NBER WORKING PAPER SERIES

DOES SORTING STUDENTS IMPROVE SCORES? AN ANALYSIS OF CLASS COMPOSITION

Courtney A. Collins
Li Gan

Working Paper 18848
http://www.nber.org/papers/w18848

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2013

Does Sorting Students Improve Scores? An Analysis of Class Composition
Courtney A. Collins and Li Gan
NBER Working Paper No. 18848
February 2013
JEL No. I21

ABSTRACT

This paper examines schools' decisions to sort students into different classes and how those sorting processes impact student achievement. There are two potential effects that result from schools creating homogeneous classes—a "tracking effect," which allows teachers to direct their focus to a more narrow range of students, and a peer effect, which causes a particular student's achievement to be influenced by the quality of peers in his classroom. In schools with homogeneous sorting, both the tracking effect and the peer effect should benefit high performing students. However, the effects would work in opposite directions for a low achieving student; he would benefit from the tracking effect, but the peer effect should decrease his score. This paper seeks to determine the net effect for low performing students in order to understand the full implications of sorting on all students.

We use a unique student-level data set from Dallas Independent School District that links students to their actual classes and reveals the entire distribution of students within a classroom. We find significant variation in sorting practices across schools and use this variation to identify the effect of sorting on student achievement. Implementing a unique instrumental variables approach, we find that sorting homogeneously by previous performance significantly improves students' math and reading scores. This effect is present for students across the score distribution, suggesting that the net effect of sorting is beneficial for both high and low performing students. We also explore the effects of sorting along other dimensions, such as gifted and talented status, special education status, and limited English proficiency.

Courtney A. Collins
Rhodes College
1400 Coleman Avenue
Macon, GA 31027
ccollins1670@gmail.com

Li Gan
Department of Economics
Texas A&M University
College Station, TX 77843-4228
and NBER
gan@econmail.tamu.edu

1. Introduction

While a large body of literature exists on education inputs such as teacher quality, school expenditure, and class size, a related but lesser-studied issue is how students are actually divided into classes. Schools may use several different strategies to allocate a given number of students within a grade across different classrooms. Some schools may choose to sort students by ability level and create classes of relatively homogeneous students. Alternatively, schools may choose to sort students with varying abilities evenly across classes. Other schools may try to match students and teachers, while taking into account individual students' learning styles. Analyzing these types of sorting decisions is important because if a particular sorting mechanism is found to be especially beneficial, many schools would have the ability to implement valuable changes without the need for large amounts of additional resources. Other changes to educational inputs—including technology available to students, reducing class size, and attracting more qualified teachers—require substantial increases in spending. Changing the sorting guidelines for a particular school only requires rearranging the students among the classes that already exist and could be accomplished, in many situations, with minimal additional resources.

The decision to sort students by ability has two distinct effects on students. The first, the direct tracking effect, produces efficiency gains for teachers in the way that they are able to structure their classrooms and pedagogy. Sorting allows teachers to narrow their instruction to a particular group of students and to tailor their teaching styles to meet the needs of those specific students. This should be beneficial to all students, including both high and low performers. Teachers in low ability classrooms will be able to focus on foundational skills that are imperative to the continued progress of their students, and teachers in high ability classrooms will be able to spend time on more advanced material that would be otherwise omitted in a more heterogeneous setting. Both groups should see improvements in performance as a result of this effect.

The second consequence of ability sorting is the resulting peer effect. Schools that homogeneously sort by ability necessarily place high ability students with high ability peers and low ability students with low ability peers. If students are directly influenced by the quality of their classmates, this would benefit students at the high end of the score distribution while hurting students who are already at the low end. The net effect of sorting, then, is positive for high ability individuals but undetermined for low ability individuals. If the peer effect is greater than the tracking effect, then low scoring students are on balanced harmed by schools' sorting procedures. However, if efficiency gains from tracking outweigh the peer effect, then sorting should have an overall beneficial impact on students across the score distribution.

The purpose of this study is threefold. First, we determine how schools sort students into various classes using student-level data from Dallas Independent School District. This data is unique in that it allows a student to be linked not only to his school, but also to his individual class. This allows us to observe the entire distribution of students within a class and evaluate sorting along a number of dimensions. We examine sorting primarily by previous test scores, in addition to gifted and talented status, special education status, and limited English proficiency (LEP). We find evidence of substantial variation in sorting policies across schools.

Second, we use the variation that exists across schools to identify the effect of sorting on student performance. We construct several "sorting indices" that measure how homogeneous or heterogeneous the classes within a particular school are, based on the different dimensions described above. Applying an innovative instrumental variable technique that uses one grade's sorting index as an instrument for the sorting index of another grade, we estimate the impact of sorting on student achievement.

Third, we consider that the effects of sorting are not necessarily the same across different types of students and we allow for heterogeneity in the sorting effect across a distribution of students. We determine the differences in the effects between high and low scoring students in addition to considering effects for students who are classified as gifted and talented, special education, or LEP. We pay careful consideration to students at the bottom end of the score distribution to determine if sorting has an overall positive or negative effect for them, revealing the relative impacts of the tracking effect compared to the peer effect.

2. Related Literature

There are two important strains of literature related to this paper: the peer effects literature and the tracking literature. Studies within the tracking literature typically characterize the classes in a school or group of schools as homogeneous or heterogeneous and attempt to analyze the effects on student outcomes. The definition of tracking may vary from study to study, as there is a wide assortment of actual tracking practices in reality, ranging from explicitly creating clearly defined and curriculum-distinct "tracks" for particular types of students to simply dividing students into groups based on ability level and providing them all with roughly the same program of study. Tracking also exists at different levels of instruction. Some school systems practice within-school tracking and others divide students across schools based on various measures. These methods vary largely across countries. Betts (2011) provides a thorough review of this literature.

Several papers provide descriptive statistics using large, nationally representative samples. Rees et al. (1996) report data from the National Education Longitudinal Study of 1988 (NELS). Using teacher-reported information, they conclude that the majority of 8th and 10th grade classes appear to be tracked by performance. For example, only about 14 percent of 8th grade students were enrolled in a heterogeneous math class. They also find that students from low income groups and minority students are more likely to be enrolled in lower-level classes and less likely to be enrolled in advanced classes. Betts and Shkolnik (2000), using data from the Longitudinal Study of American Youth (LSAY), examine the qualities of tracked classes in middle school and high school and find differences in average teacher characteristics between low ability and high ability classes. Low ability classes tend to be smaller, but they also tend to have teachers with less experience and education.

In estimating the causal effect of tracking on student achievement, one of the key issues confronted by researchers is how to deal with the selection problem that arises because students are not randomly assigned to a tracked or non-tracked school—or to a particular class within a school. Hoffer (1992) compares tracked and non-tracked schools at the middle school level. Using a propensity score approach to attempt to combat the selection issue, he finds heterogeneous impacts by student type: tracking seems to be beneficial for high-scoring students, but he finds negative effects for low-scoring students. Betts and Shkolnik (2000) use principal-level survey data and student testing scores and compare achievement for students of several different ability levels across schools that track and those that do not. They find some evidence for heterogeneous effects but argue that they are relatively small and, in most cases, not significantly different from zero.

Hanushek and Woessmann (2006) exploit variations in tracking practices across countries to determine if a country-wide system of creating curriculum-specific tracks at the secondary school level is beneficial for students. Using a difference-in-difference method that examines changes in outcomes between primary and secondary schools for tracked and non-tracked countries, they find strong evidence of increased inequality in test scores as a result of tracking. They also suggest that tracking may actually *decrease* average performance overall, so that both high and low achieving students in a tracked system lose relative to students in a non-tracked system.

Figlio and Page (2002) explore tracking in conjunction with school choice using data from NELS. They address the endogeneity problem by instrumenting for whether or not a school tracks students by using several county-level instruments based on graduation requirements, number of schools, and presidential voting data. They find no evidence that tracking benefits high-scoring students at the cost of low-scoring students; instead they suggest that tracking may in fact *benefit* low-scoring students as well.

Duflo et al. (2011) also provide support for the idea that tracking may help both high-scoring and low-scoring students. In order to avoid the selection problem, they conduct a randomized trial in Kenya by funding an extra teacher for primary schools with only one class in a particular grade. After creating the extra class, they divide the students into either homogeneous or heterogeneous classes, with respect to their previous testing scores. The authors find that being enrolled in the homogeneous classes significantly increases students testing scores, and that the effect holds for both high-scoring and low-scoring students.

