to students' labor force outcomes in years 2009–17 using data from the Texas Workforce Commission (TWC). This contains information on quarterly earnings, employment, and industry of employment for all workers covered by Unemployment Insurance (UI).[25]

3.1. Constructing the sample

The sample used for this analysis consists of students who took their third-grade state examinations for the first time between 1995 and 2007.[26] We focus on students taking their third-grade exam for the first time to alleviate concerns regarding the endogenous relationship between class rank and previous grade retention. We focus on students taking their exams in English, rather than Spanish. During this period, third-grade students took annual reading and math assessments, although the testing regime changed.[27] Consequently,

[23] For more information on the ERC, see https://research.utexas.edu/erc/

[24] We do not observe out-of-state enrollment or enrollment at private institutions in Texas. Hence, if higher rank tends to cause students to leave the state, we will underestimate the effect of rank on college attendance.

[25] Unemployment insurance records include employers who pay at least $1,500 in gross earnings to employees or have at least one employee during twenty different weeks in a calendar year regardless of the earnings paid. Federal employees are not covered. We do not observe out-of-state employment. Hence, if higher rank tends to cause students to leave the state, we will underestimate the effect of rank on earnings.

[26] A student is defined as taking their third-grade exam for the first time if the student was observed not being in the third grade in the previous year.

[27] Until 2002, the Texas Assessment of Academic Skills (TAAS) was used. Starting in 2003, the Texas Assessment of Knowledge and Skills (TAKS) was used. The primary differences had to do with which grades offered which subject tests. This does not affect this study substantively as all students took exams in math and reading for third and eighth grade.

we percentilize student achievement by subject and cohort. This ensures that the test score distribution for each subject is constant and uniform for each cohort.[28] For each student, we generate a rank within their elementary school cohort for math and reading based on their test scores, including those who had been retained in their grade.

We link students to subsequent outcomes, including performance in reading and math in eighth grade. Because rank may affect grade retention, we estimate the impact on eighth grade test scores only for students who took the test on time. We also consider classes taken in high school, including Advanced Placement (AP) courses, and graduation from high school. We then consider whether students enroll in a public college or university in Texas (separately by two-year and four-year schools), if they declare a Science Technology Engineering or Mathematics (STEM) major, and whether they graduate from college. Lastly, we examine earnings and the probability of having positive UI earnings.

For binary outcomes, such as AP course taking, high school graduation, and college enrollment, we define the variable as 1 for the event occurring in a school covered by our data and 0 otherwise.[29] For earnings, we consider average earnings including zeroes as well as excluding zeroes.

To maximize the sample, we consider as many cohorts as possible for each outcome. This means that we have more cohorts for outcomes closer to third grade and fewer cohorts for later outcomes.[30] For K–12 and initial college attended outcomes, we have 13 cohorts of students who took their third-grade tests between 1995 and 2007, giving 6,117,690 student-subject observations. For graduating college in four years we have 10 cohorts (1995–2004), totaling 4,573,672 student-subject observations. For graduating in six years and post-college outcomes, for individuals aged 23–27 we have 8 cohorts (1995–2002), or 3,597,340 student-subject observations, while we have 6 cohorts, or 2,647,240 students, for graduating with a BA within eight years. This explains the discrepancy in sample size across different outcomes.

[28] Murphy & Weinhardt (2020) show that when estimating rank effects, conditional baseline achievement should be uniformly distributed to prevent the possibility of a certain type of measurement error (one that increases multiplicatively further from the mean), generating a spurious rank effect.

[29] As an example, for college enrollment a student who does not attend any college, or attends a college out of Texas, will be coded as a zero.

[30] Results are similar when considering a consistent set of cohorts; see Appendix Figure 4.

Table 1 presents summary statistics. The sample is 47 percent white, 35 percent Hispanic, and 15 percent black. Seventy percent of students in the sample eventually graduate from a Texas public high school, while 46 percent of students attend a public university or college in the year after "on-time" high school graduation.[31] Within three years of on-time high school graduation, 23 percent attend a public four-year institution in Texas, and 31 percent attend a Texas community college. When students are 23–27 years old, 65 percent have non-zero earnings, where the average non-zero earnings income is $24,818 in 2016 dollars.

3.2. Rank measurement

We rank each student among their peers within their grade at their school according to their state percentile in standardized tests in each tested subject.[32] Put simply, a student with the highest test score in their grade will have the highest rank. However, a simple absolute rank measure would be problematic, because it is not comparable across schools of different sizes. Therefore, as with state test scores we will percentilize the rank score for individual *i* with the following transformation:

$$R_{ijsc} = \frac{n_{ijsc}-1}{N_{jsc}-1}, \qquad R_{ijsc} \in [0,1]$$

where N_{jsc} is the cohort size of school *j* in cohort *c* of subject *s*. An individual *i*'s ordinal rank position within this group is n_{ijsc}, which is increasing in test score. Here, R_{ijsc} is the standardized rank of the student that we will use for our analysis and equals the percentile rank in their class. We refer to a student's rank within their school-subject-cohort (SSC) as their class rank.[33] For example, a student who had the second-best score in math from a school cohort of twenty-one students ($n_{ijsc} = 20$, $N_{jsc} = 21$) will have $R_{ijsc} = 0.95$. This rank measure will be approximately uniformly distributed, and bounded between 0 and 1, with the lowest- (highest-) ranked student in each school cohort having $R = 0$ (1). In the case of ties in test scores,

[31] On-time graduation is defined as graduation if a student did not repeat or skip any grades after third grade.

[32] Rank effects likely operate not only though the score on this particular exam, but also through more general performance in the class over a longer period of time. We are interested in the effect of rank in academic achievement and so use third-grade test scores as a convenient summary measure of class interactions.

[33] This is done because there are no data on class assignment during our analysis period. In Figure 11 we present estimates using a sample of schools with 30 of fewer students per grade. Here an SSC grouping would represent a single class. Our parameter estimates are not meaningfully changed by using this subsample.

each of the students with the same score is given the mean rank of all the students with that test score in that school-subject-cohort.[34] We will calculate this rank measure for each student within a test administration.

Note that this is our measure of the academic rank of a student within their class. Students will not necessarily be told their class rank in these exams by their teachers, nor do we believe that students care particularly about their ranking in these low-stakes examinations. Rather, we interpret our test score rank measure as a proxy for students' day-to-day academic ranking in their class. Students learn about their rank through repeated interactions throughout elementary school with their class peers (e.g., by observing who answers the most questions or gets the best grades in assignments). Similar arguments can be made for teachers or parents learning about the rank of students.

4. Model Specification

As described in Section 2, it is imperative that the correct functional form is used to prevent specification error or passive sorting from causing a spurious rank effect. Therefore, a key choice is the proper way to model $g()$. There are two features that this function should meet: it should 1) make comparisons among similar classrooms to avoid passive sorting and account for potential heterogenous effects of achievement and classroom distributions; and 2) flexibly model the relationship between current achievement and future outcomes to avoid specification error. Put another way, what specification satisfies the conditional independence assumption and yields rank that is conditionally as good as random?

A natural way to test for passive sorting, similar to a balance test in a randomized controlled trial, is replacing Y_{ijsc} in our specification with a predetermined characteristic. If the conditional independence assumption is satisfied, predetermined characteristics should not be correlated with rank. If observable predetermined characteristics are not correlated with rank, then it plausible that unobserved predetermined characteristics are also not correlated with rank.

[34] Our main analysis limits the sample to students who took their third-grade test on time. However, their actual classroom consists of students who are on time and students who are not. In Appendix Table 1 we show that our results are very similar when we calculate rank using all students and using different methods to break ties.

We first start with a specification akin to Equation 4, omitting student characteristics, with a flexible function of achievement, $g(T_{ijsc})$, which does not vary by school. Specifically, we include 19 indicators for all but the tenth ventiles of student achievement. This allows for nonlinear relationships between prior test scores and future outcomes.[35] To allow for comparison across classes with similar distributions we then interact student achievement with indicators for the school's location within the distribution of schools according to mean and variance in achievement, akin to Equation 5.

Table 2 shows the correlations between student rank and the following student characteristics: Male, Low Income (FRPL proxy), English as a Second Language, race/ethnicity categories. Each panel shows the correlations for a different specification: Panel A shows constant impacts of ventiles of achievement, which then we allow to vary by quartiles of school mean achievement (Panel B); deciles of school mean achievement (Panel C); quartiles of school variance in achievement (Panel D); deciles of school variance in achievement (Panel E); and the interactions of the quartiles of school mean and variance in achievement (Panel F).[36] Panels A–E show that rank is correlated with predetermined student characteristics. This suggests that these specifications have issues with passive sorting. However, Panel F shows that interacting 16 indicators for school variance and mean quartiles eliminates any imbalance we have. Therefore, our preferred specification is

$$Y_{ijsc} = \sum_{r=1,r\neq 10}^{20} \mathbb{I}_n(R_{ijsc} = r)\,\rho_r + \sum_{D=1}^{16} \sum_{t=1,t\neq 10}^{20} \mathbb{I}_n\big(T_{ijsc} = t\big)\,\mathbb{I}_d(d_s = D)\mu_{nd} + \theta_{jsc} + X_i\beta + \epsilon_{ijsc} \qquad (7)$$

where $\mathbb{I}_n(R_{ijsc} = r)$ denotes an indicator function, which takes a value of one when the ventile of class rank R_{ijsc} matches r. This allows for potential nonlinearities in the effect of rank on later outcomes by estimating parameters (ρ_r) for each ventile in rank (omitting the tenth ventile). We have defined $g_d(.)$ in a similar manner, by allowing for separate impact for each

[35] In Figure 11 we show our results are robust to alternate functional forms. We chose ventiles as our main specification as we did not want to allow the rank function to have more flexibility than that of baseline achievement.

[36] Using school mean has two advantages. First, it allows a different mapping from achievement to future outcomes in different settings, which deals with the issues of misspecification. Second, due to the bounded nature of test scores, the mean is generally correlated to the skewness of a distribution.

of the ventiles of baseline achievement T_{ijsc}, which can vary at the school distribution level, d_s. Schools are characterized by their quartiles in the distributions of mean achievement and variance, providing 16 school distribution groups, D.

Appendix Figure 2 shows the remaining variation in class rank for a given test score within each of these distributions. This figure plots the class rank of students against their achievement de-meaned by school within a subject, for each of the 16 types indexed by D. The relationship between rank and achievement resembles that presented in Appendix Figure 1, but we now see that as the mean and variance of classes change, the expected rank for a given test score also changes. For example, distributions in the top row (with the smallest class variance) have a shallower gradient than those in the fourth row. Regardless, there still exists sufficient variation in rank conditional on achievement to estimate the effects of rank even within these distribution types.

In summary, using this specification (Equation 7) we do not observe imbalance in observable characteristics, and so assume the remaining variation in rank to be likewise orthogonal to unobservable factors that determine our outcomes. We show that our main findings are robust to these choices in Section 6.1 below. We now present our main findings using Equation 7 as our model.[37]

5. Results

We will primarily present results for Equation 7 by plotting the estimate coefficients, ρ_n, along with the 95 percent confidence intervals. All estimates will be relative to the tenth ventile that includes students ranked from the 45th to the 50th percentile in their class. Because there are many estimates for each outcome, we present the results visually.[38]

5.1. K–12 outcomes

We first consider the probability of repeating third grade. Panel A of Figure 4 shows that lower-ranked students are more likely to repeat third grade even after conditioning on

---

[37] Note that we observe two test scores, one each for students' math and reading. For most outcomes we stack observations so that each student has two observations. For outcomes that are subject specific, we estimate specifications that have both math and reading rank and achievement entered separately to investigate whether rank in math and reading have different effects.

[38] The corresponding estimates and standard errors are available in table form upon request.

achievement. The impact of rank on retention decreases as the rank of the student increases. Those in the lowest ventile of the rank distribution are 6 percentage points more likely than the median ranked student to be retained, this retention effect halves for the second lowest ventile. Moving a student from being ranked last to being ranked in the 25th class percentile reduces the probability of retention by roughly 5 percentage points. This is approximately twenty times as large as the increase in probability of retention associated with FRPL (0.003) or ESL (0.002) status. Given the mean retention rate of 1.6 percent, this represents a sizable shift in how students are treated by their schools, independent of their human capital. Once students reach the 40th class percentile, rank has no significant impact on retention. This suggests that at least some of the rank effect is likely coming from the way the school treats low-ranked students.

We next examine the effect of third-grade rank on achievement in the eighth grade as measured in state test score percentiles. Figure 4 Panel B shows an approximately linear effect of rank in the third grade on academic performance. Moving from the 25th percentile to the 50th percentile in rank (or from the 50th to the 75th) improves grade eight performance by approximately 2.5 percentiles. For comparison, students with non-FRPL have on average 4 percentiles higher achievement than their FRPL counterparts, and which is the same as the gap between White and Non-Hispanic Black students.

This is similar to the estimates in Murphy and Weinhardt (2020) that consider outcomes at comparable ages in England, finding that the same change in rank at the end of primary school (age 10/11) improves performance in national tests at age 13/14 by 1.9 percentiles. This is estimated on students who took the test in eighth grade "on time." As we have seen, however, rank causes some students to be retained and will also determine whether students are included in this estimating sample.[39] Hence, when interpreting the estimates of the effect on eighth-grade test scores, one should bear in mind that the sample is selected on the basis of the treatment.[40]

[39] Appendix Figure 3 illustrates that rank affects the probability of taking the eighth-grade tests on time.

[40] To bound the estimates, one could consider not taking the test on time as the lowest possible score; our estimates would then be a lower bound of the effect of third-grade rank on eighth-grade test scores.

The results for rank's impact on eighth-grade test scores are not novel, but they do corroborate that similar rank effects occur in different educational systems, establishing the external validity of each estimate. Moreover, they provide a mechanism for the later outcomes we observe. In particular, student achievement in eighth grade is correlated with many outcomes including high school achievement, class taking, college enrollment and success, and labor market outcomes.

The first high school outcome we consider is whether a student takes Advanced Placement courses. In the first two panels of Figure 5 (A and B), we use our standard specification where the two observations for math and reading for each student are stacked, such that we are estimating the mean rank coefficients across subjects. We see that elementary school achievement rank positively impacts the probabilities of taking AP Calculus and AP English. In both cases, these effects are driven by being highly ranked among classmates during elementary school. The pattern and magnitude of the estimates are similar for taking both AP subjects, with students around the 75th percentile being around 2 percentage points more likely to take AP compared with the median student. Note that the baseline rates for taking AP Calculus and AP English for our sample are 8.4 percent and 19.0 percent respectively. The exception that there is a discontinuous jump in the probability of taking AP Calculus if the student is in the top ventile of their elementary school. Students in the top ventile are 6 percentage points more likely to enroll in AP Calculus than the median-ranked student.

The second two panels of Figure 5 (C and D) consider the effect of elementary school rank in each subject. To do so we control for achievement and rank in third-grade math and reading separately and simultaneously. Panel C shows the impact on taking AP Calculus. A higher rank in math causes more students to take AP Calculus. Most of this effect occurs for students above the median in rank, whereas below the median there are only small differences in the probability of taking AP Calculus. In contrast, a student's rank in reading has very little effect on taking AP Calculus for students with rank above or below the median. The final panel (Panel D) shows the impact of third-grade rank on taking AP English courses. As before, rank in math has a stronger effect than rank in reading. Here, however, having a high rank in reading does positively impact the probability of taking AP English courses. This is evidence that some rank effects are subject specific while some have spillover effects into other subjects.

The final set of K–12 outcomes we consider is whether a student graduated from high school. The time frame we consider is within three years of on-time high school graduation, defined as graduating nine years after third grade to avoid issues of grade retention. In our sample, 70 percent graduate from high school by this definition (i.e. at most twelve years after third grade). The impact of third-grade rank on high school graduation can be seen in Figure 6 Panel A. A higher rank makes students more likely to graduate from high school. The effect is approximately linear, with impact of rank gradually declining as rank increases. Moving from the 25th to 75th percentile in rank increases the probability of graduating from a public Texas high school by 4 percentage points. In comparison, Asian students are 4.6 percentage points more likely to graduate than White students, and FRPL students are 11 percentage points less likely to graduate than non-FRPL students.

In summary, a student's rank in third grade independently affects grade retention, test performance five years later, class selection, and ultimately graduation. As we examine longer-term outcomes, these changes throughout schooling will be some of the channels that affect outcomes such as college education and earnings.

5.2. College outcomes

After high school, students face the choice of whether to enter post-secondary education. Figure 6 Panel B presents enrollment in any public college in Texas. Rank has a largely linear effect on the probability of enrolling in college in Texas. Moving from the 25th to 75th percentile in rank leads to an approximately 4-percentage-point increase in college-going. However, there is a discontinuous fall in the likelihood of attending college if students were in the lowest ventile, compared to the second-lowest-ventile of 1.5 percentage points. This is the same magnitude as the FRPL coefficient.

To understand this pattern better, we consider enrollment in two-year and four-year schools separately in Figure 6 Panels C and D. Figure 6 Panel C considers enrollment in two-year institutions. Having a low rank reduces the probability of going to community college, whereas high rank is not strongly correlated with attending community college. Again, there is a discontinuous fall in the likelihood of attending college if students were in the lowest ventile: those in the bottom ventile are 4 percentage points less likely to enroll in community college compared the baseline enrollment rate of 31 percent. Figure 6 Panel D shows the impact on enrollment in a four-year institution. Enrollment is increasing in rank, and this is driven by primarily by students ranked above the median with the effect of rank increasing in rank. The contrast of the effects of enrollment by school type is likely driven by who is at the margin of attending a community college versus a four-year institution.

Once at college, students can declare a major. Given the significant financial returns to STEM majors, we now estimate the ultimate impact of third-grade rank on the probability of a student declaring a STEM subject as their first major. Figure 7 Panel A shows that there is a comparatively small positive relationship between students' rank in elementary school and their likelihood of choosing a STEM major, an effect which is driven largely by top-ranked students.[41] Like AP choice in high school, major choice is likely to be impacted by students' rankings in particular subjects, and this previous estimate is the average effect of both reading and math ranks. To explore this, Panel B of Figure 7 presents the impacts of third-grade math and reading rank on declaring a STEM major. Here we find that the relationship is entirely driven by math rank, with top-ranked math students over 2 percentage points more likely to choose STEM over the mean STEM uptake rate of 4 percent. In contrast, neither top or bottom reading rank impacts the probability of declaring a STEM major.

Continuing with post-secondary outcomes, the panels in Figure 8 consider graduation with a bachelor's degree within various time frames—4 years (Panel A), 6 years (Panel B), and 8 years (Panel C) after "on-time" graduation from high school. We find that students with higher rank in elementary school are more likely to graduate with a bachelor's degree. Consistent with the effects on enrollment, the effect of rank is largely driven by rank above the median. The impact of being top of class in terms of percentage points is smallest for graduating within 4 years at 2.4 percentage points relative to the median-ranked student, compared to 3.4 and 3.3 percent for 6 and 8 years respectively.

[41] We code students who declare a STEM major as 1 and students who do not (including students who do not enroll in college) as 0. Hence, estimates will conflate the effect of rank college enrollment and declaring a STEM major.

5.3. Labor market outcomes

We consider the effects of rank on a range of labor market outcomes. First, we examine employment outcomes for students aged 23–27 (or 15–18 years after third grade). We consider average annual earnings between the ages for eight cohorts of students who took their third-grade examinations between 1995 and 2002 (employment years 2009–2016).[42] The first panel in Figure 9 considers the probability of having positive earnings at age 23–27. The outcome is the proportion of years between 23-27 with a positive earning measure. A value of 1 would mean having positive earnings recorded in each of these years, a value of 0 would mean having no recorded earnings between ages 23 and 27. The pattern in Panel A suggests there is little effect of rank on the probability of having positive earnings except for the lowest-ranked students, who are 1 percentage point less likely to record positive earnings than the median-ranked student.[43]

The remaining panels of Figure 9 refer to the effect of rank on earnings. Panel B shows that increasing rank increases average annual earnings between the ages of 23 and 27. Low-ranked students have meaningful earnings penalties, earning on average $1,850 per year less than the median-ranked student. High-ranked students likewise see increases in earnings, with the highest-ranked students earning $1,200 more per year on average than the median student. The lower two panels show the effect on non-zero earnings and log earnings (Panels C and D respectively). Conditional on having positive earnings, we find the negative impact of having a low rank is exacerbated. Students at the bottom now earn $1,250 less per year, compared to students in the middle of the rank distribution.

With regards to the log of average annual earnings, we see that rank affects average earnings throughout the distribution of rank. Taken together, a student's rank in third grade affects labor market outcomes. Moving from the 25th percentile to the 75th percentile in rank causes log earnings to increase by approximately 7 log points. This is comparatively smaller in magnitude than the parameter for other student characteristics. For example, conditional

[42] Note, not all individuals have four years over which to average employment. Those from the most recent cohort only have one year of labor outcomes, for instance. We average over all years from age 23 to 27 for which we have earnings. Appendix Figure 4 Panel D shows the impact on log earnings for the 8 cohorts of students for which we have all five years of earnings (who took third grade 1995–2002).

[43] This shows rank does not cause differential attrition from our earnings sample.

on third grade achievement and SSC effects males on average earn 20 log points more than females and Non-Hispanic Blacks earn 10 log points less than Whites. However, it is important to consider that these characteristics are contemporaneous with employment, where as a student's academic rank in third grade occurred up to 19 years ago.

6. Robustness

In Section 2 we discussed the concerns relating to correctly specifying the functional form so that omitted factors are not driving the results. In Section 4 we used balancing regressions to empirically inform the choice of functional form for our regressions. In this section we show that our results are robust to several alternative specifications, as well as to different samples, rank definitions and measurement error in baseline ability.

6.1. Functional form choices

First, we model the relationship between baseline achievement and outcomes using various functional forms. Figure 10 presents results for four main outcomes: eighth-grade test scores, graduation from high school, the probability of attending any college, and log earnings. The point estimates are displayed for various functional forms of student achievement, interacted with sixteen indicators for school mean and variance quartiles. In addition to our main specification (which controls for achievement using ventiles in student achievement interacted with indicators for classroom distribution type), we model student achievement using various polynomials from first order to a sixth-order interacted with indicators for classroom distribution type. The results are substantively similar once achievement is controlled for with a quadratic—the magnitude does vary somewhat across specifications, but the direction of the relationship is consistent.[44] Hence, our results showing that rank has a lasting impact on student outcomes are not dependent on the exact functional form chosen to model achievement.

In addition to choices of nonlinearity in $g_d()$, we also allow of the impact of prior achievement to vary by the school achievement distribution. We have shown in Section 4 and

[44] The fact that only allowing for a linear impact of percentile rank leads to different estimates may be reflective of the predetermined human capital distribution being normally distributed, which we have transformed into a uniform distribution through percentalization.

Table 2 that the interaction of achievement with 16 indicators for school mean and variance results in observable covariate balance. As a result, we use this as our main specification. However, we show that our results are not dependent on this specific functional form $g_d()$ for most outcomes.

Appendix Table 2 repeats the six interactions of achievement with ways of classifying school distributions and shows how these impact the estimates on outcomes. Looking at either end of our outcomes chronologically, we can see that the positive impact of rank on not repeating third grade, eighth-grade test scores, high school graduation, and log wages manifests regardless of exact specification. The coefficients tend to increase in magnitude as we allow for more flexible interactions with achievement. This is consistent with the imbalance in rank found in Table 2, where more disadvantaged students appeared to have higher rank conditional on achievement, and this imbalance decreased once we allowed for more flexible functional form. Notwithstanding these slight differences between specifications, the estimates are largely consistent.

Two outcomes are different when comparing our preferred specification to the other specifications: enrollment in any college and graduation with a bachelor's degree. In both cases, the coefficient on rank becomes negative if we do not allow the impact of achievement to vary by test score distribution. We investigate the sources of this bias in Appendix Figure 5. For the uninteracted specification where there is imbalance, estimates of the effect of rank are very similar when controlling for predetermined characteristics relative to not controlling for them. This suggests that imbalance in predetermined covariates is not generating this bias in the estimates, which highlights the importance of functional form, because the bias in these estimates likely arises due to misspecification of the $g_d()$ function, where there are different mappings of achievement to outcomes across schools. While the imbalance in student characteristics is not driving the bias we observe, imbalance may be correlated to school unobservables which impact the mapping into future outcomes. Therefore, while misspecification and balance are independent concerns, they may often be linked.

6.2. Miscellaneous robustness checks: school size, breaking ties, measurement error

One data limitation is that we do not observe which classroom students are taught in for third-grade students. Hence, our main results use fixed effects for school-subject-cohort

(SSC), which we have been referring to as a class. Therefore, as a robustness check, we estimate the effects separately by the number of students in the third-grade school-subject-cohort. The smallest of these groups is under 30 students in a cohort, in which case the SSC would represent a single class. Note that estimates from such subsamples make the school distribution-type indicators effectively distribution-type-by-school-size indicators.

Estimating the impact of rank by school size will provide information on two other potential concerns. First, given that we are using the sampling variation in the test score distribution, one may be concerned that the results are driven by schools with smaller cohorts as they would experience more sampling variation. Second, as we use percentile rank, a given percentile rank will represent different absolute rankings in schools of different sizes. Estimating separately by school size ensures that the percentile rank and absolute rank approximate each other within each subsample.

Figure 11 presents the point estimates for under 30, under 50, 50–75, 75–100, and 100+ students in an SSC. While the effects tend to be larger for the larger schools, the pattern of results remain regardless of school size. Moreover, the students in schools with fewer than 30 students per school-subject-cohort are likely to all be in the same class during third grade, and the results hold.

In our main specification, we handle ties in rank by assigning students the mean of the rank. We consider other methods for breaking ties, including assigning the lowest rank, randomly breaking ties, and a rank only among students who are "on time" in third grade. Our results are qualitatively similar regardless of our method of dealing with ties. The results tend to be slightly smaller when we break ties randomly, which we attribute to the introduction of noise into our measure of rank (see Appendix Table 1).

We use third-grade student achievement on externally graded state examinations as a measure of each student's human capital. The rank of each student within their class is derived from their test score and all the test scores of their class peers. This means that any individual non-systematic measurement error in student test scores has impacts on the two key explanatory variables, rank and achievement.[45] This is potentially problematic for our

[45] We address any systematic measurement error at the class level, due to class-level shocks, with the inclusion of class fixed effects. Examples of this type of shock are a class having a particularly good or

estimation strategy, as both are subject to the same measurement error but to different degrees; this could generate a spurious relationship between student rank and gains in test scores or conversely attenuate a genuine relationship, depending on the extent of the mismeasurement.

When measurement error is large both the test scores and the rank of the student could be impacted. One example is that if a student "got lucky" in an examination, we would record them as having a higher rank than they actually have. Subsequently, students with mistakenly high ranks would have inferior outcomes compared to students that were not mismeasured, which would downward-bias the rank parameter. In contrast, when measurement error in student test scores is low, rank may not be impacted at all. In this case measurement error only impacts our measure of human capital and not rank, meaning that it is possible for the rank measure to pick up some information about human capital that is lost in the test score measure. This could generate a spurious correlation between rank and future outcomes, as here achievement is just acting as a proxy for innate ability.

Note that in addition, any measurement error in peers' test scores may also generate additional measurement error in the rank determination. The combination of these factors makes this a nonstandard measurement error problem. Specifically, we have nonlinear and nonadditive measurement error, in rank which will be correlated to the measurement error in human capital, and it is unclear how this would impact the rank parameter in practice.[46]

To gauge the importance of this problem in our setting we perform a data-driven bounding exercise following the approach of Murphy and Weinhardt (2020).[47] We implement a set of Monte Carlo simulations which add random noise to our baseline measure, recompute

bad teacher, or a disturbance on examination day. This is critical as such shocks would be rank-preserving.

[46] We discuss the nature of measurement error in this context in the Appendix. Two solutions to nonclassical measurement error, repeated measurements and instrumental variables (Schennach, 2016), are not available to us. We are using the first standardized test scores of each student; all later test score measures are potential outcomes and so could not be considered a repeated measurement. Using the test score from the other subject as an instrument is also not valid, as it does not meet the exclusion restriction. As we show in Figure 5, the rank (and test scores) in other subjects have spillover effects across subjects.

[47] Murphy and Weinhardt (2020) address the same problem and show that additive random noise would nonlinearly downward-bias the effect of class rank depending on the extent of the measurement error.

the student ranks based on this (more) noisy measure, and estimate our main results. We do this for three types of measurement with three levels of measurement error, performing 100 simulations of each.

First, we add measurement error drawn from a normal distribution $n(0, \vartheta)$, where ϑ takes the values of 10%, 20%, and 30% of the standard deviation in the third-grade achievement. Appendix Figure 6 shows the resulting estimates for our main outcomes. All the point estimates using data with additional measurement error are smaller in magnitude than our main estimates, and the estimated effect sizes continue to decrease as the amount of measurement error increases. The intuition for such a downward bias is that the "lucky student" will have higher baseline achievement and rank than they should have had; we therefore expect to observe these students (who are artificially highly ranked) having lower growth in test scores than expected.

Second, we add measurement error drawn from a uniform distribution $u[-\vartheta, \vartheta]$, where ϑ takes the values of 10%, 20%, and 30% of the standard deviation in the third-grade achievement. Appendix Figure 7 presents the simulation results for our main estimates with this additional uniform noise. These estimates are all smaller than our main effects, but increasing amounts of uniform noise in student test scores has no significant impact on the rank estimates, with the exception of eighth-grade test scores in the top half of the rank distribution.

Third, we add heterogenous measurement error, which increases the further we go from the state median achievement. This reflects the possibility that examinations are less accurate at the extremes and are targeted to measuring differences in ability within the middle part of the distribution. The additional measurement error was drawn from a normal distribution transformed by twice the student's absolute distance from the median plus 0.5, $(2|\alpha|) + 0.5$, namely $n(0, \vartheta) \times (2|\alpha| + 0.5)$. This is to ensure the mean of the scaling factor is 1, making it comparable to the normally distributed measurement error (in mean and variance). As before, ϑ takes the values of 10%, 20%, and 30% of the standard deviation in the third-grade achievement. Despite being scaled by the same amount on average, the impact of this type of noise on our estimates is considerably smaller than additional normally distributed measurement error (see Appendix Figure 8).

In summary, measurement error in test scores, which are then used to determine rank, generates nonlinear and nonadditive measurement error in rank, which will be correlated to the measurement error in human capital. A priori it is unclear how this would impact the rank parameters. We show that in practice increasing measurement error in student test scores attenuates the rank effect.

6.3. Heterogeneity

In this section we explore the heterogeneity of rank effects on the four main outcomes: (1) eighth-grade test scores; (2) graduating high school; (3) enrolling in college; and (4) log average earnings ages 23–27. We explore this for three predefined variables; race (white/non-white), gender (male/female), and Free and Reduced-Price Lunch (FRPL) eligibility in third grade (eligible/non-eligible). We present the estimates for each of these categories in Figures 12, 13, and 14, respectively.

First, is the impact of third-grade rank different for white and non-white students? Figure 12 shows that the effects of rank on eighth-grade test scores are similar by race. In contrast, the effects of rank on high school graduation are more pronounced for non-white students, for both low and high ranks. There is a more distinct difference when considering the impact on college attendance by race. Having a low rank in elementary school has a considerably larger negative effect on college attendance for non-white students compared to white students. The lowest-ranked non-white students are 7.5 percentage points less likely to attend college than the median-ranked student; in contrast, the lowest-ranked white students are only 1 percentage point less likely to attend. Once above the median rank, white and non-white students react in in a similar manner to their rank. We find that the effects of rank for low and high ranks are larger for non-white students, similar in pattern to high school graduation.

In contrast to the large differences by race, there is no evidence for heterogeneity with respect to gender for test scores, high school graduation, and college attendance. (Figure 13). This finding is different than Murphy and Weinhardt (2020), who show that boys' secondary school test scores have a more positive reaction to high elementary rank and a less negative reaction to having a low rank. However, we find that the case of males experiencing a low rank does have a less negative impact on earnings compared to females.

Finally, the heterogeneity of estimates with regards to FRPL students mirror those of white/non-white students, with the more disadvantaged group being more strongly affected by rank than their counterparts (Figure 14). This is true for test scores, graduation, college enrollment, and earnings. This finding is similar to Murphy and Weinhardt (2020) who find that in England, Free School Lunch students gain more from being highly ranked. One explanation is that these sets of students have low academic confidence or a different information set about achievement, and weight their school experience more heavily than non-disadvantaged students. This has important consequences for optimal classroom composition, which we discuss below.

7. Implications for School Choice

We estimate the effects of rank, net of SSC fixed effects. Traditional peer effects suggest that better peers should help performance. However, we show that having *more* better peers also has a negative effect by lowering rank. In this section we quantify the effects of rank as compared to the benefits of having better peers and school environments, and how these effects relate to other findings in the literature.

Consider the following thought experiment. A parent may move their child to a "better" school. This would come with a decrease in their child's rank and a likely increase in the quality of their child's peers. What would be the net gain in test scores from such a move?

To operationalize this, we categorize elementary schools into "good", "bad", and "average" according to two different measures of quality, mean achievement and value added. First, mean third-grade achievement is simply third-grade achievement estimated on indicators for a student's third-grade school (Columns 1–4, Table 3). Each school fixed effect represents the average third-grade achievement at that school. This is clearly not the causal effect of attending the school, although it may reflect what parents consider when choosing an elementary school, as such effects are relatively easy to observe. For our second quality measure, we calculate elementary schools' third- to eighth-grade unconditional value added by controlling for students' third-grade achievement (Columns 5–8, Table 3), which yields the average growth in test scores for students who attended a particular school. Each school fixed effect in this specification will capture many factors including student's own human capital, peer effects, resource differences, and parental investments. The standard deviation of these school value-added measures is 0.055. So, a student in a one standard deviation–better school has on average 5.5 percentile higher growth in test scores. Note, this is for benchmarking purposes only, which is reflected in our use of unconditional school value added in order to keep the analysis and interpretation as simple as possible (using estimates of school value added while conditioning on school demographic proportions provides very similar results).