Related work in the peer effects literature attempts to identify the effects of a student's peer group on his own behavior. If peer effects exist, then an individual is influenced by the characteristics or decisions of those around him. This has clear implications for sorting because dividing students into separate classes based on student characteristics automatically places them in a particular peer group. While homogeneous sorting would result in gains for high ability students who would now have a larger proportion of high achieving classmates, there are potential harmful effects for low ability students who now have more low achieving classmates.

Lefgren (2004) uses variation in tracking policies to evaluate peer effects and highlights the relationship between the two. Using student-level data from Chicago Public Schools, he measures the degree of tracking within a school by the amount of variation in students' incoming test scores that can be explained in a regression of classroom dummy variables. He implements an instrumental variable strategy based on interactions between the level of tracking within a school and observed student ability. His findings provide evidence of significant but very small peer effects.

Other papers examine peer effects beyond the context of tracking. Sacerdote (2001) analyzes peer effects at the college level by exploiting random variation in peers caused by random assignment in college housing. Using a dataset of about 2,000 college freshmen from Dartmouth College, all of whom have randomly assigned roommates, he examines the impact of peer groups—both at the roommate level and at the dorm level—on several variables, including choice of major, GPA, and fraternity membership. He finds convincing evidence of peer effects for several outcomes. Specifically, a student's peers have a strong impact on whether he will join a fraternity, and which fraternity he will join. This effect is particularly strong at the dorm level. He also finds effects for GPA, especially between roommates. An individual's GPA is positively and significantly impacted the GPA of his roommate.

Ding and Lehrer (2007) provide an analysis of peer effects in secondary schools using data from Jiangsu Province in China. They are able to avoid the selection problem inherent in many peer effects studies because the students in their data set are assigned to a particular school based only on observable test scores, rather than on unobservable characteristics. Using a panel of about 1,300 students, whom they follow from middle school completion into college admission, they find strong evidence of peer effects. They conclude that peer effects are heterogeneous with respect to student type: high-performing students benefit significantly more from having high-performing peers than do low-performing students. They also conclude that all students benefit from having more homogeneous peer groups (or peer groups with less score variation), although the impact is again larger for high-performing students.

Lavy et al. (2011) use administrative data from high schools in Israel to determine the mechanisms through which peer effects work. Using students' birthdays to qualify them as "repeaters" and potentially of lower ability, they exploit variation in the proportion of low ability students within a classroom. They then use that variation to determine the peer effect and conclude that more low ability students within a class negatively impacts the performance. Using a student-level questionnaire, the authors describe several channels through which the peer effect may work, documenting effects on disruptive classroom behavior and relationships among students.

3. A Sorting Model

3.1 Sorting by Test Score

To examine the effect of sorting on students' academic achievement, we begin by considering the following model:

$$s_{ijkt} = \rho s_{ijkt-1} + \phi\gamma_{jkt} + X_{ijkt}\beta + C_{jkt}\eta + \varepsilon_{ijkt}, \quad (1)$$

where s_{ijkt}represents the test score of student *i* in class *j* at school *k* in time *t* and s_{ijkt-1}represents the same student's previous year score. We include X_{ijkt}, a vector of student-level controls, and C_{jkt}, a vector of classroom characteristics.

The variable γ_{jkt} is a sorting index for a class *j*, describing the dispersion of the students in the classroom based on observable characteristics, such as test score. Higher levels of γ_{jkt} indicate a class that is sorted in a more homogeneous way. Lower levels of γ_{jkt} indicate more heterogeneous sorting, or that the test scores in the class are more evenly dispersed. It follows that positive values of ϕ suggest that homogeneous sorting is helpful in improving student performance, and that negative values provide support for hypothesis that heterogeneous classes are more likely to improve achievement.

It is possible that a particular type of sorting will have different effects for different groups of students. For example, sorting high-scoring students into one class and low-scoring students into another class may allow the classes to move at different paces, which may benefit both groups of students. The teacher in the low-scoring class may be able to focus on foundational skills necessary to the improvement the students, while the teacher in the high-scoring class may have the opportunity to move on to new, more challenging material without the fear of losing the understanding of the class.

However, this type of sorting may not necessarily benefit both groups. An alternative hypothesis is that by creating evenly distributed groups, students with more understanding of the material may be able to help those with less understanding. In this situation, low-scoring students might benefit without causing a cost for high-scoring students. It may even be plausible that this situation could benefit *both* high scorers and low scorers.

In order to examine how sorting may affect different types of students, we allow φ to vary by students' observable characteristics, as shown in the following model:

$$s_{ijkt} = \rho s_{ijkt-1} + \phi_1\gamma_{jkt}(high_{ijkt}) + \phi_2\gamma_{jt}(low_{ijkt}) + X_{ijkt}\beta + C_{jkt}\eta + \varepsilon_{ijkt}, \quad (2)$$

where students are ranked by their previous test score and placed into one of two groups. The variable $high_{ijkt}$ is equal to one if student *i* is a high-scoring student, and low_{ijkt} is equal to one if student *i* is a low-scoring student.

3.2 Other Methods of Sorting

In addition to examining schools' decisions to sort by previous score, we also consider their decisions to sort according to other characteristics, such as gifted and talented (GT) classification or special education status. For example, some schools may group all of its GT students into a single class to allow them to move at their own pace, while other schools may divide them into several classes with other non-GT students.[3] Having GT students included in a regular classroom could potentially help or hurt non-GT students in the same ways that high-scoring students could affect low-scoring students.

Similar logic holds for special education students. Some schools create separate classrooms for special education students, while other schools include those students in regular education classrooms. Either of these sorting processes may have implications for both types of students—and those implications may be different for special education students, compared with regular education students.

we empirically examine both GT and special education sorting and allow the sorting effect to vary across student type, in a model similar to model (2). we also include measures of sorting by limited English proficiency (LEP) status, which may be particularly relevant for the state of Texas, which serves a large population of LEP students[4].

3.3 A Sorting Index

To empirically determine the effects of both types of sorting, we first construct a measure defining how "sorted" a class is. Consider the following measure for each grade within a school:

$$\alpha_{1k} = \sqrt{\frac{1}{N}\sum\left(s_{ijk} - \bar{s_k}\right)^2} \qquad (3)$$

[3] Even if GT students are divided into classrooms with many non-GT students, they still may be "pulled out" for several hours during the school day or during the week. Unfortunately, the Dallas ISD data contains only one classroom per student, so it is not possible to tell if the students participate in this type of program.
[4] For example, more than 20 percent of students in Dallas ISD are classified as LEP students.

where s_{ijk} represents the score of student *i* in class *j* and school *k*, $\bar{s_k}$ is the score average in school *k*, and *N* represents the total number of students in school *k*. The index α_{1k}, then, measures overall score dispersion within a school. Likewise, consider the following measure which represents the level of dispersion within an individual classroom:

$$\alpha_{2jk} = \sqrt{\frac{1}{N_j}\sum\left(s_{ijk} - \bar{s_j}\right)^2} \tag{4}$$

where $\bar{s_j}$ is the score average within class *j* and N_j represents the total number of students in class *j*. Using the two above measures, we define the following index for each classroom:

$$sort_{jk} = \frac{\alpha_{1k}}{\alpha_{2jk}}, \tag{5}$$

which reflects the score dispersion in classroom *j* relative to overall score dispersion in school *k*. Higher values of $sort_{jk}$ indicate less variation in the classroom scores relative to the overall variation in the school and suggest that the class is homogeneously sorted. Lower values of $sort_{jk}$ indicate that more variation in the classroom and suggest that students are distributed in a more heterogeneous way.

We define a similar measure to gauge sorting along other dimensions, such as gifted and talented, special education, or LEP status. Consider the following index:

$$sort_{jk}^{z} = \frac{\alpha_{1k}^{z}}{\alpha_{2jk}^{z}}, \tag{6}$$

for $z = GT, SP$, and LEP. This index is analogous to equation (5), except that the testing scores are replaced with a binary variable indicating GT, special education, or LEP status. Consider the sorting process for GT students. Higher values of $sort_{jk}^{GT}$ indicate that schools place gifted and talented students into a separate class (or classes), rather than dispersing them evenly across all classes, which would be suggest by a low value of $sort_{jk}^{GT}$.

3.4 Endogeneity of the Sorting Index

It is essential to consider not only the effect of sorting on students' scores but also why they are sorted into their given classes at the outset. Although the dataset allows identification of characteristics such as previous score and other student classifications, teachers and principals certainly observe many other variables which may be used to divide students into different classrooms. Principals may attempt to "match" certain students with certain teachers, or they may have policies whereby parents can request a certain teacher for their children.

Unobserved variables such as behavior may also play an important role in the classroom assignment process. For example, if a principal observes that several students have had behavior problems in the past, he may try to divide those students evenly across the classes within a grade, or he may assign them to a particular teacher who has had previous success with behavioral problems. In this case, behavior is an unobserved variable that affects a school's sorting index. An endogeneity problem arises if behavior, or other unobserved variables correlated with sorting, impact students' test scores.

In order to deal with this endogeneity, we propose to instrument for one grade's sorting index using the sorting index from another grade at the same school. If the administration at school k uses certain guidelines in assigning students to classes in grade g, it is likely that those guidelines are also used for other grades in school k. Therefore, the sorting indices for classes in grade g should be correlated with the sorting indices for grade $g+1$. However, there is no reason to believe that the way in which classes are sorted in grade $g+1$ should impact the scores of students in grade g. Therefore, sorting indices in grade $g-1$ should provide valid instruments for sorting indices in grade g.

The problem that arises when trying to match indices from individual classes across grades is that there is no way to map the classes from fourth grade to specific fifth grade classes. Sorting indices for each class within a grade can be created, but they cannot be mapped one-to-one across grades. Instead, we create a grade-specific sorting measure that can be used for all classes within a grade. Instead of using the classroom-specific α_{2jk}, we create a grade-specific measure $\overline{\alpha_{2k}}$, which is simply the average score dispersion across all classes within the grade:

$$\overline{\alpha_{2k}} = \sqrt{\frac{1}{J}\sum \frac{1}{N_j}\sum \left(s_{ijk} - \bar{s}_j\right)^2}. \qquad (7)$$

The variable J indicates the number of classrooms in school k. The sorting index we use in the empirical analysis, then, is

$$sort_k = \frac{\alpha_{1k}}{\alpha_{2k}}, \quad (8)$$

and its interpretation is similar to that of equation (5). Higher values of $sort_k$ indicate less dispersion of scores within classes relative to score dispersion with the school, which means more sorting (or more homogeneous classes). Lower values of $sort_k$ indicate more dispersion of scores within classes, which means less sorting (or more heterogeneous classes). In the empirical estimation, we use $sort_k$ for the fifth grade to instrument for $sort_k$ for the fourth grade.

Table 8 reports the correlation between fourth and fifth grade sorting indices, with one observation per school. Correlation coefficients range from 0.37 to 0.57.

4. Data

4.1 Student-Level TAKS Data from Dallas ISD

One drawback to many datasets used to explore how school inputs or characteristics affect achievement is that students cannot typically be linked to their actual classes. For example, the Texas Education Agency (TEA) collects student-level testing data from the Texas Assessment of Knowledge and Skills (TAKS) for all public school students starting in third grade. The dataset is a rich panel of information on student achievement and characteristics. However, while students' schools and grade levels are available in the dataset, their specific classes are not. This makes it impossible to link a particular student to his classroom and to actually determine how schools have divided students across classes within a grade.

While students are not linked to specific classes in the statewide dataset, several school districts do collect student-level data that may be linked to a class variable. We employ a unique dataset from Dallas Independent School District that contains both class and grade identifying information. This allows me not only to track a student to his class, but also to examine the distribution of scores and other characteristics within and across classrooms for a particular grade.

The dataset includes student-level math and reading TAKS scores for two school years. we examine all third grade students in the 2003-2004 school year who become fourth graders in 2004-2005, a total of 9,325 children from 135 different schools in Dallas ISD. In addition to achievement scores for both years, the dataset contains race and gender variables and identifiers for students qualifying for programs such as free or reduced lunch, gifted and talented, special education, and limited English proficiency. Summary statistics are shown in table 1.

4.2 Test Score Variables

Texas reports students' scores in two ways. The first score is a student's raw score, which corresponds to the number of questions he answered correctly on the exam. For the 2004-2005 exam, the maximum raw score is 42 points. The second score measure is a student's scale score, which is scaled using the Rasch partial credit method to control for the difficulty of the exam across different administrations of the test. Scale scores are used to compare two different cohorts' scores. For example, scale scores could be used to compare fourth graders in 2004 with the following group of fourth graders, who took the exam in 2005.

Although the scores allow for direct comparison in this way, they are not meant to be vertically linked. That is, a third grader's 2004 score should not be directly compared to his fourth grade 2005 score in order to gauge improvement. Because that is precisely the comparison we want to make, we convert the scale scores into z-scores, by subtracting out the mean score and dividing by the standard deviation in a given year. A student's z-score is given by

$$s_{it} = \frac{scale_{it} - \mu_t}{\sigma_t}, \tag{9}$$

where $scale_{it}$ is student *i*'s scale score in period *t*, and μ_t and σ_t represent the mean and standard deviation of the scale scores. A student's score is now a representation of where he lies along the distribution of scores. we generate z-scores for both the current year (2004-2005) and the previous year (2003-2004).

5. Empirical Results

5.1 Do Schools Sort?

Before examining any effects of sorting or class size, it is first important to determine whether any schools appear to sort students based on observable characteristics and how prevalent this type of sorting is. We explore potential sorting based on several different observable characteristics: previous math score, previous reading score, gifted and talented status, special education status, and LEP status.

To investigate sorting based on students' previous scores, we create dummy variables for each class and compare the mean scores by running the following regression:

$$s_{ijt-1} = \beta_1 + \sum_{j=2}^{J} \beta_j D_j + \varepsilon_{ij}, \tag{10}$$

where s_{ijt-1} is student *i*'s test score in the previous year and D_j is a dummy variable for class *j*. Therefore, β_1 gives the mean score for the first class and $\beta_2, \beta_3, \ldots, \beta_J$ show the differences in score relative to the first class. If schools divide their students into classes randomly, then there should be no difference in the previous year score means for any of the classes. That is, $\beta_2, \beta_3, \ldots, \beta_J$ should not be significantly different from zero or from each other.

Alternatively, if schools do divide students into classes based on their previous year scores, then there should be significant differences in the average scores. Consider the case in which a school has three classes within a single grade. The administration may choose to sort students into three groups—low-scoring students who need additional assistance to improve their grades, average-scoring students who are achieving at grade-level, and high-scoring students who are ready to move on to more challenging material. In this case, β_2, and β_3 would be significantly different from zero, as well as different from each other.

We also examine basic evidence of how schools sort according to students classifications into GT, special education, or LEP categories. We run the following linear probability model:

$$Z_{ij} = \beta_1 + \sum_{j=2}^{J} \beta_j D_j + \varepsilon_{ij}, \tag{11}$$

where $z = GT, SP,$ and LEP and is a dummy variable indicating that classification. The right hand side of this equation is analogous to equation (9), where D_J is a dummy variable for class *j*.

It should be noted that schools may face constraints related to which teachers are certified to teach students who fall into these particular categories. For example, if a principal's strategy included dispersing GT students equally among all the classes within a grade, he would be forced to deviate from that strategy if some of the fourth grade teachers were not certified. Ideally, teacher characteristics would be included in the analysis to reveal potential sorting constraints. However, because the data allows linkage to a specific class but not to a teacher, this is not possible.

We run regressions (9) and (10) for each of the 135 schools in the district to determine which schools potentially sort by the various dimensions. Examples of the results for two particular schools are shown in tables 2A and 2B[5]. Consider, for example, the results for school 186, which are given in the first table. This school has four classes of fourth graders—two with lower average math scores and two with higher average math scores. The average score for class 1, given by the constant, is 26.1 (the maximum raw score is 42 points). The coefficient for class 2 is not significantly different from zero, and the point estimate is only 1.2 points, suggesting that there is no substantial score difference between the two classes. However, the estimates for class 3 and 4 are both statistically significant and indicate a 4.9 point and 6.1 point difference in score from class 1. Class 1 also has significantly lower reading scores than any of the other classes.