To gauge the benefit of attending an elementary school with better attainment, we also record the mean value added of "good", "bad", and "average" schools in terms of attainment. It appears that schools with higher mean achievement also have higher third-to-eighth-grade value added, although these gains are only half the size compared to the effects if parents were selecting schools on the basis of value added: 0.026 versus 0.055 (final row of Table 3).

To ascertain the net benefits of attending these schools net of rank effects, we need to consider how a student's rank would change. The values in the columns of Table 3 are the mean rank of students in each state achievement ventile in each school type. For example, consider the median student at the tenth achievement ventile. If they attended an average elementary school in terms of third-grade achievement, their expected class rank would be 0.479, whereas they would have an expected rank of 0.613 (0.346) if they had attended a "bad" ("good") school. We can see that there is a clear trade-off in terms of rank and the quality of school when measured in absolute achievement. As may be expected, this rank–quality tradeoff is higher when elementary school quality is measured in mean achievement compared to value added. When parents move their child from a "bad" value-added school to a "good" one, the loss in rank is only −0.155. If they instead use mean achievement to decide, the change in rank is larger at −0.267 (Table 3, row 10).

We can see that students from higher up in the state achievement distribution also have higher ranks in their classes. There is not much difference in terms of expected rank at an average school, independent of whether school quality is measured in terms of value added or absolute attainment throughout the achievement distribution. However, this thought experiment clearly shows that the decrease in rank from moving from a "bad" to "good" school is always smaller when considering schools in terms of their value added. Moreover, the loss of rank from attending a "better" school is largest for the students near the middle of the distribution.

How would these changes in ranks and school environments impact a student's overall attainment? For this, we require one last piece of information, the linear relationship between rank and eighth-grade test score, which is 0.09. We estimate this by modifying Equation 5 to have a linear effect of rank. We can now calculate the impact on test scores.

Let us consider the case of parents of a median student (tenth ventile) considering "good" or "bad" schools in terms of mean third-grade achievement. Sending the child to the "better" school would lead to an increase in eighth-grade test scores by 6.1 percentiles (using the associated value-added scores from the bottom of Table 3, $0.026 - (-0.035)$). However, sending the child to the "better" school would reduce their expected class rank by 0.267. This would tend to reduce the student's eighth-grade test score by 2.4 percentiles (0.09×0.267). Therefore, the rank effect has reduced the gains from attending a school whose performance is two standard deviations better by 39 percent, resulting in a net gain of 3.7 percentiles.

Alternatively, if parents were better informed and selected elementary schools on the basis of value added, then there would be a smaller trade-off in class rank (−0.155) and larger increases in future test scores (0.11). In this case, a median student attending a "good" school rather than a "bad" one would gain 11 percentile points in the eighth-grade test score distribution. In contrast, the student would lose 1.4 percentiles due to their lower rank ($-0.014 = (-0.15) \times 0.09$). Hence, if parents were to choose on the basis of value added, there would be a net gain of 9.6 percentiles.

In the case of parents choosing on the basis of value added, the effect of school quality is roughly eight times the size of the rank effect. Note, while this rank effect is relatively small, this school quality measure encapsulates *all* observable and unobservable factors that contribute to student value added.

Ideally, we could describe the contribution of rank relative to all other peer effects; however, estimating all other types of peer effects is implausible in our setting. Moreover, while choosing the best school for their child in value-added terms is what most parents try to achieve, value added is difficult to observe. Basing school choice on the basis of mean achievement, we see that the importance of class rank is substantial, reducing any perceived benefits by up to 39 percent. The main message for parents weighing the trade-off of rank is that choosing schools based on value added is the best strategy.

Finally, what does the presence of rank effects mean for optimal classroom composition, given the heterogeneity of the effects? We find that disadvantaged groups such as non-white students or students eligible for FRPL are more affected by rank. Therefore, unlike linear-in-means peer effects, where moving students between groups would have no net impact, rearranging students with rank in mind could improve overall outcomes. This would require variation in test score distributions across classes, such that some low ranked disadvantaged students have the same achievement as high ranked advantaged students in other classes. These two groups of students exchanging classes would generate a net gain in achievement as the disadvantaged students would gain more from the higher rank than the non-disadvantaged students would lose from having a lower rank. However, this sort of exercise warrants caution because the changes in classroom distribution may be out of sample for the estimates in this paper (Carrell et al., 2013).

This finding illustrates a trade-off for programs that move disadvantaged third-grade students into situations where they will be the lowest-ranked student. The extensive literature on selective schools and school integration has shown mixed results from students attending selective or predominantly non-minority schools. There are generally two quasi-experimental approaches used for the impact of getting into a "high-quality" school, lottery admissions and regression discontinuities around admission thresholds. With respect to our paper, the key difference between these approaches is that only in regression discontinuities does attending a "higher quality" school necessarily come at the at the cost of a lower rank. Being admitted to a high value-added school via lottery would not necessarily cause a decrease in rank, but going to a selective school by just making the admission threshold would do so. In light of this, we now discuss our findings with relation to the existing literature on school quality.

The use of regression discontinuities in estimating the impact of attending an academically selective school ensures that students are similar across the boundary. The preponderance of evidence is that there are no benefits in terms of academic achievement, despite large improvements in peer quality (Angrist and Lang, 2004; Clark, 2010; Kling et al., 2007; Dobbie and Fryer, 2014 , Abdulkadiroğlu et al., 2014)[48]. In crossing the threshold, at least two types of peer effects would be at play in addition to other school factors. Students have better peers on average, but will have a lower rank. One can think of this as attending a "better school" as defined by mean achievement in our policy example, but more extreme. In this case the marginal student would mechanically go from highest ranked in one school to the lowest ranked in another.

The effectiveness of schools has also been estimated through lottery admissions to oversubscribed charter schools and voucher programs. In contrast to studies that use academic cutoffs, lottery admissions to charter schools do not necessarily pose a rank–school quality trade-off; in fact, in some cases students may gain in both rank and quality. We can think of this as a more extreme version of attending a better school, as defined by value added.

Entry to an oversubscribed charter school often improves student performance (Angrist et al., 2010, Angrist et al., 2013, Abdulkadiroğlu et al., 2011). Critically, attendance at these "high-quality" schools may not come at the expense of rank if they are recruiting low-attaining students. This can be seen in the heterogeneity of charter school effects for low-achievement students, which are positive in lower-achieving districts (Angrist et al., 2013), and are decreasing in the achievement of peers (Abdulkadiroğlu et al., 2011, Cullen et al., 2006). In summary, our results on class rank can help reconcile the different measured effects of "high-quality" schools across studies. Essentially, rank may "undo" the benefits of attending a "good" school for students who will be the lowest ranked in their class.

8. Conclusions

We make two contributions in this paper. First, we discuss identification of the effects of class rank. In particular, we discuss active sorting, passive sorting, and misspecification of the achievement mapping as potential sources of bias in this literature. We provide guidance on how to address these issues and reinforce the utility of balance and specification checks. Notably, the issues of misspecification and measurement error also apply to the literature that

[48] There *are* improvements in non-academic outcomes or college attendance.

estimates effects of rank, as well as other types of peer effects, using experiments that randomly allocate students of different abilities into classrooms.

Second, we demonstrate that a student's rank among their peers at a young age has long-lasting impacts. This affects a student's performance in school, including tests, courses taken, and progress through toward graduation. Ultimately, it also affects the chance of student graduation from high school. Relative position affects the decision to enroll in post-secondary education. Strikingly, it affects a student's real earnings in their mid-twenties. We find that a student enrolling in a class where they are at the 75th percentile rather than the 25th in third grade increases their real wages between ages 23 and 27 by $1500 per annum, or approximately 7 percent.[49] For comparison, Carrell et al. (2018) look at the long-run impact of peers at the same ages, and find that being in a class of 25 with a student who was exposed to domestic violence reduces an individual's earnings by 3 percent.

Our findings add to a growing list of papers demonstrating that conditions for young children have long-lasting consequences. In contrast to other papers that focus on policy differences faced by students, we document the effect of an unavoidable phenomenon in groups: relative rank. Some of the effect of rank may come via teachers and administrator interactions with students. We document that students are more likely to be retained in third grade which is a decision made not by the student but by teachers, administrators, and families.

Moreover, we find that disadvantaged groups gain more from being highly ranked and lose more from being lowly ranked among their peers. Therefore, unlike linear-in-means peer effects, where moving students between groups would have no net impact, grouping students with rank and these heterogeneous effects in mind could improve overall outcomes. Schools may also influence student performance by manipulating the salience of rank.

Finally, we examine if and to what extent parents should consider rank effects when choosing the best school for their children. Critically, we document a trade-off from attending

[49] Using the present discounted value of earnings of $522,000 as in Chetty et al. (2014), who follow Krueger (1999) in discounting earnings gains at a 3 percent real annual rate, we calculate that these rank differences would increase life-time earnings by $36,540 in net present value. This figure is based on the point estimates from Chetty et al. (2014), Figure 10 Panel 5. The 5th ventile has a coefficient of −0.032 while the 15th ventile has an estimate of 0.035.

a school with high-achieving peers: this mechanically lowers the class rank of your own child. We examine this trade-off in detail based on the observed student and school allocations in Texas. We find that rank offsets about 40 percent of the benefits of school value added for the median-performing student, if parents choose schools based on mean peer achievement. If parents instead choose schools based on value added, the offsetting effects of rank from attending a better school are much smaller.

Future research on rank should focus on the interaction between rank and policies that exaggerate or mediate the effects of rank. Future research should also consider the effect of rank in groups outside of school settings.

References

Abdulkadiroğlu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J., & Pathak, P. A. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. The Quarterly Journal of Economics, 126(2), 699-748.

Abdulkadiroğlu, A., Angrist, J., & Pathak, P. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. Econometrica, 82(1), 137–196.

Angrist, J. D. & Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston's Metco program. American Economic Review, 94(5), 1613–1634.

Angrist, J. D., Dynarski, S. M., Kane, T. J., Pathak, P. A., & Walters, C. R. (2010). Inputs and impacts in charter schools: KIPP Lynn. American Economic Review, 100(2), 239-43.

Angrist, J. D., Pathak, P. A., & Walters, C. R. (2013). Explaining charter school effectiveness. American Economic Journal: Applied Economics, 5(4), 1-27.

Azmat, G., Bagues, M., Cabrales, A., & Iriberri, N. (2019). What you know can't hurt you, for long: A field experiment on relative performance feedback. Management Science, Vol. 65, No. 8, 3714–3736.

Azmat, G. & Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. Journal of Public Economics, 94(7–8), 435–452.

Bergman P. (2018). "The Risks and Benefits of School Integration for Participating Students: Evidence from a Randomized Desegregation Program" IZA DP No. 11602.

Bertoni, M., & Nisticò, R. (2019). Ordinal Rank and Peer Composition: Two Sides of the Same Coin?.

Bietenbeck, J. (2020). "The Long-Term Impacts of Low-Achieving Childhood Peers: Evidence from Project STAR" Journal of the European Economic Association, Vol. 18, No. 1 (February 2020), 392-426.

Black, S., Devereux, P., & Salvanes, K. (2011). "Too young to leave the nest? The effects of school starting age" Review of Economics and Statistics, 92(2), 455–467.

Black, S., Devereux, P. & Salvanes, K. (2013). "Under Pressure? The Effect of Peers on Outcomes of Young Adults" Journal of Labor Economics, Vol. 31, No. 1 (January 2013), 119-153.

Booij, A., Leuven, E., & Oosterbeek, H. (2017). "Ability peer effects in university: Evidence from a randomized experiment" Review of Economic Studies, 84(2), 547-587.

Brown, G., Gardiner, J., Oswald, A. & Qian, J. (2008). “Does wage rank affect employees’ well-being?” Industrial Relations: A Journal of Economy and Society, 47(3), 355–389.

Card, D., Mas, A., Moretti, E., & Saez, E. (2012). “Inequality at work: The effect of peer salaries on job satisfaction” American Economic Review, 102(6), 2981-3003.

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

Carrell, S. E. & Hoekstra, M. L. (2010). “Externalities in the classroom: How children exposed to domestic violence affect everyone's kids.” American Economic Journal: Applied Economics, 2(1), 211-28.

Carrell, S. E., Hoekstra, M., & Kuka, E. (2018). “The long-run effects of disruptive peers” American Economic Review, 108(11), p.3377-3415

Carrell, S. E., Sacerdote, B. I., & West, J. E. (2013). “From natural variation to optimal policy? The importance of endogenous peer group formation” Econometrica, 81(3), 855-882.

Chetty, R., Friedman, J., Hilger, N., Saez, E., Schanzenbach, D.W., & Yagan, D. (2011). “How does your kindergarten classroom affect your earnings? Evidence from Project STAR” The Quarterly Journal of Economics, 126(4), 1593-1660.

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

Chetty, R., Hendren, N., & Katz, L. (2016). “The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment” American Economic Review, 106(4), 855-902.

Clark, D. (2010). “Selective schools and academic achievement” The B. E. Journal of Economic Analysis & Policy, 10(1), 1–40.

Cullen, J. B., Jacob, B. A., & Levitt, S. (2006). “The Effect of School Choice on Participants: Evidence from Randomized Lotteries” Econometrica, 74(5), 1191–1230.

Delaney, J. & Devereux, P. J. (2019). The Effect of High School Rank in English and Math on College Major Choice.

Dobbie, W. & Fryer, R. (2014). “The impact of attending a school with high-achieving peers: Evidence from the New York City exam schools” American Economic Review, 6(3), 58–75.

Elsner, B. & Isphording, I. (2017a). “A Big Fish in a Small Pond: Ability Rank and Human Capital Investment” Journal of Labor Economics, 35(3).

Elsner, B. & Isphording, I. E. (2017b). "Rank, Sex, Drugs and Crime" Journal of Human Resources, 0716-8080R.

Elsner, B., Isphording, I.E., & Zölitz, U. (2018). "Achievement Rank Affects Performance and Major Choices in College" SSRN working paper.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation (No. w7867). National Bureau of Economic Research.

Hoxby, C. & Weingarth, G. (2006). "Taking race out of the equation: School reassignment and the structure of peer effects" Working paper, Harvard University.

Kinsler, J. & Pavan, R. (2021). Local Distortions in Parental Beliefs over Child Skill. Journal of Political Economy, 129(1), 000-000.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). "Experimental Analysis of Neighborhood Effects" Econometrica, 75(1), 83–119.

Kremer, M., & Levy, D. (2008). Peer effects and alcohol use among college students. Journal of Economic perspectives, *22*(3), 189-206.

Krueger, A. B. (1999). Experimental estimates of education production functions. The Quarterly Journal of Economics, 114(2), 497-532.

Lavy, V., Silva, O., & Weinhardt, F. (2012). "The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools" Journal of Labor Economics, 30(2), 367–414.

Luttmer, E. F. P. (2005). "Neighbors as negatives: Relative earnings and well-being" The Quarterly Journal of Economics, 120(3), 963–1002.

Marsh, H., Seaton, M., Trautwein, U., Ludtke, O., Hau, O'Mara, A., & Craven, R. (2008). "The Big Fish Little Pond Effect Stands Up to Critical Scrutiny: Implications for Theory, Methodology, and Future Research" Educational Psychology Review, 20(3), 319–350.

Murphy, R. & Weinhardt, F. (2014). "Top of the Class: The Importance of Ordinal Rank" CESIFO Working Paper NO. 4815.

Murphy, R. & Weinhardt, F. (2020). "Top of the Class: The Importance of Ordinal Rank" The Review of Economic Studies, vol. 87(6), 2777–2826. https://doi.org/10.1093/restud/rdaa020.

Oreopoulos, P., Stabile, M., Walld, R., & Roos, L. L. (2008). "Short-, medium-, and long-term consequences of poor infant health an analysis using siblings and twins" Journal of Human Resources, 43(1), 88-138.

Payne, A. & Smith, J. (2018). "Big Fish, Small Pond: The Effect of Rank at Entry on Post-Secondary Outcomes" Mimeo.

Ribas, R .P., Breno, S., & Giuseppe, T. (2018). “Can Better Peers Signal Less Success? The Disruptive Effect of Perceived Rank on Career Investment” SSRN working paper, http://dx.doi.org/10.2139/ssrn.3135824.

Ribas, R. P., Sampaio, B., & Trevisan, G. (2020). “Short-and Long-term Effects of Class Assignment: Evidence from a Flagship University in Brazil” *Labour Economics*, 101835.