While classes in this school appear to be sorted by previous testing scores, they do not seem to be sorted along other dimensions. There is no significant difference in the number of gifted and talented students across the classes. About 11.8 percent of the students in class 1 are classified at GT; while the percents are higher in magnitude for the other classes, none of the differences is statistically significant. There is also not a significant difference between the number of special education students across the classes, as shown in column (5) of the table. It does appear, however, that the students are sorted by LEP status. Class 1 has significantly more LEP students than the other classes, particularly classes 2 and 4.

Table 2B shows an analogous example for school 235, a non-sorting school. There are no significant differences in the reading and math scores for any of the classes in this grade. The magnitude differences in the reading scores are particularly small. There is also no significant difference in the number of GT, special education, or LEP students across classes.

Table 3 shows a summary of the results from all of the schools. A school is tagged as a "sorting school" along the math or reading score dimension if the average score for any of the class dummies is statistically different from the others. Similarly, a school is considered a "sorting school" along the GT, special education, and LEP dimensions if any of the class dummies is statistically different from the others in equation (10). Of the 135 schools, almost three-fourths sort along at least one dimension. Almost 19 percent are math score sorters and

[5] Complete results from all schools are available upon request.

24 percent are reading score sorters. About 28 percent appear to sort by GT status, 57 percent sort by LEP status, and 13 percent sort by special education classification. Many schools appear to sort along multiple dimensions; almost 40 percent sort by at least two dimensions, and more than 20 percent sort by at least three.

About one-fourth of the schools in the dataset have no significant difference in any of the variables across different classes. It is important to note that even if score averages are not significantly different, schools may still be considering score in a strategic division of students into classes. Some schools may be purposefully allocating students of different abilities equally among classes. If administrators believe that an equal division of student ability is beneficial to some or all students, then there should be no significant score average score difference between classes, even if the school is acting strategically.

Table 4 presents a summary of schools that sort along at least two dimensions. Of the group of math sorting schools, 11 also sort by reading score, 15 also sort by GT status, 15 also sort by LEP status, and 5 sort by special education status. The diagonals of the table represent the total number of schools that sort along that particular dimension.

Table 5 presents results showing how observable characteristics predict whether or not schools sort, and along what dimensions. The dependent variable in each regression is equal to one if the school is classified as a sorting school. For example, schools with a higher variation in reading scores are more likely to sort by previous reading score. Schools with more gifted and talented and special education students are more likely to sort along those dimensions. Many of the coefficients are insignificant, and there appears to be a substantial amount of sorting present in schools that is unrelated to school or student characteristics.

In addition to comparing average scores and characteristics across classes, we also examine how two individual students' characteristics affect the probability that they will be in the same class. Consider the following regression for two students *i* and *j*:

$$same_{ij} = \Delta_{ij}\psi + \varepsilon_{ij}\,, \tag{12}$$

where $same_{ij}$ is equal to one if students *i* and *j* are in the same class. The vector Δ_{ij} includes various measures of differences between the students, including difference in math score, reading score, GT status, LEP status, and special education status. Negative values of ψ indicate more homogeneous classes, because higher differences in scores (or other characteristics) would decrease the likelihood of being in the same class. Positive values of ψ indicate more heterogeneous classes, because higher differences in scores would increase the likelihood that the two students are in the same class.

In order to run equation (11) we first construct all potential pairs of students within a grade in a given school. The regressions are evaluated separately for each school. An example is presented in table 6. This school appears to sort homogeneously. Students with larger differences in scores—both math and reading—are less likely to be placed in the same class. This is also true for LEP, GT, and special education status. None of the other coefficients are significantly different from each other. For example, the female coefficient is negative, but very small in magnitude and insignificant, indicating that classes do not appear to be sorted by gender. Table 7 summarizes the regressions from each school. The results are qualitatively consistent with the mean comparisons across classrooms. Difference in LEP status is the most commonly significant coefficient, confirming that many schools tend to divide students along this dimension.

5.2 Effect of Sorting by Test Score

It is not immediately clear whether sorting students will be beneficial for them or which types of sorting will be most beneficial for different types of students. As described earlier, an intuitive argument can be made for the benefits of tracking students into homogenous classes, as well as for evenly dividing them into heterogeneous classes. To explore this issue empirically, we create a sorting index for each class within a school measuring how dispersed its students are when compared to the overall school population at a single grade level. Following the formulas described in equations (5) and (6), we construct two indices: α_{1k}, which measures overall score dispersion within a grade in school *k*, and α_{2k}, which is a measure of score dispersion within the classes of school *k*. As described in section 3, the variable $sort_k = \frac{\alpha_{1k}}{\alpha_{2k}}$ reflects how “sorted” the classes of school *k* are, relative to the overall score dispersion in the school. Higher levels of $sort_k$ indicate classes that are more homogeneously sorted, and lower levels of $sort_k$ indicate that classes are more heterogeneous. we construct this variable separately for math scores and reading scores.

The baseline results for math scores are found in column (1) of table 9A. The dependent variable is a student’s 2005 math TAKS score, measured as a z-score. The results suggest a positive, statistically significant relationship between the sorting variable and math score, indicating that, sorting is beneficial to students. This suggests that, on average, students gain by being placed into a more homogeneous classroom, compared to one with more dispersion in testing scores.

Column (2) presents the same regression excluding the tails of the "score gain" distribution as a robustness check[6]. Students who showed extreme gains or extreme losses from 2004 to 2005 (moving from a raw score of 1 to 42, for example) are removed from the sample in this specification, because a change this large raises concerns about the validity of one or both of the scores. With these observations removed, the sorting effect is slightly larger in magnitude, but still strongly significant.

We also run the same regression with a student's gain in math score on the left hand side, rather than the level 2005 math score. The results, given in table 9B, are similar to the level results. The sorting coefficients are positive and significant in both specifications, suggesting that more homogeneous classes produce an increase in math scores for the average student.

We examine similar effects for reading scores. A summary of the math and reading results is presented in columns (1) and (2) of table 10. Note that each value represents the coefficient on the sorting index from a separate regression. All regressions include the same controls listed in tables 9A and 9B. The results for sorting by reading scores are similar, although smaller in magnitude, compared with the effects on math score.

As explained in section 3.4, there are potential endogeneity problems inherent in estimating the effect of sorting on student performance. Schools choose how to divide students into classes, and it is likely that they make this determination using variables that are unobserved to the researcher. Unobservable characteristics such as behavior may affect both schools' sorting decisions and student performance, causing the sorting coefficient to be biased. To correct for this, we create $sort_k = \frac{\alpha_{1k}}{\alpha_{2k}}$ for the fifth grade in every school to be used as an instrument for the fourth grade sorting index. The two indices should be correlated if schools' sorting guidelines are similar across grades and administrators use common mechanisms within a school to divide students into classes. The instrument exogeneity condition requires that there be no effect of the fifth grade's sorting index on the scores of the fourth grade

[6] The top and bottom 2.5 percent of the distribution is excluded from this regression.

students. This is intuitively plausible; there is no reason to believe that the way fifth grade students are divided into classes would have any impact on the academic performance of the fourth grade students in their school.

Columns (3) and (4) of table 10 show the results of the 2SLS estimations for math score and reading score, respectively. The estimates are positive and significant in for both math and reading scores, and generally larger in magnitude than the OLS estimates, suggesting there was a downward bias in the original results. These results hold across various specifications—for both level scores and score gains, and when the tails of the score gain distribution are excluded. These estimates confirm that, on average, more homogeneous classes are beneficial for students in increasing both math and reading achievement.

5.3 Allowing for Heterogeneity in the Sorting Effect

One of the concerns within the literature regarding sorting is that it may benefit one group of students at the cost of hurting another group. For example, several researchers[7] have suggested that while sorting may raise achievement levels for high achieving students, it actually lowers the performance of low achieving students. To test for this possibility, we rank students according to their previous year testing score, create dummy variables for high and low scoring students, and allow the sorting effect to vary across the two groups.

Table 11 presents the results allowing for heterogeneity in the sorting effect by student ability. The first rows report the coefficients for math score. While the results suggest slightly larger effects for high scoring students, there are still large, positive, and significant results for the low scoring group. The estimates for the two groups are not significantly different from each other. This suggests that it is not the case that sorting causes high ability students to gain at the cost of low achievers; on the contrary, both groups of students benefit from more homogeneous classes. These results are consistent with the conclusions of Duflo et al. (2008) and give credence to the line of reasoning that suggests that more homogeneous classes allow teachers to teach to a more narrow range of students, which is beneficial for both high and low scoring individuals. The results for reading score are similar. While the coefficients overall are still slightly smaller than the ones for the math results, all 2SLS specifications indicate a positive significant effect of sorting on reading score.