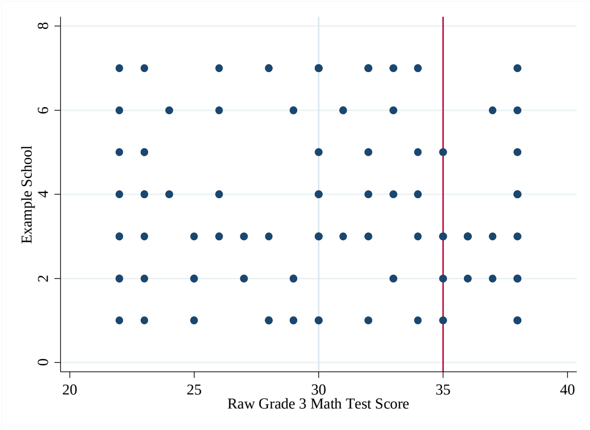
Sacerdote, B. (2001). “Peer Effects with Random Assignment: Results for Dartmouth Roommate” The Quarterly Journal of Economics 116 (2), 681-704.

Tversky, A. & Kahneman, D. (1974). “Judgment under uncertainty: Heuristics and biases” Science, 185(4157), 1124–1131.

Whitmore, D. (2005). "Resource and peer impacts on girls’ academic achievement: Evidence from a randomized experiment" American Economic Review P & P, 199-203.

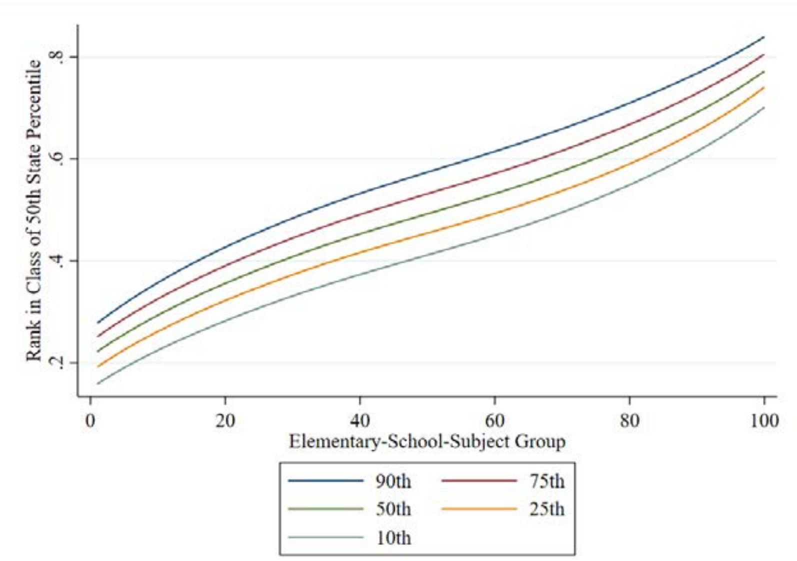
Zax, J. S. & Rees, D. I. (2002). IQ, academic performance, environment, and earnings. *Review of Economics and Statistics*, *84*(4), 600-616.

Figure 1: Class Test Score Distributions



Note: These figures are based on hypothetical data, based on schools in Texas. This example shows seven classrooms with the same minimum, maximum, and mean scores, with students who achieve the same score having different ranks.

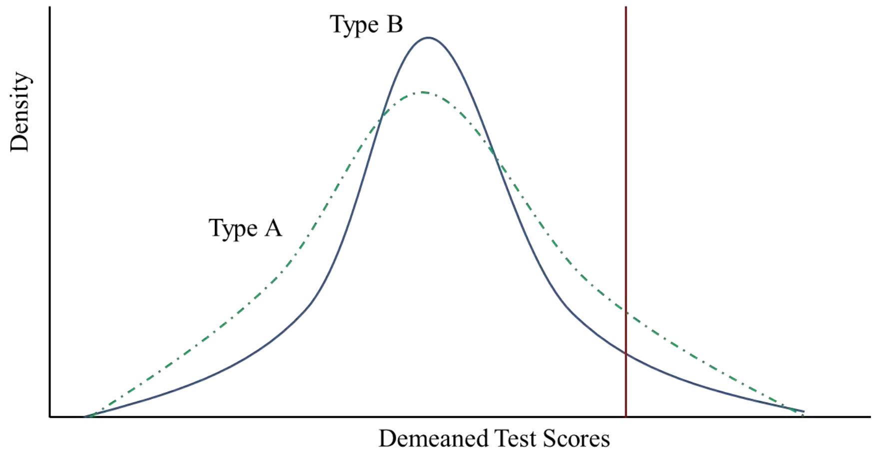
Figure 2: Variation of Local Rank of Median Student Within School-Subject Groups



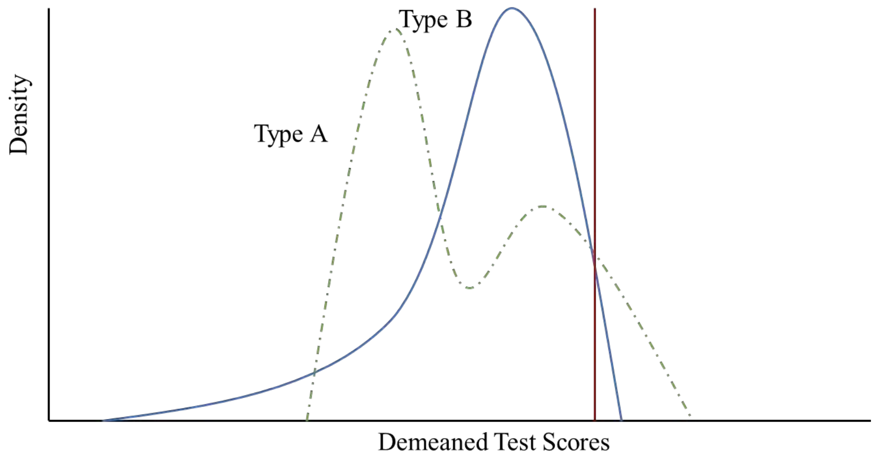
Note: This figure plots the distribution of rank for students at the 50*th* percentile of the statewide achievement distribution within different school-subject groups. The horizontal axis plots elementary school-subjects as percentiles so that the school with the lowest (highest) rank for the statewide median student has a value of 1 (100). The variation in class rank for the state median student comes from the 13 cohorts from 1995 to 2007. For each school-subject percentile, we first plot the class rank of the 10th percentile of students with state median achievement. The equivalent curves are also drawn for the 25th, 50th, 75th, and 90th percentiles.

Figure 3: Passive Sorting and Rank

A. High/Low Variance



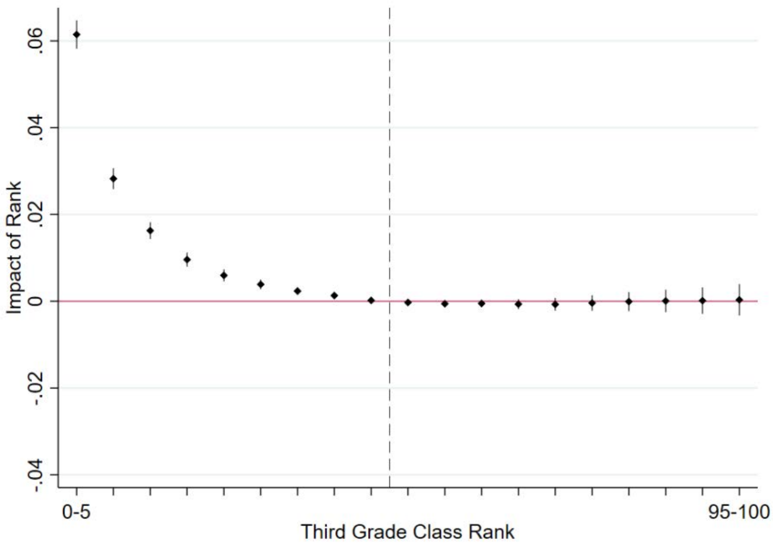
B. Nonstandard Distribution



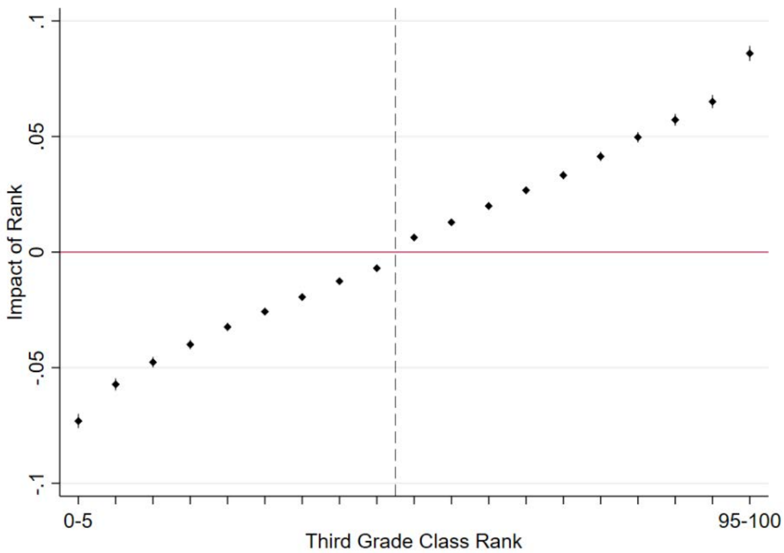
Note: These figures display several hypothetical test score distributions to illustrate the issue of passive sorting. Each panel shows the test score distribution of two types of schools, Type A and Type B, that have been transformed such that they have the same mean. The red line indicates an arbitrary test score of students a set distance from the type average. In both panels, due to the shape of the test score distributions, a student with the same test score relative to the mean (red line) would have a higher rank in in Type B compared to Type A. Passive sorting could occur if students with different characteristics systematically attended schools with different distribution types. Passive sorting would then cause a bias if the characteristics directly impact future outcomes.

Figure 4 – Effect of Third-Grade Rank on K–12 Outcomes, 1

A. Repeat Third Grade



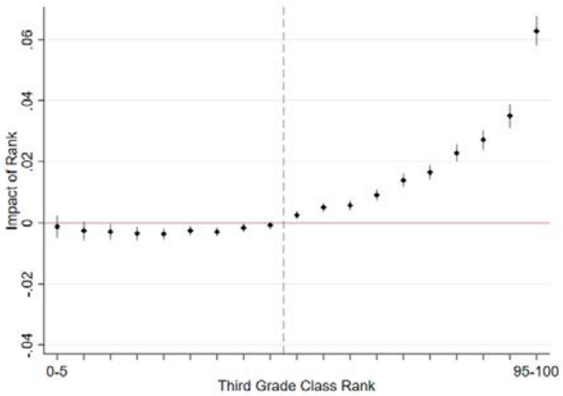
B. Eighth-Grade Test Scores



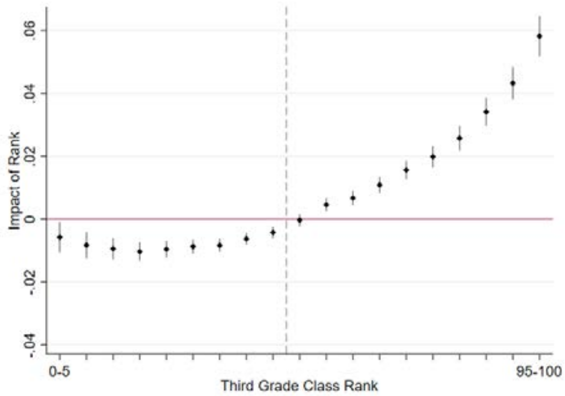
Note: These figures plot the coefficient for ventiles of class rank with 95% confidence intervals calculated using standard errors clustered at the school level. The 45th to 50th percentile interval is the omitted category. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement interacted with type of elementary school test-score distribution. The mean retention rate is 1.6 percent.

Figure 5 – Effect of Third-Grade Rank on K–12 Outcomes, 2

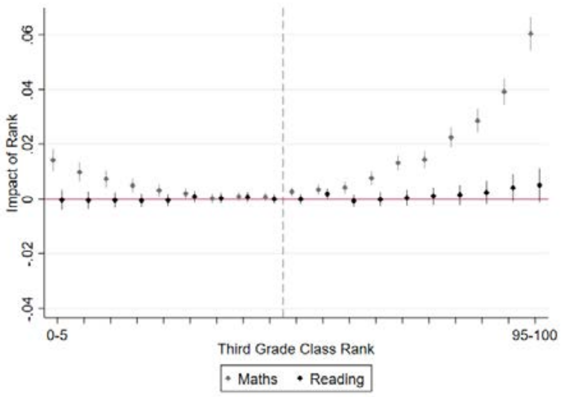
A. AP Calculus



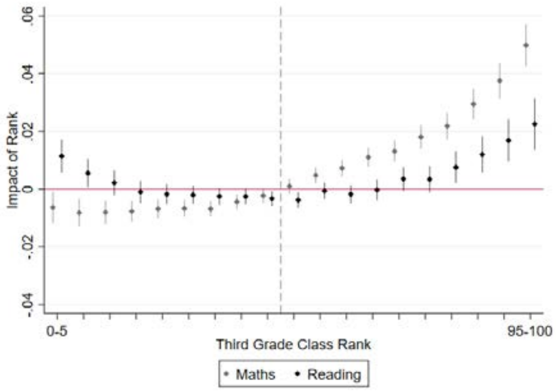
B. AP English



C. AP Calculus, Subject-Specific Rank



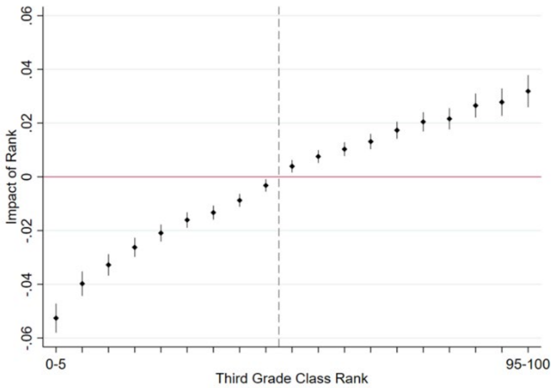
D. AP English, Subject-Specific Rank



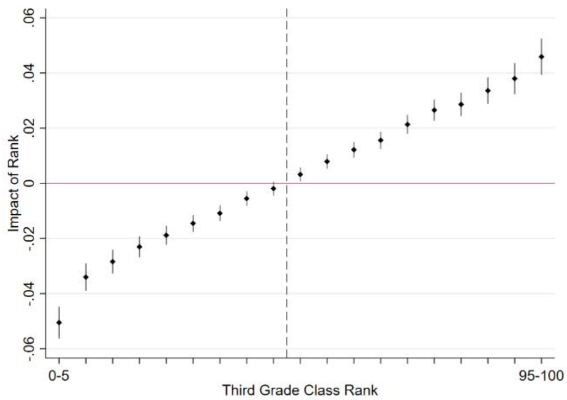
Note: These figures plot the coefficient for ventiles of class rank with 95% confidence intervals calculated using standard errors clustered at the school level. The 45th to 50th percentile interval is the omitted category. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement interacted with type of elementary school test score distribution. Panels C and D come from a specification which controls for achievement and rank in the third grade in both subjects simultaneously.