[7] See Hoffer (1992) and Betts and Shkolnik (2000).

5.4 Other Types of Sorting

In addition to examining how schools sort in regards to previous testing score, we also examine sorting along several other dimensions—gifted and talented status, special education status, and LEP status. The estimates for gifted and talented sorting are presented in tables 12-13. Overall, the results indicate effects that are positive but not statistically significant in the 2SLS regressions. Allowing for heterogeneity between GT students and non-GT students reveals that most of the positive results are being generated for non-GT students, although the point estimates are still not significant.

Results for special education students, given in tables 14-15, suggest negative (but not significant) sorting effects for math and negative, significant effects for reading. The results that allow for heterogeneity in the effect, as presented in table 15, suggest negative effects for non-special education students and positive (but not significant) effects for special education students.

Regressions for limited English proficiency students are reported in tables 16-17. While the 2SLS estimates are generally not estimated precisely, most of the specifications indicate that the sorting coefficient is positive, suggesting that more homogeneous classes are useful in increasing performance.

5. Conclusion

The purpose of this study is to examine how schools sort students into classes, how those sorting mechanisms affect student achievement, and whether there are heterogeneous sorting effects across a distribution of students. Using detailed student-level data that allows a student to be linked to his classroom, we find evidence of a wide variation in sorting practices across schools. Many schools appear to sort along various dimensions, including previous math and reading scores, gifted and talented or special education status, and limited English proficiency.

We find strong evidence that sorting students into more homogeneous groups is beneficial, particularly for sorting by previous testing score. Interestingly, when allowing for heterogeneity in the sorting effect across a distribution of students, we find positive and significant results for both high scoring and low scoring students, suggesting that *both* groups benefit from sorted classes. This is consistent with the hypothesis that dividing students into more homogeneous groups allows teachers to direct their focus to a more narrow range of students and meet the needs of their particular classroom more efficiently.

This study has valuable policy implications because unlike many school policy variables, the composition of classes can often be changed with little need for increased funds. A school with a fixed number of classrooms and teachers can increase efficiency by rearranging students in the most effective way possible. This study suggests that creating classes with lower levels of dispersion of score or ability level may improve the achievement outcomes for students across the score distribution.

REFERENCES

Alexander, Karl L., Entwisle, Doris R., Dauber, Susan L., 1996. Children in Motion: School Transfers and Elementary School Performance. Journal of Educational Research, 90 (1), 3-12.

Angrist, Joshua D., Lavy, Victor, 1999. Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. The Quarterly Journal of Economics 114 (2), 533-575.

Argys, Laura M., Rees, Daniel I., Brewer, Dominic J., 1996. Detracking America's Schools: Equity at Zero Cost?" Journal of Policy Analysis and Management 15 (4), 623-645.

Betts, Julian R., Shkolnik, Jamie, 2000. The effects of ability grouping on student achievement and resource allocation in secondary schools. Economics of Education Review 19 (1), 1-15.

Brown, Byron W., Saks, Daniel H., 1975. The Production and Distribution of Cognitive Skills within Schools. The Journal of Political Economy 83 (3), 571-594.

Coleman, James S. et al., 1966. Equality of Educational Opportunity (EEOS). US Government Printing Office, Washington, D.C.

Cullen, Julie B., Jacob, Brian A., Levitt, Steven D., 2005. The impact of school choice on student outcomes: an analysis of the Chicago Public Schools. Journal of Public Economics 89 (5-6), 729-760.

Cullen, Julie, Reback, Randall, 2006. Tinkering toward accolades: school gaming under a performance accountability system. In: Gronberg, Timothy J., Jansen, Dennis W. (Eds.), Advances in Applied Microeconomics. Improving School Accountability, vol. 14. Elsevier Science, Amsterdam.

Deere, Donald, Strayer, Wayne, 2002. Competitive Incentives: School Accountability and Student Outcomes in Texas. Working Paper.

Ding, Weili, Lehrer, Steven, 2007. Do peers affect student achievement in China's secondary schools? The Review of Economics and Statistics 89 (2), 300-312.

Dubin, Jeffrey A., McFadden, Daniel L., 1984. An Econometric Analysis of Residential Electric Appliance Holdings and Consumption. Econometrica 52 (2), 345-362.

Duflo, Esther, Dupas, Pascaline, Kremer, Michael. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. American Economic Review 101 (5).

Eide, Eric, Showalter, Mark H., 1998. The effect of school quality on student performance: A quantile regression approach. Economic Letters 58 (3), 345-350.

Ehrenberg, Ronald G., Brewer, Dominic J., 1994. Do School and Teacher Characteristics Matter? Evidence from High School and Beyond. Economics of Education Review 13 (1),1-17.

Figlio, David, Getzler, Lawrence S., 2006. Accountability, Ability, and Disability: Gaming the System? In: Gronberg, Timothy J., Jansen, Dennis W. (Eds.), Advances in Applied Microeconomics. Improving School Accountability, vol. 14. Elsevier Science, Amsterdam.

Figlio, David, Page, Marianne, 2002. School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality? Journal of Urban Economics 51, 497-514.

Hanushek, Eric A., 1986. The Economics of Schooling: Production and Efficiency in Public Schools. Journal of Economic Literature 24 (3), 1141-1177.

Hanushek, Eric A., 1997. Assessing the Effects of School Resources on Student Performance: An Update. Education Evaluation and Policy Analysis, 19 (2), 141-164.

Hanushek, Eric A., Kain, John F., Rivkin, Steven G., 2004. Disruption versus Tiebout improvement: the costs and benefits of switching schools. Journal of Public Economics 88 (9-10), 1721-1746.

Hoffer, Thomas B., 1992. Middle school ability grouping and student achievement in science and mathematics. Educational Evaluation and Policy Analysis 14 (3), 205-227.

Hoxby, Caroline M., 2000. The Effects of Class Size on Student Achievement: New Evidence from Population Variation. The Quarterly Journal of Economics 115 (4), 1239-1285.

Ingersoll, Gary M., Scamman, James P., Eckerling, Wayne D., 1989. Geographic mobility and student achievement in an urban setting. Educational Evaluation and Policy Analysis 11 (2), 143-149.

Jacob, Brian A., 2005. Accountability, incentives, and behavior: the impact of high-stakes testing in the Chicago Public Schools. Journal of Public Economics 89 (5-6), 761-796.

Jacob, Brian A., Levitt, Steven, 2003. Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. Quarterly Journal of Economics 118 (3), 843-878.

Kerbow, David, 1996. Patterns of Urban Student Mobility and Local School Reform. Journal of Education for Students Placed at Risk 1 (2), 147-169.

Kain, John F., O'Brien, Daniel M., 1999. A Longitudinal Assessment of Reading Achievement: Evidence from the Harvard/UTD Texas Schools Project. University of Texas at Dallas, UTD Texas Schools Project.

Krueger, Alan B., 2003. Economic Considerations and Class Size. The Economic Journal 113 (485), F34-F63.

Ladd, Helen F., Walsh, Randall, 2002. Implementing value-added measures of school effectiveness: getting the incentives right. Economics of Education Review 21(1), 1-17.

Medina, Jennifer, 2003. New York measuring teachers by test scores. New York Times (January 21, 2008).

Mishel, Lawrence, Rothstein, Richard (Eds.), 2002. The Class Size Debate. Economic Policy Institute, Washington, D.C.

Reback, Randall, 2008. Teaching to the rating: school accountability and the distribution of student achievement. Journal of Public Economics, 92 (5-6), 1394-1415.

Rees, Daniel I., Argys, Laura M., Brewer, Dominic J., 1996. Tracking in the United States: Descriptive Statistics from the NELS. Economics of Education Review 15 (1), 83-89.

Rees, Daniel I., Argys, Laura M., Brewer, Dominic J., 2000. How should we measure the effect of ability grouping on performance? Economics of Education Review 19 (1), 17-20.

Rivkin, Steven G., Hanushek, Eric A., Kain, John F., 2005. Teachers, Schools, and Academic Achievement. Econometrica 73 (2), 417-458.

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

Todd, Petra E., Wolpin Kenneth we., 2007. The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. Journal of Human Capital 1 (1), 91-136.