Figure 6 – Effect of Third-Grade Rank on High School Graduation and College Enrollment

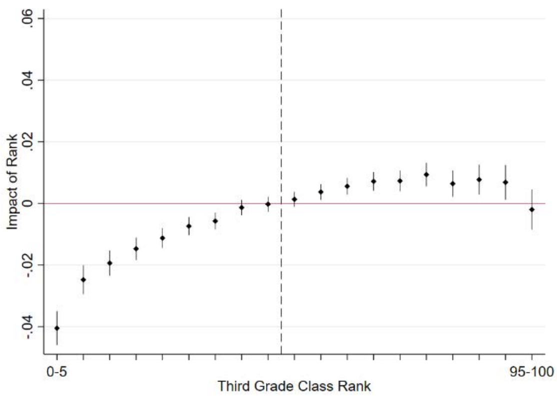
A. High School Graduation



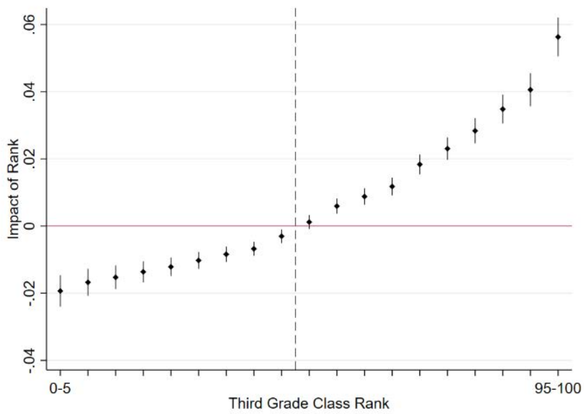
B. College Enrollment



C. Community College Enrollment



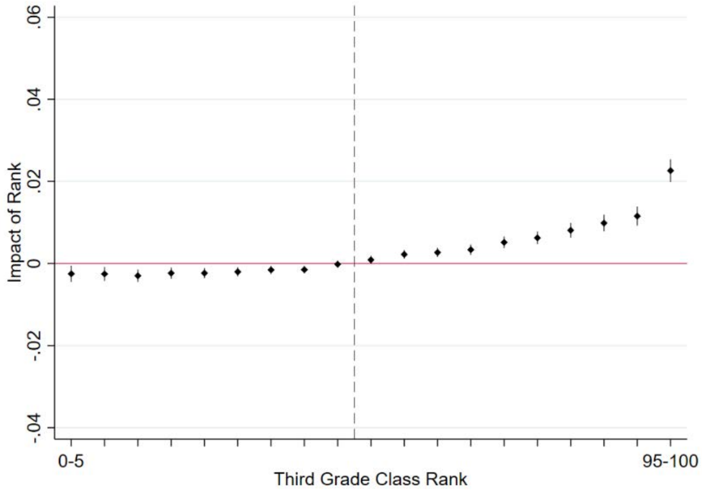
D. 4-Year College Enrollment



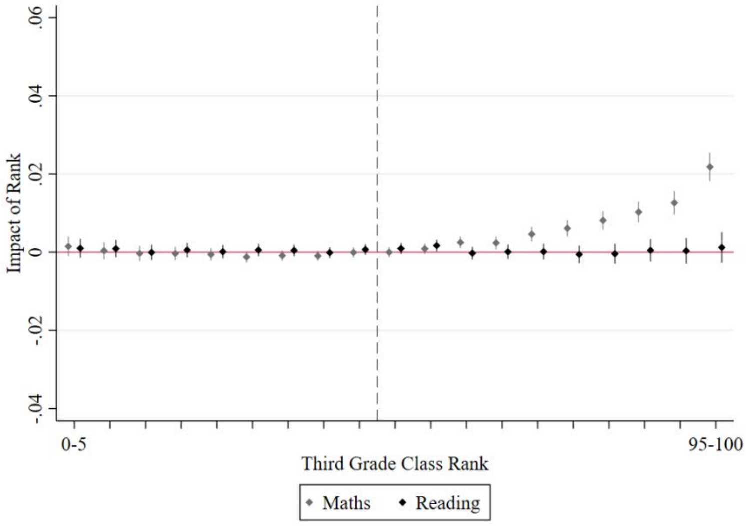
Note: These figures plot the coefficient for ventiles of class rank with 95% confidence intervals calculated using standard errors clustered at the school level. The 45th to 50th percentile interval is the omitted category. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement interacted with type of elementary school test score distribution. The mean high school graduation rate is 71 percent, with a college enrollment rate of 47 percent. The mean two-year college enrollment rate is 31 percent, and the mean four-year college enrollment rate is 23 percent, as some attend both two- and four-year colleges.

Figure 7 – Effect of Third-Grade Rank on Major Choice

A. Declaring a STEM Major



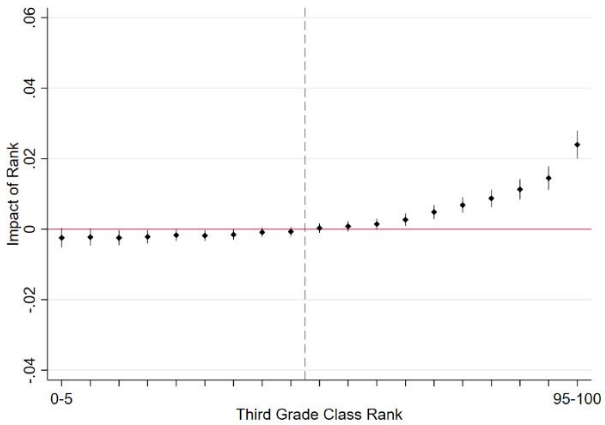
B. Declaring a STEM Major by Subject-Specific Rank



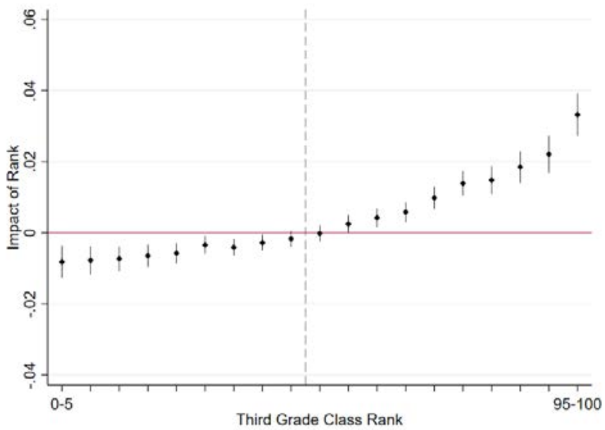
Note: These figures plot the coefficient for ventiles of class rank with 95% confidence intervals calculated using standard errors clustered at the school level. The 45th to 50th percentile interval is the omitted category. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement interacted with type of elementary school test score distribution. Panel B comes from a specification which controls for achievement and rank in third grade in both subjects simultaneously. Mean STEM enrolment is 0.04.

Figure 8 – Effect of Third-Grade Rank on Bachelor's Degree Receipt

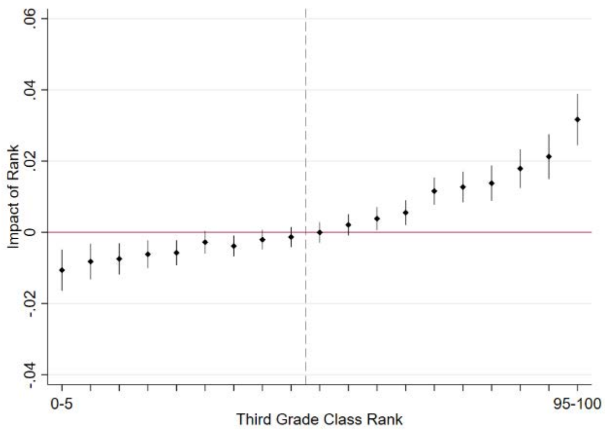
A. Graduate 4-Year College in 4 years



B. Grad 4-Year College in 6 years



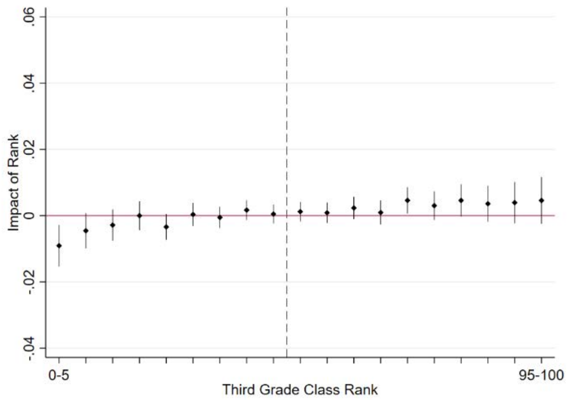
C. Grad 4-Year College in 8 years



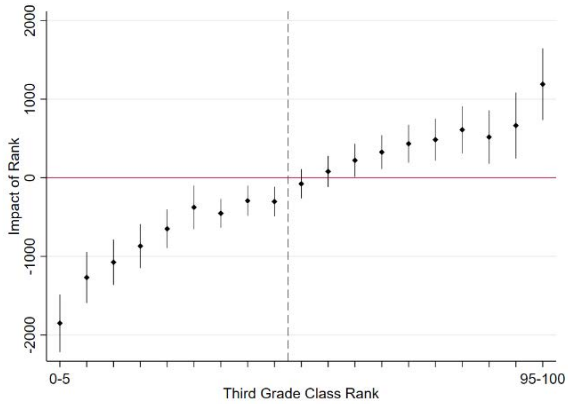
Note: These figures plot the coefficient for ventiles of class rank with 95% confidence intervals calculated using standard errors clustered at the school level. The 45th to 50th percentile interval is the omitted category. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement interacted with type of elementary school test score distribution. Mean graduation rate in 4/6/8 years is 4/14/16 percent.

Figure 9 – Effect of Third-Grade Rank on Labor Market Outcomes (Age 23–27)

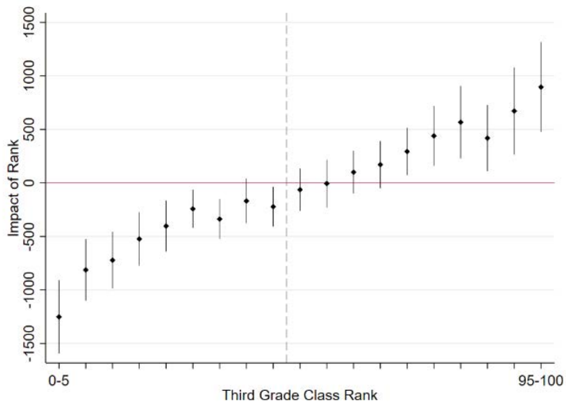
A. Positive Earnings



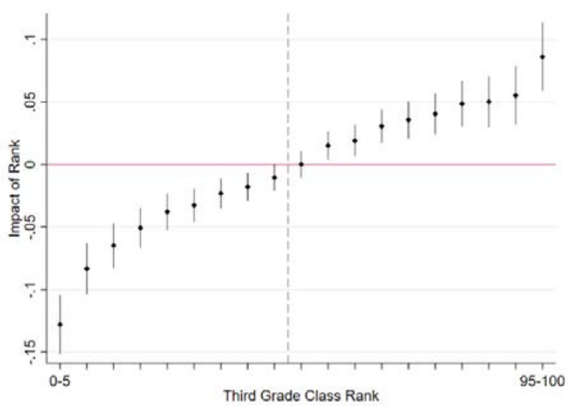
B. Average Earnings



C. Average Nonzero Earnings

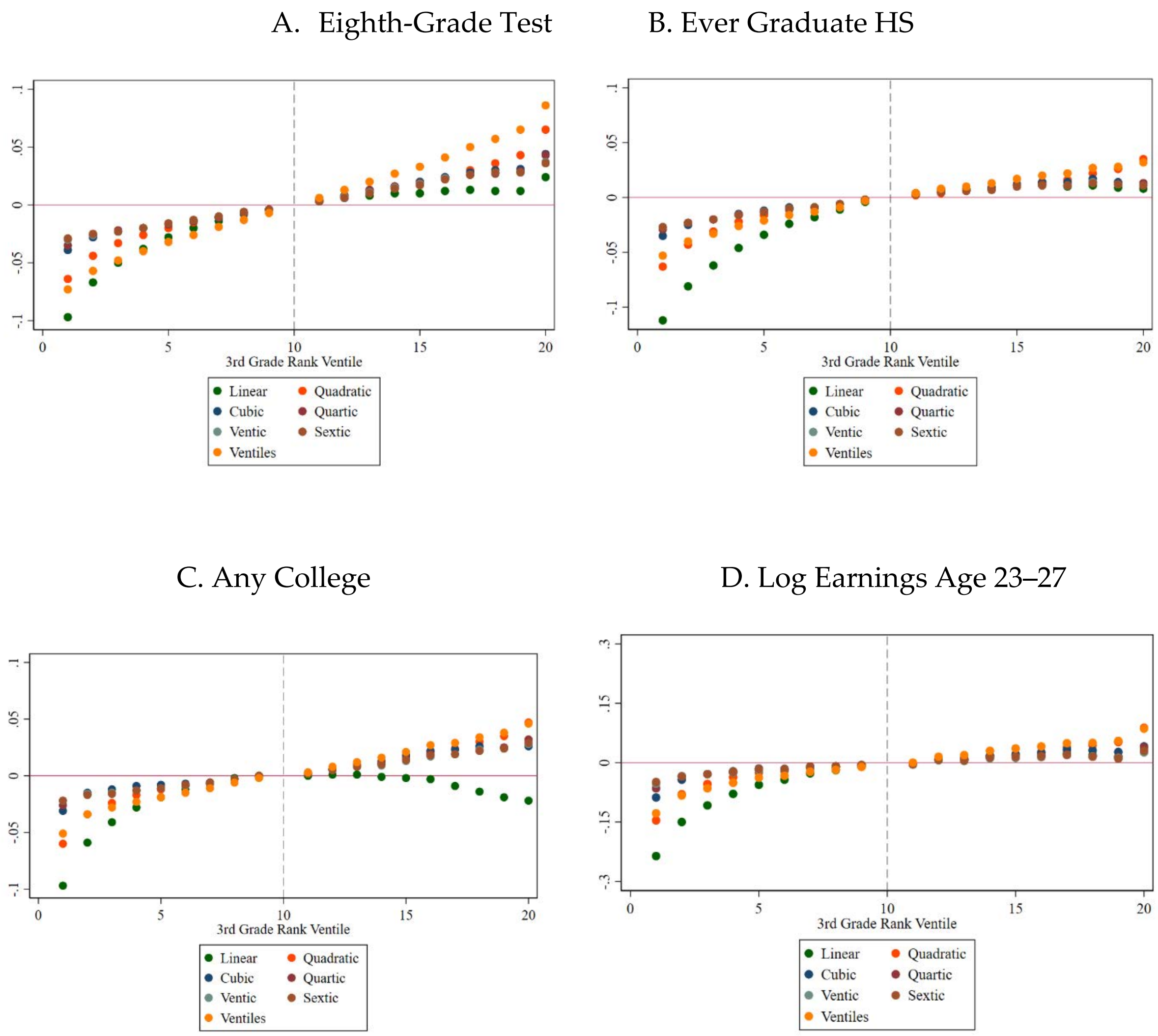


D. Log Average Earnings



Note: These figures plot the coefficient for ventiles of class rank with 95% confidence intervals calculated using standard errors clustered at the school level. The 45th to 50th percentile interval is the omitted category. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement interacted with type of elementary school test score distribution. The mean positive annual earnings between 23–27 are $24,912. Mean earnings including zeroes are $17,365.

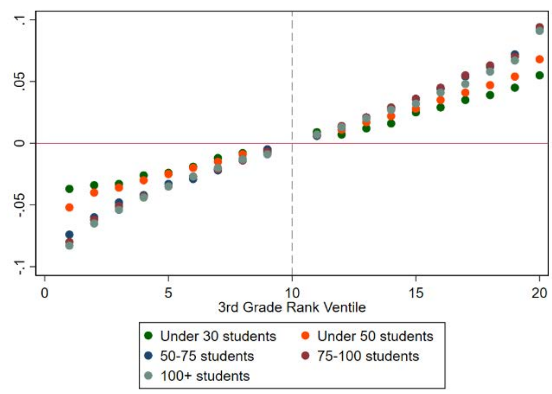
Figure 10 – Flexible Controls for Achievement



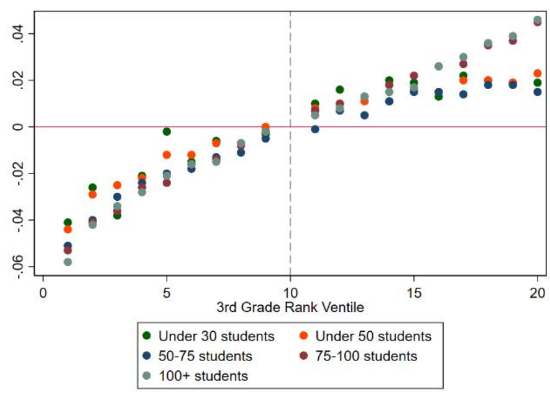
Note: These figures plot the coefficient for ventiles of class rank. The 45th to 50th percentile interval is the omitted category. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement.

Figure 11 – Results by Class Size

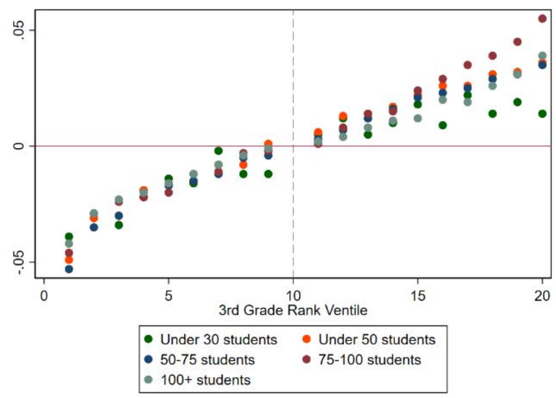
A. Eighth Grade Test



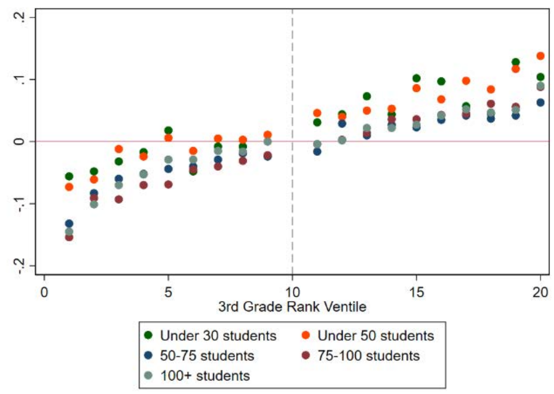
B. Ever Graduate HS



C. Any College



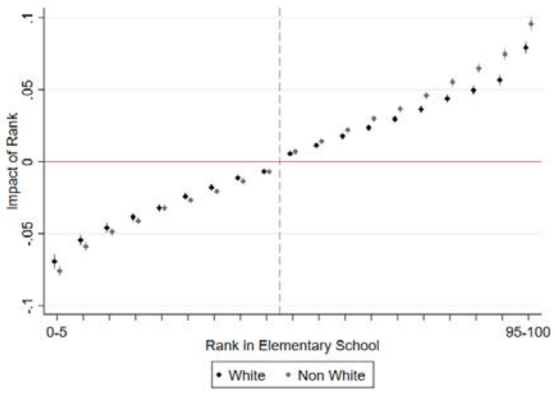
D. Log Earnings Age 23–27



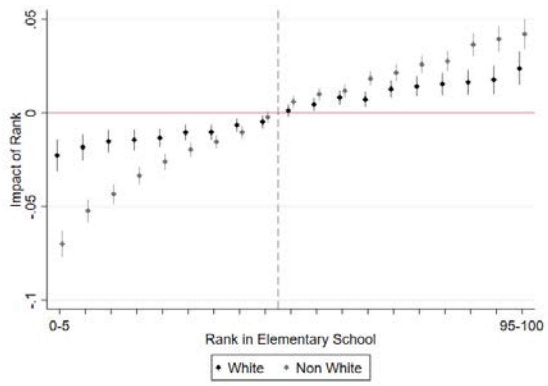
Note: These figures plot the coefficient for ventiles of class ran with 95% confidence intervals calculated using standard errors clustered at the school level. The 45th to 50th percentile interval is the omitted category. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement.