Word, Elizabeth, Johnston, John, Bain, Helen Pate et al., 1990. The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project Technical Report 1985-1990. Tennessee Department of Education, Nashville, TN.

Table 1: Summary Statistics (4th graders)

student characteristics					
math scale score	10751	2182.7660	193.4539	1280	2684
reading scale score	10534	2159.7990	175.7893	1319	2614
math scale score 2004	9633	2213.1780	180.0879	1228	2697
reading scale score 2004	9430	2226.6030	176.3924	1356	2588
female	12015	0.4807	0.4996	0	1
lunch	12236	0.8634	0.3435	0	1
black	12015	0.2925	0.4549	0	1
hispanic	12015	0.6382	0.4805	0	1
asian	12015	0.0113	0.1058	0	1
indian	12015	0.0027	0.0515	0	1
GT	12236	0.1993	0.3995	0	1
special education	12236	0.0991	0.2987	0	1
LEP	12236	0.2118	0.4086	0	1
school characteristics					
enrollment (grade level)	138	88.66667	38.4298	9	181
number of classes	138	4.695652	1.95425	1	12
class size	138	18.8913	3.59579	5	27
teacher experience	138	11.47863	2.98164	4	18.656
teacher salary	138	46980.52	2459.14	41879	54072
sort (by math score)	138	1.048918	0.07335	0.97109	1.5636
sort (by reading score)	138	1.067374	0.10903	0.95884	1.7697
sort (by GT status)	136	1.045837	0.08011	0.98858	1.6879
sort (by special ed status)	136	1.042658	0.08709	0.92489	1.8003
sort (by LEP status)	132	1.137354	0.17592	0.920	1.9837

Table 2A: Mean Characterisitcs by Classes (Sorting School Example)

	(1)	(2)	(3)	(4)	(5)
VARIABLES	math score	reading score	gifted/talented	LEP	special educ
class 2	1.233	4.027*	0.149	-0.404***	0.0235
	(2.682)	(2.328)	(0.147)	(0.106)	(0.120)
class 3	4.900*	6.157***	0.0824	-0.471***	-0.0431
	(2.593)	(2.206)	(0.147)	(0.106)	(0.120)
class 4	6.054**	6.633***	0.149	-0.471***	-0.176
	(2.634)	(2.281)	(0.147)	(0.106)	(0.120)
constant (class 1)	26.10***	25.70***	0.118	0.471***	0.176**
	(1.981)	(1.685)	(0.101)	(0.0724)	(0.0818)
classification	sorted	sorted	not sorted	sorted	not sorted
Observations	49	47	62	62	62
R-squared	0.143	0.196	0.024	0.328	0.052

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2B: Mean Characterisitcs by Classes (Non-Sorting School Example)

	(1)	(2)	(3)	(4)	(5)
VARIABLES	math score	reading score	gifted/talented	LEP	special educ
class 2	3.167	1.170	0.0870	0.0435	-0.0870
	(2.457)	(2.175)	(0.138)	(0.0355)	(0.0772)
class 2	0.881	0.0110	0.0435	0	-0.0870
	(2.541)	(2.289)	(0.138)	(0.0355)	(0.0772)
constant (class 1)	23.33***	27.14***	0.261***	-0	0.130**
	(1.765)	(1.588)	(0.0978)	(0.0251)	(0.0546)
classification	not sorted	not sorted	not sorted	not sorted	not sorted
Observations	45	43	69	69	69
R-squared	0.041	0.01	0.006	0.029	0.025

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Summary of Sorting Schools

sorting type	number of schools	percent
math score	25	0.185
reading score	33	0.244
gifted/talented	38	0.281
LEP	77	0.570
special educ	17	0.126
<u>Number of sorting dimensions</u>		
At least one	100	0.741
At least two	53	0.393
At least three	28	0.207
At least four	9	0.067
At least five	0	0.000
No sorting	35	0.259

Table 4: Summary of schools by sorting type

	math score	reading score	gifted/talented	LEP	special educ
math score	**25**	11	15	15	5
reading score	11	**33**	19	25	3
gifted/talented	15	19	**38**	28	5
LEP	15	25	28	**77**	10
special educ	5	3	5	10	**17**

<u>Notes</u>: Includes only schools that sort by at least one type. Table shows number of schools that sort by an additional type, given the original sorting type.

Table 5: Characteristics of Sorting Schools (dependent variable=1 if school is a sorting school)

VARIABLES	(1) Math Score	(2) Reading Score	(3) Gifted	(4) Special Educ	(5) LEP
average math score	-0.0005	-0.0003	0.0002	-0.0001	0.0001
	(0.0009)	(0.0010)	(0.0010)	(0.0008)	(0.0012)
std dev of math score	0.0020	-0.0018	0.0020	0.0028*	-0.0026
	(0.0018)	(0.0021)	(0.0022)	(0.0016)	(0.0025)
average reading score	0.0007	0.0007	-0.0013	-0.0006	0.0003
	(0.0011)	(0.0013)	(0.0013)	(0.0009)	(0.0015)
std dev of reading score	0.0005	0.0047**	0.0010	-0.0042**	0.0045*
	(0.0019)	(0.0022)	(0.0023)	(0.0017)	(0.0026)
special educ	0.8945	0.0244	0.9884	1.7203**	0.8963
	(0.8065)	(0.9394)	(0.9812)	(0.7086)	(1.1080)
gifted	0.0733	-0.0042	1.0858**	-0.2075	-0.4547
	(0.3903)	(0.4546)	(0.4749)	(0.3429)	(0.5362)
LEP	-0.1281	0.1623	0.5557*	0.1859	0.3291
	(0.2706)	(0.3152)	(0.3292)	(0.2378)	(0.3718)
free lunch	-0.5088	0.1594	-0.8995	0.2743	0.0205
	(0.4558)	(0.5309)	(0.5545)	(0.4005)	(0.6262)
female	1.1976**	0.7126	0.6336	-0.1891	-0.6947
	(0.5552)	(0.6467)	(0.6754)	(0.4878)	(0.7627)
enroll	0.0020	0.0026	0.0035	-0.0013	0.0014
	(0.0019)	(0.0023)	(0.0024)	(0.0017)	(0.0027)
number of classes	0.0093	-0.0227	-0.0622	0.0272	-0.0281
	(0.0351)	(0.0409)	(0.0428)	(0.0309)	(0.0483)
black	0.7042	-0.3101	1.3623**	-0.6916	-0.2151
	(0.4770)	(0.5556)	(0.5803)	(0.4191)	(0.6553)
hispanic	0.5803	-0.0272	1.2920**	-0.8543**	-0.3679
	(0.4899)	(0.5706)	(0.5960)	(0.4304)	(0.6731)
teacher experience	0.0672	0.0108	0.0617	0.0199	0.0090
	(0.0437)	(0.0510)	(0.0532)	(0.0384)	(0.0601)
average salary	-0.1176**	0.0101	-0.0626	-0.0229	-0.0369
	(0.0523)	(0.0609)	(0.0636)	(0.0459)	(0.0718)
constant	2.8995	-2.0773	3.2939	3.1598	1.5381
	(2.8664)	(3.3389)	(3.4874)	(2.5184)	(3.9381)
Observations	131	131	131	131	131
R-squared	0.2138	0.1164	0.1408	0.1959	0.0966

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Pairwise Comparisons (Example)

	Dependent variable = 1 if students are in the same class
Δ math score	-0.0659**
	(0.0301)
Δ reading score	-0.141***
	(0.0338)
Δ LEP	-0.138***
	(0.0468)
Δ gifted	-0.152***
	(0.0471)
Δ special educ	0.261***
	(0.0984)
Δ black	-0.0367
	(0.0515)
Δ hispanic	0.0391
	(0.0518)
Δ free lunch	-0.0358
	(0.0477)
Δ female	-0.00342
	(0.0455)
constant	-0.628***
	(0.0597)
Observations	4,305

Notes: An observation is a pair of students within a grade
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Summary of Pairwise Comparison Coefficients

VARIABLES	Negative	Positive
Δ math score	19	6
Δ reading score	25	11
Δ LEP	44	7
Δ gifted	16	11
Δ special educ	18	21
Δ black	25	10
Δ hispanic	22	5
Δ free lunch	12	12
Δ female	6	5

Table 8: Correlation of Sorting Indices Across Grades

sort (math score)	0.5134
sort (reading score)	0.5041
sort (gifted)	0.3729
sort (special ed)	0.5655
sort (LEP)	0.4054

Note: Measures correlation between sorting indices for fourth grade and fifth grade. One observation per school.