Figure 12 – Heterogeneity by Race

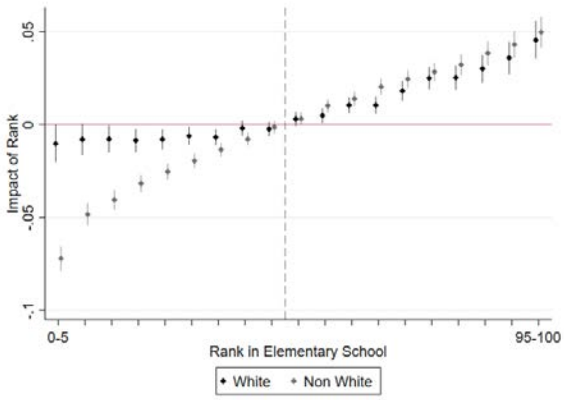
A. Eighth-Grade Test



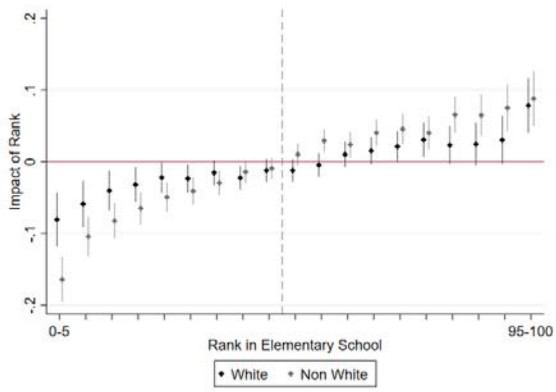
B. Ever Graduate HS



C. Any College



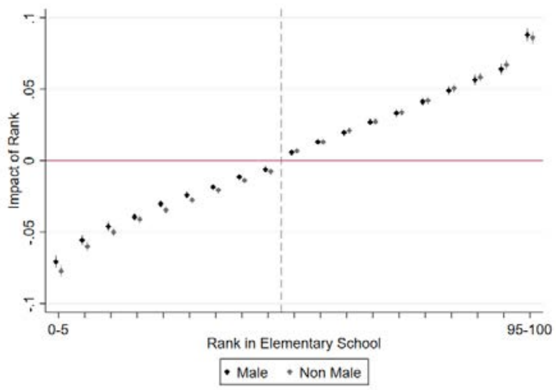
D. Log Earnings Age 23–27



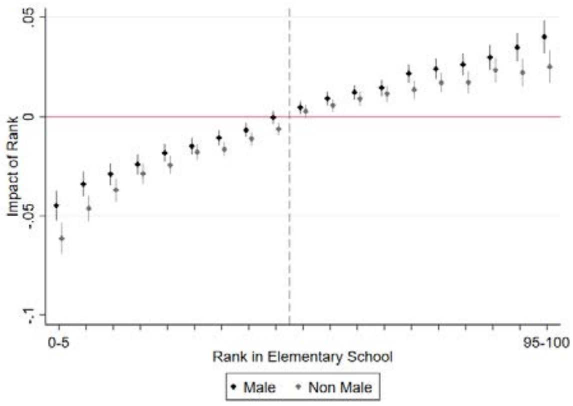
Note: These figures plot the coefficient for ventiles of class rank with 95% confidence intervals calculated using standard errors clustered at the school level. The 45th to 50th percentile interval is the omitted category. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement.

Figure 13 – Heterogeneity by Gender

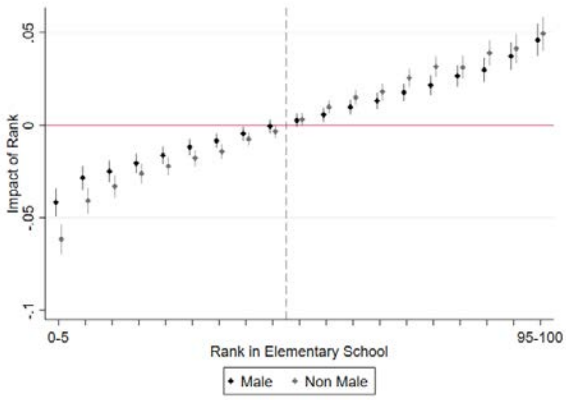
A. Eighth-Grade Test



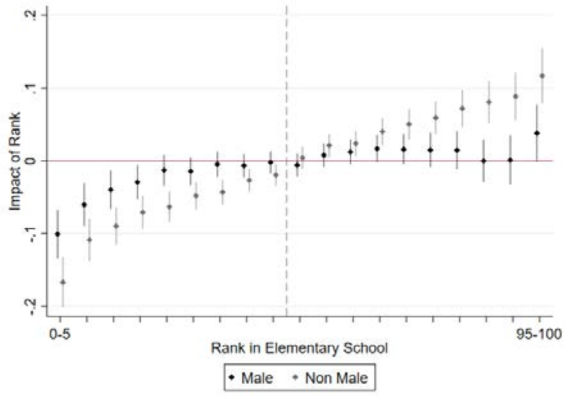
B. Ever Graduate HS



C. Any College



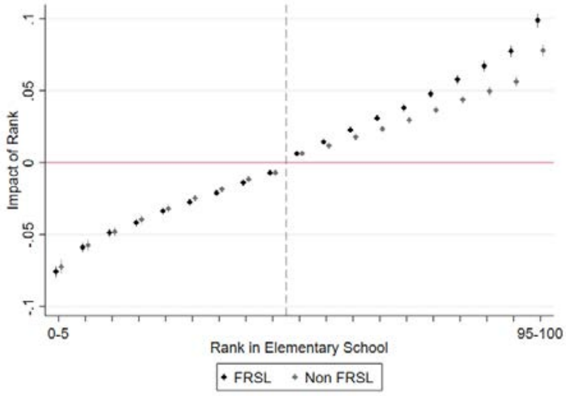
D. Log Earnings Age 23–27



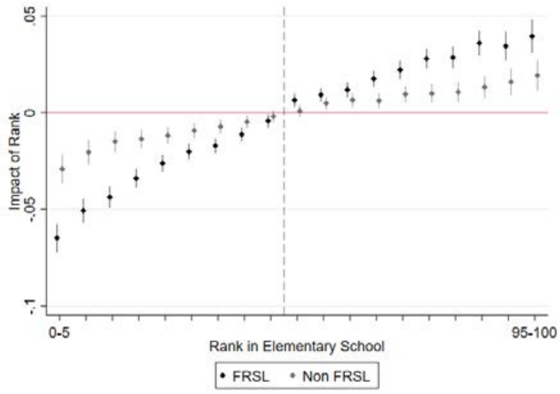
Note: These figures plot the coefficient for ventiles of class rank. The 45th to 50th percentile interval is the omitted category with 95% confidence intervals calculated using standard errors clustered at the school level. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement.

Figure 14 – Heterogeneity by Free and Reduced-Price Lunch

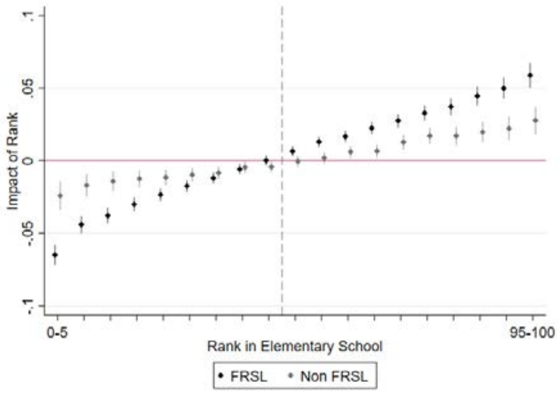
A. Eighth-Grade Test



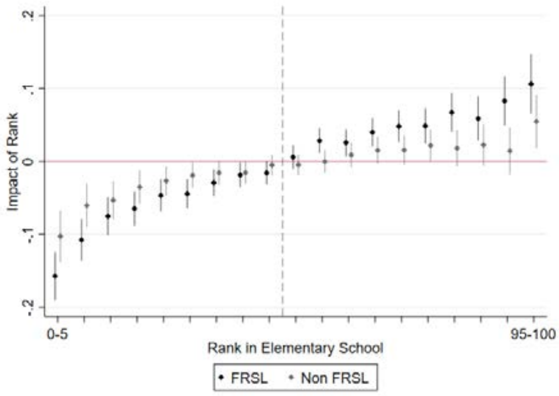
B. Ever Graduate HS



C. Any College



D. Log Earnings Age 23–27



Note: These figures plot the coefficient for ventiles of class rank. The 45th to 50th percentile interval is the omitted category with 95% confidence intervals calculated using standard errors clustered at the school level. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement.

Table 1: Summary Statistics

	Mean	Standard Deviation	Observations
Demographics – 13 Cohorts			
Male	0.50	0.50	6,117,690
Economic Disadvantage	0.49	0.50	6,117,690
English as a Second Language	0.12	0.32	6,117,690
White	0.47	0.50	6,117,690
Asian	0.03	0.17	6,117,690
Black	0.15	0.36	6,117,690
Hispanic	0.35	0.48	6,117,690
Size of Third Grade SSC	93.7	46.6	6,117,690
Repeat Third Grade	0.02	0.13	6,117,690
K–12 Outcomes – 13 Cohorts			
State Test Percentile, Eighth Grade	0.55	0.28	4,919,673
Ever Graduate High School	0.70	0.46	6,117,690
AP Calculus	0.08	0.28	6,117,690
AP English	0.18	0.39	6,117,690
Any College	0.46	0.50	6,117,690
Enroll, 4-Yr College	0.23	0.42	6,117,690
Enroll, 2-Yr College	0.31	0.46	6,117,690
College Outcomes – 10/8/6 Cohorts			
Declare STEM Major	0.041	0.198	6,117,690
BA in 4 years	0.06	0.24	4,573,672
BA in 6 years	0.14	0.34	3,597,340
BA in 8 years	0.16	0.37	2,647,240
Age 23–27 Labor Outcomes – 8 Cohorts			
Nonzero Wages	0.65	0.43	3,597,340
Real Wages	17,300	24,093	3,597,340
Real Nonzero Wages	24,818	25,372	2,652,284

Note: This table contains summary statistics for the main estimating sample of third graders from 1995 to 2007. Some outcomes are only available for early cohorts which generates the differences in sample size. *Enroll, 4-Yr College* means enrollment within three years of "on-time" high school graduation and is similarly defined for two-year colleges.

Table 2 – Balance Test

	Low Income	Male	ESL	White	Asian	Black	Hispanic
				A. Uninteracted			
Rank	0.097***	−0.011**	0.030***	−0.085***	−0.001*	0.047***	0.037***
	(0.005)	(0.004)	(0.003)	(0.005)	(0.000)	(0.004)	(0.004)
				B. School Mean Quartiles			
Rank	0.030***	−0.004	0.010*	−0.028***	−0.001	0.024***	0.006
	(0.006)	(0.005)	(0.004)	(0.006)	(0.002)	(0.005)	(0.005)
				C. School Mean Deciles			
Rank	0.025***	−0.001	0.006	−0.024***	−0.004	0.023***	0.006
	(0.006)	(0.005)	(0.004)	(0.006)	(0.002)	(0.005)	(0.005)
				D. School Variance Quartiles			
Rank	0.039***	−0.015***	0.020***	−0.039***	0.000	0.021***	0.018***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.001)	(0.004)	(0.004)
				E. School Variance Deciles			
Rank	0.030***	−0.014***	0.020***	−0.031***	−0.001	0.016***	0.015***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.001)	(0.004)	(0.004)
				F. School Quartiles of Mean and Quartiles of Variance			
Rank	−0.006	−0.006	0.002	0.008	0.003*	−0.003	−0.009*
	(0.005)	(0.005)	(0.004)	(0.005)	(0.002)	(0.004)	(0.004)
N	6,117,651	6,117,651	6,117,651	6,117,651	6,117,651	6,117,651	6,117,651

Notes: This table regresses predetermined characteristics on a linear effect of rank, SSC fixed effects, and achievement controls. Panel A does not interact ventiles of achievement with anything. Panel B interacts ventiles of achievement with indicators for quartiles of school mean achievement. Panel C interacts achievement with deciles of school mean. Panel D interacts achievement with quartiles of school variance. Panel E interacts achievement with deciles of school variance. Panel F is our preferred specification and interacts achievement with 16 indicators for school mean and variance (quartiles of mean × quartiles of variance). Standard errors are clustered at the school level.

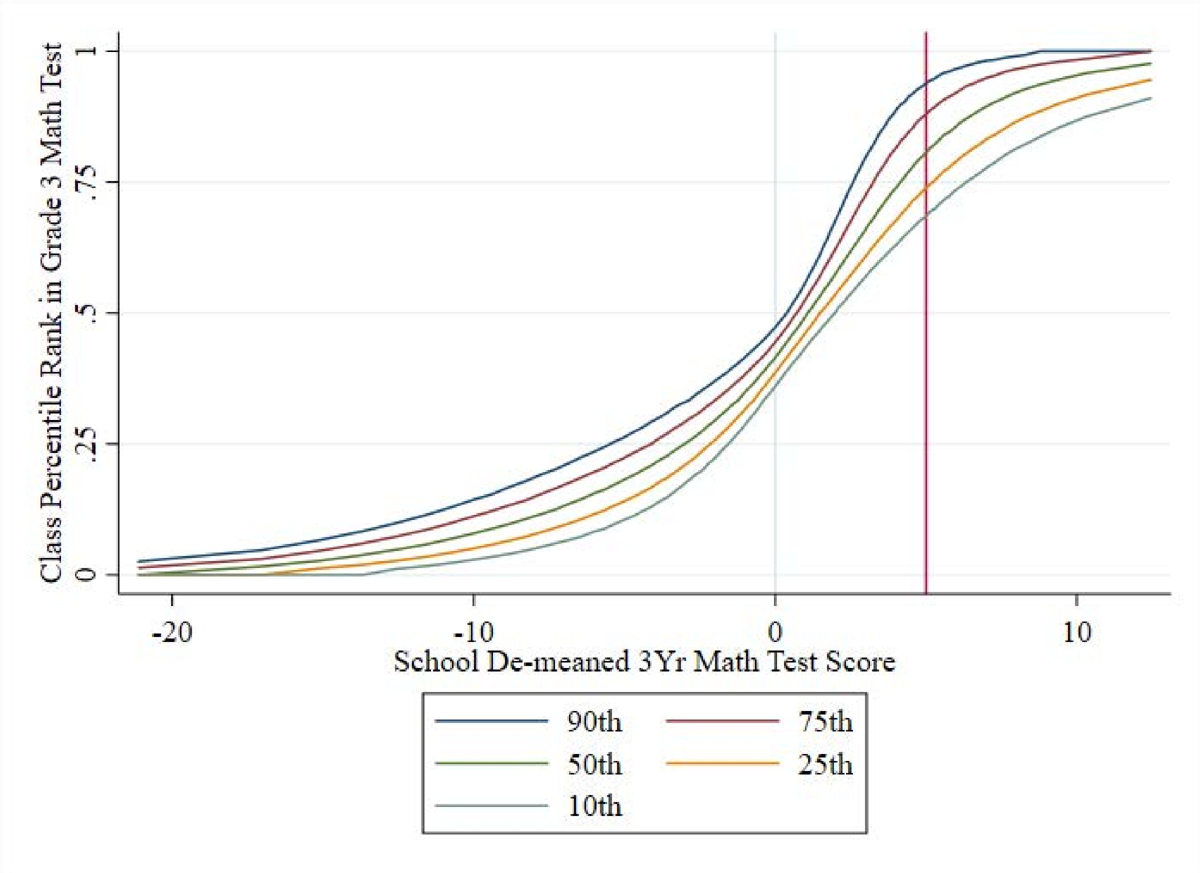
Table 3: The Distribution of Rank by Elementary School Effectiveness Measures

	Third-Grade Attainment				Third- to Eighth-Grade Value Added			
State Ventile	"Average" School Rank	"Bad" School Rank	"Good" School Rank	(3)-(2) Change	"Average" School Rank	"Bad" School Rank	"Good" School Rank	(3)-(2) Change
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	0.028	0.054	0.014	*−0.040*	0.049	0.075	0.035	*−0.040*
2	0.074	0.137	0.039	*−0.098*	0.106	0.145	0.075	*−0.070*
3	0.123	0.211	0.068	*−0.143*	0.160	0.203	0.113	*−0.090*
4	0.172	0.279	0.100	*−0.179*	0.213	0.257	0.153	*−0.104*
5	0.222	0.341	0.133	*−0.208*	0.261	0.309	0.188	*−0.121*
6	0.272	0.401	0.170	*−0.231*	0.310	0.354	0.227	*−0.127*
7	0.323	0.458	0.210	*−0.248*	0.358	0.399	0.264	*−0.135*
8	0.374	0.509	0.250	*−0.259*	0.403	0.447	0.304	*−0.143*
9	0.426	0.565	0.294	*−0.271*	0.449	0.493	0.349	*−0.144*
10	0.479	0.613	0.346	*−0.267*	0.497	0.542	0.387	*−0.155*
11	0.527	0.650	0.390	*−0.260*	0.541	0.581	0.435	*−0.146*
12	0.577	0.703	0.452	*−0.251*	0.590	0.624	0.485	*−0.139*
13	0.627	0.737	0.495	*−0.242*	0.630	0.667	0.527	*−0.14*
14	0.685	0.787	0.561	*−0.226*	0.684	0.718	0.587	*−0.131*
15	0.726	0.810	0.613	*−0.197*	0.724	0.749	0.634	*−0.115*
16	0.784	0.858	0.685	*−0.173*	0.778	0.799	0.693	*−0.106*
17	0.828	0.889	0.745	*−0.144*	0.821	0.841	0.752	*−0.089*
18	0.876	0.916	0.808	*−0.108*	0.870	0.881	0.819	*−0.062*
19	0.930	0.995	0.889	*−0.106*	0.924	0.930	0.890	*−0.040*
20	0.962	0.973	0.934	*−0.039*	0.959	0.961	0.940	*−0.021*
Value Added	−0.001	−0.035	0.026	*0.061*	0.000	−0.055	0.055	*0.11*

Notes: This table categorizes elementary schools into "good", "bad", and "average" in terms of average third-grade attainment and third- to eighth-grade value added. "Good" ("bad") is defined as being one standard deviation above (below) the average (with tolerance of 0.0045). Each row shows the average class rank of students in this type of school for that ventile. The Rank Change column is the average rank change of students in that ventile from moving from a "bad" to a "good" school. The final row presents the third- to eighth-grade school-level value added for this type of school. Value added is calculated conditional on a cubic of third-grade percentile.

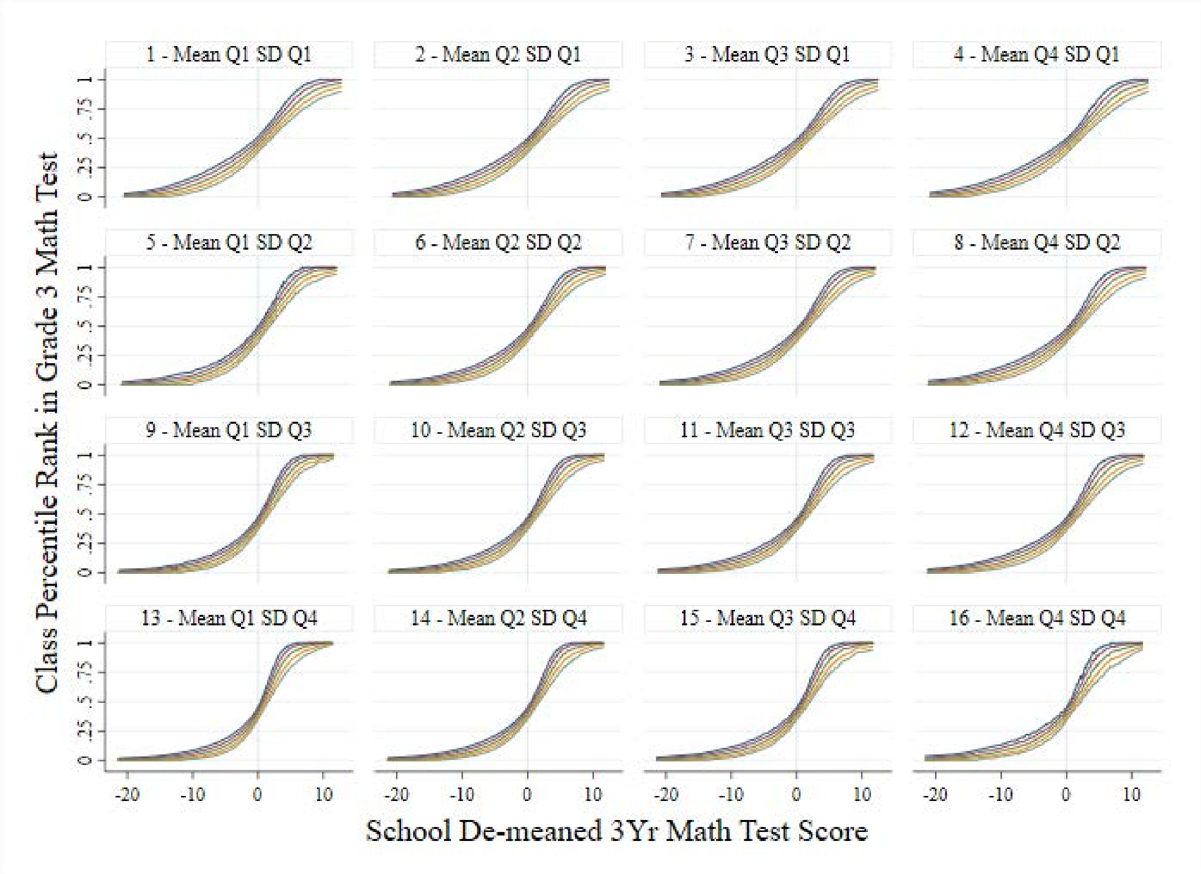
Appendix Figures and Tables

Appendix Figure 1: Common Support of Class Test Score



Notes: This figure shows a student's rank for a given test score relative to the mean. We show the differences in rank across SSCs. Each line represents a percentile of students with that test score relative to the mean. (e.g. 10th percentile). The vertical spread shows the different rank values held by students with the same relative position within their classes. The red line represents where the students from Figure 1, who are five points above their class mean would reside.

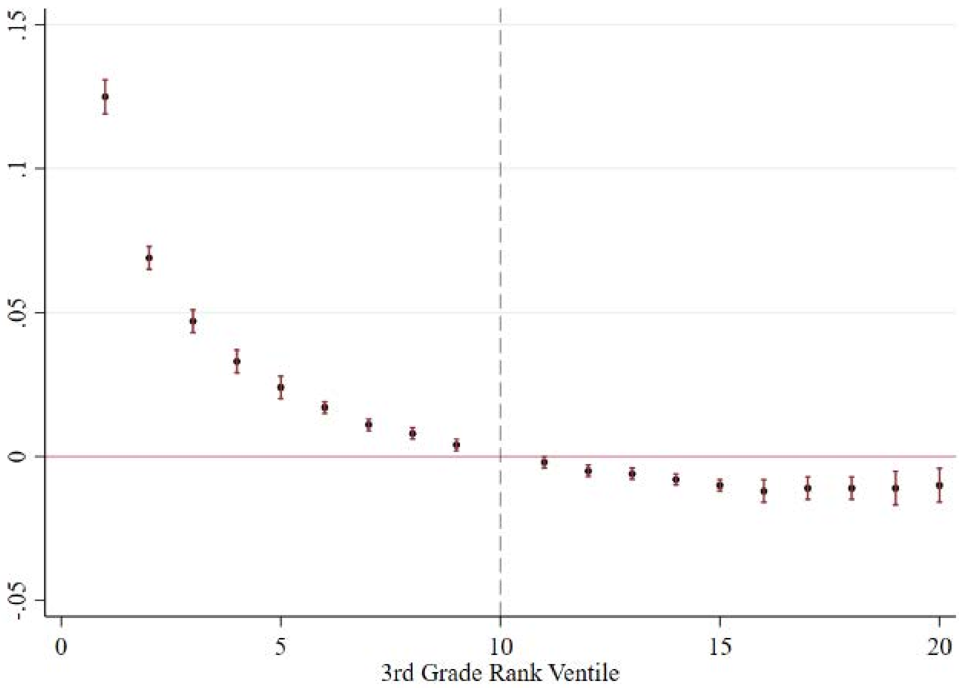
Appendix Figure 2 – Variation in Local Rank of Median Student Rank Within School-Subject Groups



Notes: These figures show the rank for a student for a given test score relative to the mean in each school distribution type $g_d(\quad)$. We show the differences in rank across SSCs. Each line represents a percentile of students with that test score (e.g. 10th percentile) relative to the mean. The vertical spread shows the different rank values held by students with the same relative position within their classes.

Appendix Figure 3 – Testing for Missingness

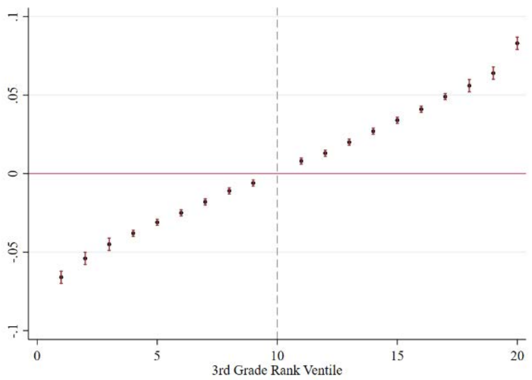
A. Missing Eighth-Grade Test



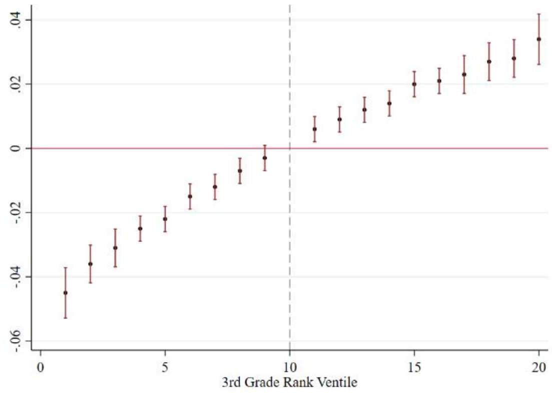
Note: This figure plots the coefficient for ventiles of class rank with 95% confidence intervals calculated using standard errors clustered at the school level. The 45th to 50th percentile interval is the omitted category. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement. The mean retention rate is 1.6.

Appendix Figure 4 – Balanced Sample for Full Set of Outcomes

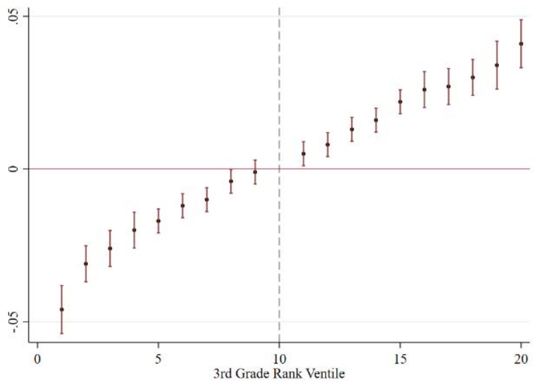
A. Eighth-Grade Test



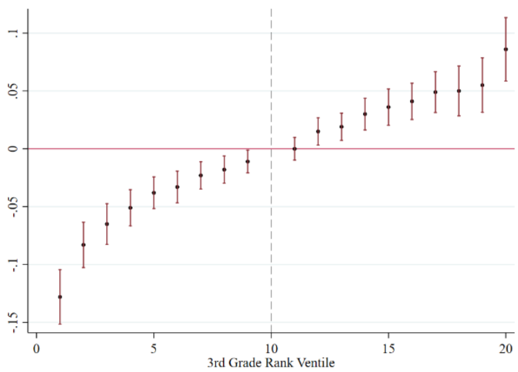
B. Ever Graduate HS



C. Any College



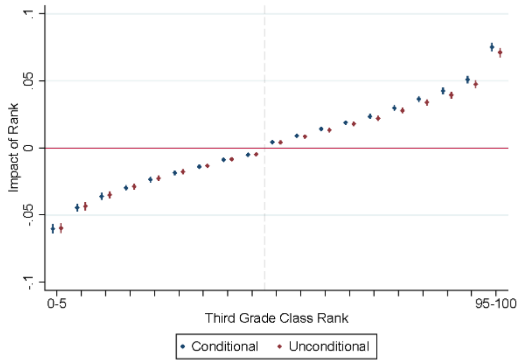
D. Log Real Earnings Age 23–27



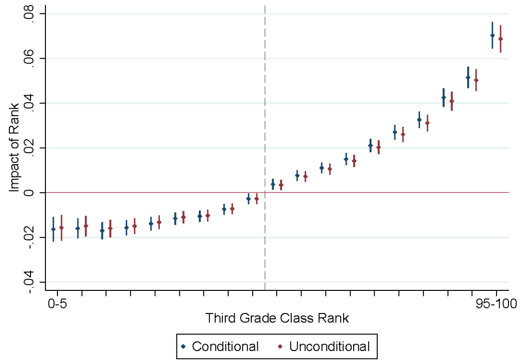
Notes: These figures are constructed on the reduced balanced Sample for Full Set of Outcomes with 8 Cohorts Third Grade 1995–2002. They plot the coefficient for ventiles of class rank. The 45th to 50th percentile interval is the omitted category. Estimates come from Equation 7, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement.

Appendix Figure 5 – Conditional/Unconditional Estimates

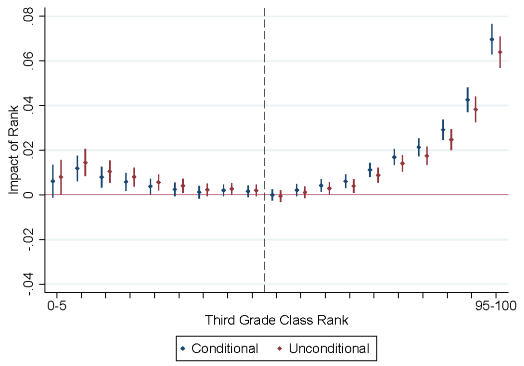
A. Eighth-Grade Test



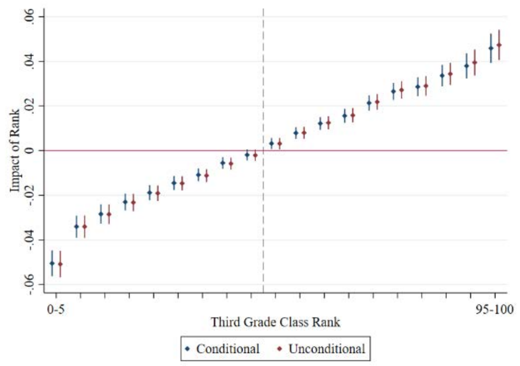
B. Ever Graduate HS



C. Any College



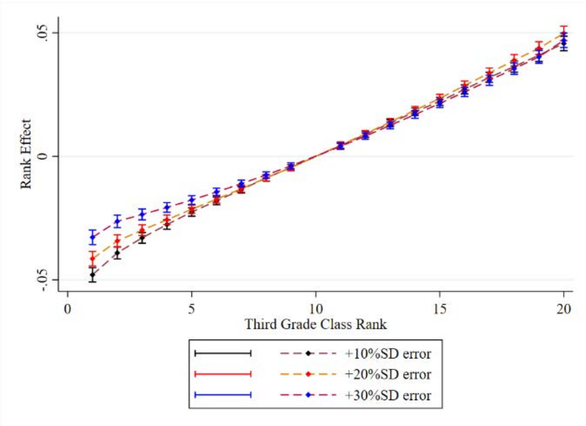
D. Log Real Earnings Age 23–27



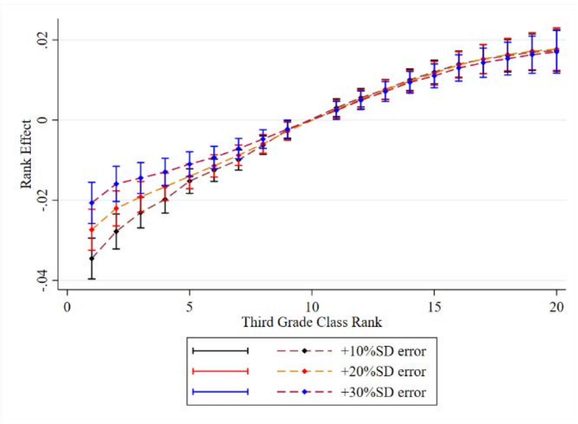
Notes: These figures plot the coefficient for ventiles of class rank with 95% confidence intervals calculated using standard errors clustered at the school level. The 45th to 50th percentile interval is the omitted category. Estimates come from Equation 4, which includes controls for race, gender, ESL status, and indicators for ventiles of student achievement, not interacted by school test score distribution. We show how the estimates change when including controls for predetermined characteristics including race, gender, and ESL status (our preferred specification) and when we do not.