Table 9A: Effect of Sorting by Previous Year Test Score
(No Heterogeneity Across Student Groups)

	Dependent Variable: Math Score			
Variables	(1) OLS	(2) OLS	(3) 2SLS	(4) 2SLS
sort (by score)	0.3613***	0.2883***	0.6447***	0.5257***
	(0.0989)	(0.0873)	(0.1632)	(0.1442)
math score 2004	0.4843***	0.6022***	0.4846***	0.6015***
	(0.0101)	(0.0093)	(0.0101)	(0.0093)
reading score 2004	0.1982***	0.1484***	0.1989***	0.1496***
	(0.0103)	(0.0091)	(0.0103)	(0.0092)
female	-0.0224	-0.0105	-0.0194	-0.0115
	(0.0146)	(0.0129)	(0.0147)	(0.0130)
lunch	-0.0489**	-0.0532***	-0.0468**	-0.0516***
	(0.0224)	(0.0198)	(0.0226)	(0.0200)
black	-0.3973***	-0.2847***	-0.4031***	-0.2899***
	(0.0377)	(0.0335)	(0.0380)	(0.0338)
hispanic	-0.1793***	-0.1272***	-0.1829***	-0.1281***
	(0.0367)	(0.0325)	(0.0371)	(0.0328)
asian	0.0786	0.0541	0.0861	0.0591
	(0.0753)	(0.0671)	(0.0768)	(0.0686)
indian	-0.0401	-0.0043	-0.0343	0.0027
	(0.1513)	(0.1362)	(0.1512)	(0.1361)
GT	0.3746***	0.2774***	0.3733***	0.2765***
	(0.0190)	(0.0170)	(0.0192)	(0.0171)
LEP	-0.0879***	-0.0768***	-0.0822***	-0.0736***
	(0.0235)	(0.0207)	(0.0237)	(0.0208)
special education	0.0041	-0.0003	0.0096	0.0039
	(0.0426)	(0.0373)	(0.0428)	(0.0375)
perc exper (0 years)	0.0081***	0.0070**	0.0088***	0.0073***
	(0.0031)	(0.0028)	(0.0031)	(0.0028)
perc exper (1-5 years)	0.0063**	0.0049*	0.0069**	0.0053**
	(0.0029)	(0.0026)	(0.0030)	(0.0026)
perc exper (6-10 years)	0.0034	0.0009	0.0030	0.0005
	(0.0027)	(0.0024)	(0.0027)	(0.0024)
perc exper (11-20 years)	0.0050**	0.0040**	0.0052**	0.0040**
	(0.0021)	(0.0018)	(0.0021)	(0.0019)
average salary	0.0215*	0.0147	0.0242*	0.0164
	(0.0128)	(0.0113)	(0.0129)	(0.0114)
Tails Excluded	NO	YES	NO	YES
Observations	9,093	8,589	8,949	8,455
R-squared	0.5142	0.6065	0.5148	0.6073
Cragg-Donald F stat			5187	4881

Notes: All regressions also include controls for average math and reading scores by school, enrollment and enrollment squared, class size, and a constant.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9B: Effect of Sorting by Previous Year Test Score
(No Heterogeneity Across Student Groups)

	Dependent Variable: Math Score Gain			
Variables	(1) OLS	(2) OLS	(3) 2SLS	(4) 2SLS
sort (by score)	0.4009***	0.3158***	0.6613***	0.5276***
	(0.1123)	(0.0963)	(0.1853)	(0.1591)
math score 2004				
reading score 2004	-0.0448***	-0.0349***	-0.0453***	-0.0350***
	(0.0103)	(0.0089)	(0.0104)	(0.0089)
female	0.0520***	0.0437***	0.0541***	0.0414***
	(0.0166)	(0.0142)	(0.0167)	(0.0143)
lunch	-0.0526**	-0.0502**	-0.0519**	-0.0500**
	(0.0255)	(0.0218)	(0.0257)	(0.0220)
black	-0.2994***	-0.1919***	-0.3072***	-0.1982***
	(0.0428)	(0.0368)	(0.0431)	(0.0372)
hispanic	-0.1639***	-0.1055***	-0.1689***	-0.1071***
	(0.0417)	(0.0359)	(0.0421)	(0.0362)
asian	-0.0322	-0.0379	-0.0209	-0.0304
	(0.0855)	(0.0739)	(0.0872)	(0.0756)
indian	-0.0862	-0.0143	-0.0825	-0.0090
	(0.1718)	(0.1502)	(0.1717)	(0.1501)
GT	0.1478***	0.0964***	0.1477***	0.0954***
	(0.0210)	(0.0181)	(0.0212)	(0.0183)
LEP	-0.0546**	-0.0526**	-0.0495*	-0.0503**
	(0.0267)	(0.0228)	(0.0269)	(0.0230)
special education	0.0481	0.0311	0.0530	0.0346
	(0.0483)	(0.0411)	(0.0486)	(0.0414)
perc exper (0 years)	0.0080**	0.0066**	0.0085**	0.0068**
	(0.0035)	(0.0030)	(0.0036)	(0.0031)
perc exper (1-5 years)	0.0057*	0.0043	0.0062*	0.0045
	(0.0033)	(0.0028)	(0.0034)	(0.0029)
perc exper (6-10 years)	0.0020	-0.0005	0.0015	-0.0010
	(0.0031)	(0.0027)	(0.0031)	(0.0027)
perc exper (11-20 years)	0.0052**	0.0041**	0.0053**	0.0041**
	(0.0024)	(0.0020)	(0.0024)	(0.0021)
average salary	0.0195	0.0118	0.0217	0.0130
	(0.0145)	(0.0125)	(0.0146)	(0.0126)
Tails Excluded	NO	YES	NO	YES
Observations	9,093	8,589	8,949	8,455
R-squared	0.0374	0.0306	0.0382	0.0309
Cragg-Donald F stat			5188	4881

Notes: All regressions also include controls for average math and reading scores by school, enrollment and enrollment squared, class size, and a constant.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 10: Effect of Sorting by Previous Year Test Score
(No Heterogeneity Across Student Groups)

Dependent Variable	(1) OLS Tails Included	(2) OLS Tails Excluded	(3) 2SLS Tails Included	(4) 2SLS Tails Excluded
Math Score	0.3613***	0.2883***	0.6447***	0.5257***
	(0.0989)	(0.0873)	(0.1632)	(0.1442)
Math Score Gain	0.4009***	0.3158***	0.6613***	0.5276***
	(0.1123)	(0.0963)	(0.1853)	(0.1591)
Reading Score	0.2383***	0.1342**	0.4387***	0.3454***
	(0.0754)	(0.0662)	(0.1201)	(0.1056)
Reading Score Gain	0.3246***	0.1880**	0.5725***	0.4220***
	(0.0852)	(0.0733)	(0.1357)	(0.1168)

Notes: Each value is a coefficient from a separate regression of score on the sorting index. Included in all regresssions are all student and school controls reported in the previous table.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 11: Effect of Sorting by Previous Year Test Score
(Heterogeneity Across Student Groups)

Dependent Variable		(1) OLS Tails Included	(2) OLS Tails Excluded	(3) 2SLS Tails Included	(4) 2SLS Tails Excluded
Math Score	Sort*High	0.4410***	0.3223***	0.7098***	0.5535***
		(0.0991)	(0.0878)	(0.1631)	(0.1445)
	Sort*Low	0.2775***	0.2550***	0.5505***	0.4881***
		(0.0992)	(0.0878)	(0.1630)	(0.1444)
Math Score Gain	Sort*High	0.1626	0.1147	0.4658***	0.3653**
		(0.1089)	(0.0934)	(0.1793)	(0.1541)
	Sort*Low	0.6220***	0.4915***	0.9294***	0.7451***
		(0.1088)	(0.0934)	(0.1791)	(0.1539)
Reading Score	Sort*High	0.3030***	0.1778***	0.4909***	0.3802***
		(0.0759)	(0.0668)	(0.1201)	(0.1058)
	Sort*Low	0.1548**	0.0818	0.3502***	0.2865***
		(0.0763)	(0.0671)	(0.1207)	(0.1062)
Reading Score Gain	Sort*High	0.0712	-0.0126	0.3290**	0.2427**
		(0.0823)	(0.0711)	(0.1305)	(0.1130)
	Sort*Low	0.5677***	0.3786***	0.8313***	0.6369***
		(0.0822)	(0.0711)	(0.1304)	(0.1128)