Appendix Figure 6 – Measurement Error Simulations with Normally Distributed Measurement Error

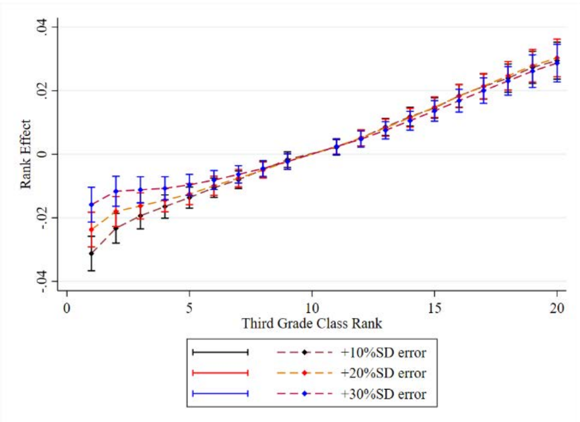
A. Eighth-Grade Test



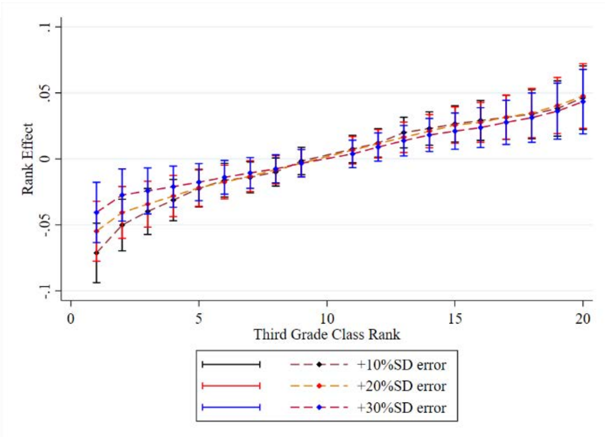
B. Ever Graduate HS



C. Any College



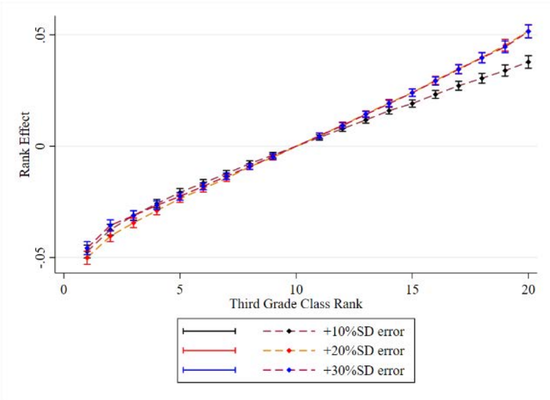
D. Log Real Earnings Age 23–27



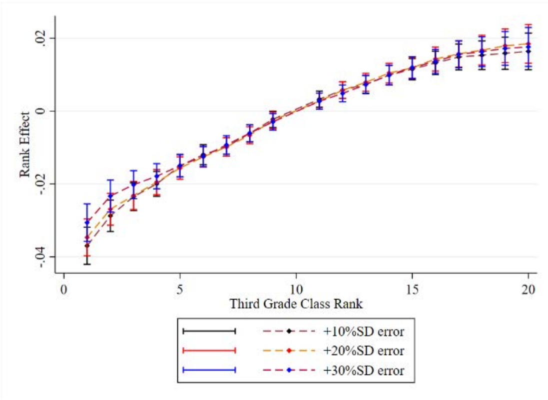
Notes: These figures show estimates for our main specification where we added normally distributed noise with zero mean and a standard deviation equivalent to 10%, 20% and 30% of the standard deviation in the third -grade achievement before calculating ranks. The resulting nonlinear and correlated measurement error in third-grade achievement and the rank measure results in a nonlinear downward bias in the rank estimate.

Appendix Figure 7 – Measurement Error Simulations with Uniformly Distributed Measurement Error

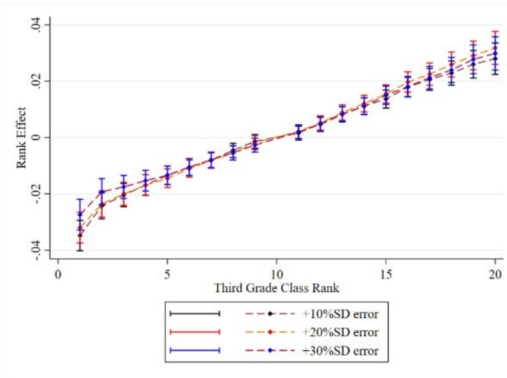
A. Eighth-Grade Test



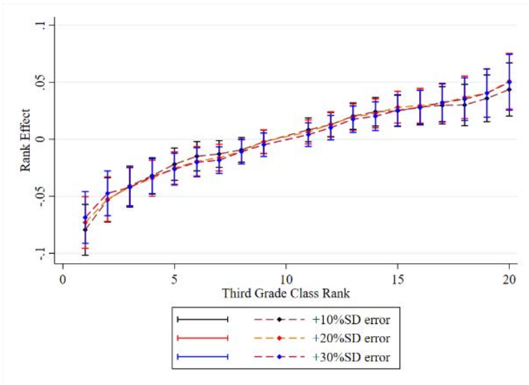
B. Ever Graduate HS



C. Any College



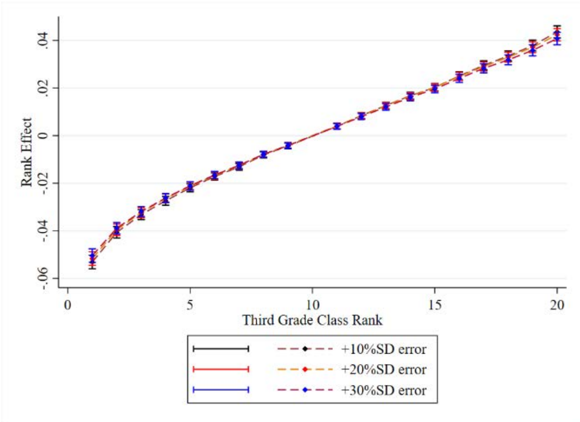
D. Log Real Earnings Age 23–27



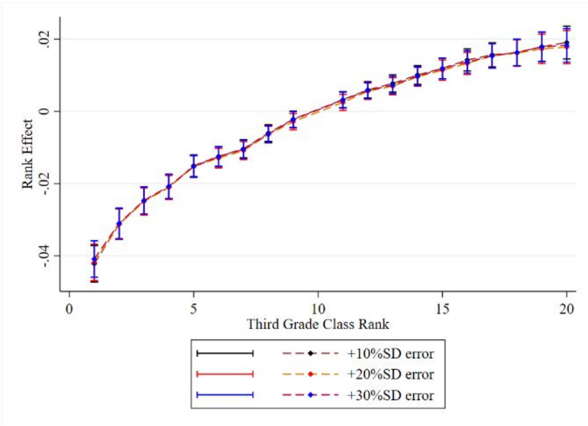
Notes: These figures show estimates for our main specification where we added uniformly distributed noise with zero mean and a maximum/minimum value equal to ±10%, ±20%, and ±30% of the standard deviation in the third-grade achievement before calculating ranks.

Appendix Figure 8 – Measurement Error Simulations with Heterogenous Measurement Error

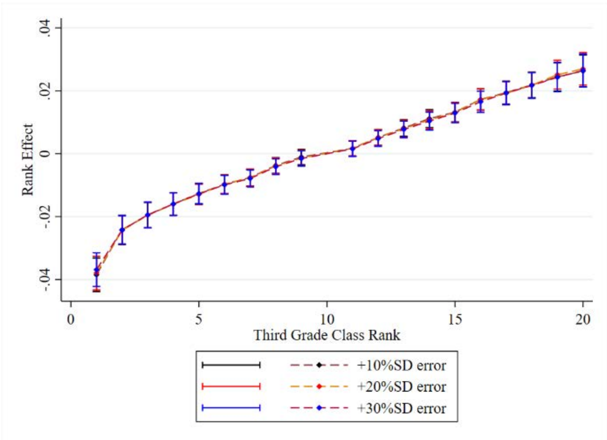
A. Eighth-Grade Test



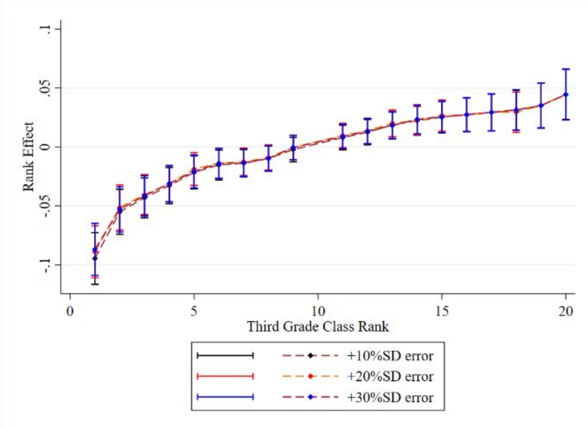
B. Ever Graduate HS



C. Any College

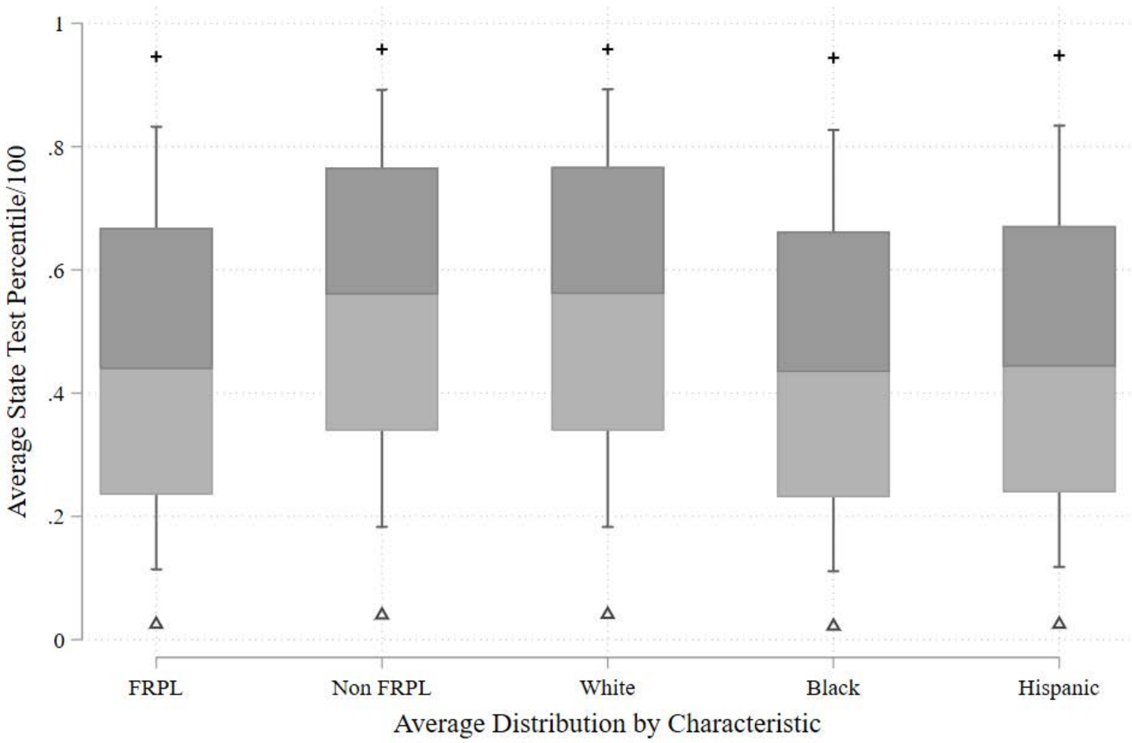


D. Log Real Earnings Age 23–27



Notes: These figures show estimates for our main specification where we added measurement error drawn from a normal distribution transformed by twice the student's absolute distance from the median plus 0.5, $(2|\alpha|) + 0.5$, namely $n(0, \vartheta) * (2|\alpha| + 0.5)$. This is to ensure the mean of the scaling factor is 1.

Appendix Figure 9 – Average Classroom Distributions by Student Characteristics



Notes: This figure displays the average classroom distribution statistics for students with different characteristics. The triangle shows the average minimum test score (in statewide percentiles), the whiskers show the class average 10th and 90th percentiles, and the plus sign shows the average maximum. The rectangles show the interquartile range with the median separating the two colors.

Appendix Table 1 – Alternate Measures of Rank on Main Outcomes

	Method for Calculating Rank			
	Mean Rank	Bottom Rank	Random Rank	On-Time Students
	(1)	(2)	(3)	(4)
Outcome				
Test Scores	0.149***	0.155***	0.100***	0.147***
	(0.003)	(0.002)	(0.002)	(0.003)
Grad High School	0.085***	0.078***	0.059***	0.083***
	(0.004)	(0.004)	(0.004)	(0.004)
Any College	0.088***	0.090***	0.060***	0.086***
	(0.005)	(0.004)	(0.004)	(0.005)
Real Earnings	1853.1***	1983.5***	1482.3***	1785.8***
	(251.1)	(244.6)	(213.7)	(247.7)

Notes: This table presents the rank estimates from 16 different regressions, using four different rank measures on four outcomes. Mean Rank assigns the average rank to all students tied with the same score (this is the measure we use in the paper). Bottom Rank assigns the bottom rank to all students tied with the same score. Random Rank assigns a random rank to all students with tied with the same score. On-time students assigns rank on a mean rank basis only among students who took their third-grade exam on time, rather than all students in their class who took the exam. Standard errors are clustered at the school level.

Appendix Table 2 – Outcomes by Function of Distribution

	Repeat 3rd	Grade 8 Test	Grad HS	Any College	Grad BA in 8 yrs	Log Wage
			A. Uninteracted			
Rank	−0.037***	0.089***	0.061***	−0.007	−0.013**	0.121***
	(0.005)	(0.003)	(0.004)	(0.005)	(0.005)	(0.017)
			B. School Mean Quartiles			
Rank	−0.038***	0.141***	0.084***	0.084***	0.019***	0.171***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.020)
			C. School Mean Deciles			
Rank	−0.040***	0.152***	0.088***	0.092***	0.027***	0.185***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.021)
			D. School Variance Quartiles			
Rank	−0.039***	0.106***	0.083***	0.041	−0.002	0.144***
	(0.003)	(0.002)	(0.004)	(0.004)	(0.005)	(0.017)
			E. School Variance Deciles			
Rank	−0.040***	0.113***	0.089***	0.050***	0.005	0.152***
	(0.003)	(0.002)	(0.004)	(0.004)	(0.005)	(0.018)
			F. School Quartiles of Mean and Quartiles of Variance			
Rank	−0.036***	0.147***	0.087***	0.088***	0.029***	0.173***
	(0.003)	(0.003)	(0.004)	(0.005)	(0.005)	(0.020)
N	6,117,651	4,919,628	6,117,651	6,117,651	2,647,234	2,652,264

Notes: This table considers the linear effect of rank for different specification of $g_d()$. Panel A controls for student prior achievement with ventiles for third-grade achievement. Panel B interacts achievement with quartile of school mean achievement. Panel C interacts achievement with deciles of school variance. Panel D interacts achievement with quartile of school achievement variance. Panel E interacts achievement with deciles of school achievement variance. Panel F is our preferred specification and interacts achievement with 16 indicators for school mean and variance (quartiles of mean × quartiles of variance). Standard errors are clustered at the school level.

Appendix: Measurement Error

Consider the simplest of situations where we have only one group and two explanatory variables, individual i's human capital $T_i{}'$ and true class ranking $R_i{}'$. Assume $T_i{}'$ cannot be measured directly, so we use a test score T_i as a baseline, meaning a noisy measure of true human capital which has measurement error $T_i = l_i(T_i' , e_i)$. We also use this observed test score T_i in combination with the test scores of all others in that group, T_{-i}, to generate the rank of an individual, $R_i = k(T_i , T_{-i})$. We know this rank measure is going to be measured with error ϵ_i , such that $R_i = h_i(R_i{}' , \epsilon_i)$. The problem is that this error, ϵ_i, is a function of the student's ability $T_i{}'$ and their own measurement error , e_i, but also depends on the ability and measurement errors of all other individuals in their group (T'_{-i}, e_{-i}, respectively): $R_i = k\big(l_i(T_i' , e_i), l_{-i}(T'_{-i} , e_{-i})\big) = f(T_i' , e_i, T'_{-i} , e_{-i})$. Therefore, any particular realization of e_i not only causes noise in measuring T_i', but also in $R_i{}'$. This means we have correlated nonlinear and nonadditive measurement error, where $COV(e_i, e_{-i}) \neq 0$. This specific type of nonclassical measurement error is not a standard situation, and it is unclear how this would impact the estimated rank parameter.