Notes: Each value is a coefficient from a separate regression of score on the sorting index.
Included in all regresssions are all student and school controls reported in the previous table.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 12: Effect of Sorting by Gifted and Talented Status
(No Heterogeneity Across Student Groups)

Dependent Variable	(1) OLS Tails Included	(2) OLS Tails Excluded	(3) 2SLS Tails Included	(4) 2SLS Tails Excluded
Math Score	0.1438*	0.0940	0.0495	0.0944
	(0.0848)	(0.0746)	(0.1805)	(0.1568)
Math Score Gain	0.2197**	0.1434*	0.0368	0.1215
	(0.0961)	(0.0822)	(0.2047)	(0.1728)
Reading Score	0.1436*	0.0270	-0.0290	-0.0806
	(0.0859)	(0.0755)	(0.1830)	(0.1615)
Reading Score Gain	0.3011***	0.1360	0.1837	0.0624
	(0.0971)	(0.0835)	(0.2066)	(0.1784)

Notes: Each value is a coefficient from a separate regression of score on the sorting index. Included in all regresssions are all student and school controls reported in the previous table.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 13: Effect of Sorting by Gifted and Talented Status
(Heterogeneity Across Student Groups)

Dependent Variable		(1) OLS Tails Included	(2) OLS Tails Excluded	(3) 2SLS Tails Included	(4) 2SLS Tails Excluded
Math Score	Sort*GT	0.0885	-0.0088	-0.0901	-0.0686
		(0.1266)	(0.1129)	(0.2428)	(0.2163)
	Sort*Non-GT	0.1863*	0.1693*	0.1440	0.1988
		(0.1113)	(0.0970)	(0.2438)	(0.2076)
Math Score Gain	Sort*GT	-0.00592	-0.0942	-0.1551	-0.0666
		(0.144)	(0.124)	(0.2753)	(0.2383)
	Sort*Non-GT	0.392***	0.317***	0.1667	0.2419
		(0.126)	(0.107)	(0.2764)	(0.2286)
Reading Score	Sort*GT	0.1733	0.0817	-0.0386	-0.0105
		(0.1281)	(0.1137)	(0.2457)	(0.2218)
	Sort*Non-GT	0.1206	-0.0136	-0.0224	-0.1271
		(0.1131)	(0.0984)	(0.2476)	(0.2151)
Reading Score Gain	Sort*GT	0.2230	0.1107	0.0393	0.0663
		(0.1448)	(0.1258)	(0.2778)	(0.2454)
	Sort*Non-GT	0.3614***	0.1548	0.2822	0.0599
		(0.1277)	(0.1088)	(0.2794)	(0.2376)

Notes: Each value is a coefficient from a separate regression of score on the sorting index.
Included in all regresssions are all student and school controls reported in the previous table.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14: Effect of Sorting by Special Education Status
(No Heterogeneity Across Student Groups)

Dependent Variable	(1) OLS Tails Included	(2) OLS Tails Excluded	(3) 2SLS Tails Included	(4) 2SLS Tails Excluded
Math Score	-0.0090	0.0664	-0.0228	0.0333
	(0.1030)	(0.0902)	(0.1632)	(0.1413)
Math Score Gain	-0.0476	0.0494	-0.0096	0.0768
	(0.117)	(0.0995)	(0.1852)	(0.1558)
Reading Score	-0.1841*	-0.2200**	-0.3132*	-0.2849*
	(0.1049)	(0.0926)	(0.1663)	(0.1471)
Reading Score Gain	-0.0965	-0.1440	-0.1715	-0.1536
	(0.1186)	(0.1025)	(0.1881)	(0.1627)

<u>Notes</u>: Each value is a coefficient from a separate regression of score on the sorting index. Included in all regresssions are all student and school controls reported in the previous table.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 15: Effect of Sorting by Special Education Status
(Heterogeneity Across Student Groups)

Dependent Variable		(1) OLS Tails Included	(2) OLS Tails Excluded	(3) 2SLS Tails Included	(4) 2SLS Tails Excluded
Math Score	Sort*SpEd	0.1951	0.3303	0.3182	0.2127
		(0.8678)	(0.7493)	(1.7150)	(1.4523)
	Sort*Non-SpEd	-0.0111	0.0636	-0.0262	0.0315
		(0.1034)	(0.0906)	(0.1630)	(0.1412)
Math Score Gain	Sort*SpEd	0.0441	0.303	0.6283	0.3434
		(0.986)	(0.826)	(1.9460)	(1.6005)
	Sort*Non-SpEd	-0.0486	0.0467	-0.0159	0.0740
		(0.117)	(0.0999)	(0.1850)	(0.1556)
Reading Score	Sort*SpEd	-0.1710	-0.3591	1.6248	1.6000
		(0.9128)	(0.8048)	(1.7013)	(1.5295)
	Sort*Non-SpEd	-0.1842*	-0.2187**	-0.3312**	-0.3017**
		(0.1053)	(0.0929)	(0.1664)	(0.1471)
Reading Score Gain	Sort*SpEd	0.4408	-0.0847	2.5336	2.2571
		(1.0326)	(0.8905)	(1.9245)	(1.6927)
	Sort*Non-SpEd	-0.1016	-0.1445	-0.1967	-0.1751
		(0.1191)	(0.1028)	(0.1882)	(0.1627)

Notes: Each value is a coefficient from a separate regression of score on the sorting index. Included in all regresssions are all student and school controls reported in the previous table.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 16: Effect of Sorting by LEP Status
(No Heterogeneity Across Student Groups)

Dependent Variable	(1) OLS Tails Included	(2) OLS Tails Excluded	(3) 2SLS Tails Included	(4) 2SLS Tails Excluded
Math Score	0.2211***	0.1746***	0.1838*	0.0730
	(0.0423)	(0.0372)	(0.1038)	(0.0927)
Math Score Gain	0.188***	0.144***	0.1145	-0.0042
	(0.0480)	(0.0410)	(0.1173)	(0.1019)
Reading Score	0.1166***	0.0535	0.0677	0.0033
	(0.0429)	(0.0378)	(0.1052)	(0.0930)
Reading Score Gain	0.1213**	0.0513	0.0834	0.0062
	(0.0485)	(0.0418)	(0.1191)	(0.1030)

Notes: Each value is a coefficient from a separate regression of score on the sorting index. Included in all regresssions are all student and school controls reported in the previous table.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 17: Effect of Sorting by LEP Status
(Heterogeneity Across Student Groups)

Dependent Variable		(1) OLS Tails Included	(2) OLS Tails Excluded	(3) 2SLS Tails Included	(4) 2SLS Tails Excluded
Math Score	Sort*LEP	0.1482	0.0811	0.2925	0.2019
		(0.1236)	(0.1084)	(0.2463)	(0.2202)
	Sort*Non-LEP	0.2305***	0.1868***	0.1647	0.0502
		(0.0449)	(0.0395)	(0.1129)	(0.1009)
Math Score Gain	Sort*LEP	0.188	0.0970	-0.0247	-0.0542
		(0.140)	(0.119)	(0.2786)	(0.2423)
	Sort*Non-LEP	0.188***	0.150***	0.1389	0.0046
		(0.0509)	(0.0436)	(0.1276)	(0.1109)
Reading Score	Sort*LEP	0.1023	0.0224	0.1799	0.2864
		(0.1266)	(0.1108)	(0.2508)	(0.2193)
	Sort*Non-LEP	0.1184***	0.0575	0.0481	-0.0483
		(0.0455)	(0.0400)	(0.1143)	(0.1013)
Reading Score Gain	Sort*LEP	0.2477*	0.1151	0.3011	0.4109*
		(0.1432)	(0.1227)	(0.2837)	(0.2427)
	Sort*Non-LEP	0.1054**	0.0431	0.0454	-0.0675
		(0.0514)	(0.0443)	(0.1294)	(0.1122)

Notes: Each value is a coefficient from a separate regression of score on the sorting index.
Included in all regresssions are all student and school controls reported in the previous table.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1