

*Department of Economics
Seminar Series*

*Liang Choon Wang
University of California, San Diego*

***“Peer Effects in the Classroom:
Evidence from a Natural Experiment
in Malaysia”***

*Friday,
January 29, 2010
3:30 p.m.
211 Middlebush*

**Peer Effects in the Classroom:
Evidence from a Natural Experiment in Malaysia***

Liang Choon Wang
University of California, San Diego

Job Market Paper
(January 2009)

Abstract

Studies of peer effects in schools often face identification problems that arise from the non-random selection of students into classrooms or schools. This paper exploits the quasi-random assignment of students into classrooms in a large secondary school in Malaysia to estimate the causal effects of peers on educational outcomes. Even with random assignment, peer effect estimates are vulnerable to a mechanical bias (Guryan et al., forthcoming). I show that existing treatments of this bias are inadequate and underestimate peer effects. I demonstrate a simple solution and find a large positive causal effect of peers on math achievement – a one standard deviation increase in the average baseline math score of classmates leads to a 0.50 standard deviation increase in a student’s current math score. Effort and behavior may account for part of this large peer effect: the presence of high achieving peers lowers absence rates and the incidence of discipline violations. Unlike previous findings, I do not find evidence that peer effects vary across different types of students.

* I thank Eli Berman, Julian Betts, Prashant Bharadwaj, Tiffany Chou, Gordon Dahl, Michael Ewens, David Figlio, Ben Gilbert, Ben Gillen, Nora Gordon, Larry Hedges, Jacob LaRiviere, Xun Lu, Craig McIntosh, Steve Raudenbush, Alex Sawyer, Yixiao Sun, and many seminar participants for helpful comments and discussions. I am grateful to the Institute for Applied Economics and the Department of Economics of the University of California San Diego for research funding, and the Spencer Foundation for fellowship support. I also thank the Malaysian secondary school for data provision and assistance that made this study possible. Any errors are the mistake of the author.

1. Introduction

Policy debates on the costs and benefits of school choice, ability grouping, and special education programs largely hinge on the existence and forms of peer effects in schools and classrooms. For example, opponents of school choice initiatives often argue that school choice may worsen the educational outcomes of students in schools from which better students have opted out. Parents opposed to ability grouping may also be concerned with the adverse effects of separating low achieving and high achieving students into different classrooms. These concerns are legitimate if peer effects are present so that reassigning high-achieving students away from low-achieving students can redistribute achievement gains. To the extent that test scores predict wages and incomes, the presence of peer effects implies that policy changes altering the ability mix of students in the same school or classroom can impact future wage and income distribution of a country (Bishop, 1989; Bound and Johnson, 1992; Gamoran and Mare, 1989). Quantifying the effects of peers on student outcomes may also provide insights into the spillover effects of any policy intervention. If positive peer effects exist, then a policy intervention directed towards some students can also benefit other students not directly targeted by the policy through interactions between students. Hence, the central questions to policy makers are: do peer effects exist, and if so, what are their sizes and do they vary across different types of students?

Studies of peer effects in schools face numerous identification and estimation problems. The first type of identification problem is what Manski (1993) referred to as the reflection problem and it occurs when researchers cannot properly isolate the effect of a person on her peers from the effect of her peers on her. This issue is essentially an endogeneity problem as outcomes of members in the group are jointly determined (Moffitt, 2001). Second, individuals in the same group may have similar behavior and outcomes because they self-select or are selected into the same group on the basis of factors such as motivation and ability, which are unobserved by researchers. In the presence of unobserved selection, group members may behave similarly because they share similar characteristics and not necessarily because they influence one another. Third, members of the same group may also have similar outcomes because they experience common shocks unrelated to their interactions. For instance, students in the same classroom may have similar achievement because they share the same teachers. Manski (1993) labeled effects due to unobserved selection and common group shocks as correlated effects. Finally, a mechanical negative correlation exists between an individual's own outcome and the average of

her peers' outcomes in a typical linear regression framework because they deviate in opposite directions from the group mean. With the exception of Guryan et al. (forthcoming), this problem has largely been neglected in the literature.¹ Failing to address these identification problems may lead to inconsistent estimates of the effect of peers.

Several studies attempt to address these identification problems. The first commonly employed strategy, especially for studies of peer effects in school, is to address unobserved selection and the reflection problem by using a fixed effects estimator and replacing current peer achievement with lagged peer achievement (Betts and Zau, 2004; Hanushek et al., 2003; Vigdor and Nechyba, 2007). Since variation across classrooms or cohorts within a school and lagged peer achievement may not be credibly exogenous, Hoxby and Weingarth (2007) used school-reassignment-induced variation in peers to implement instrumental variable estimation. In contrast, Burke and Sass (2008) adapted a fixed effects estimator proposed by Arcidiacono et al. (2005) and modeled individual outcomes as a function of peers' and teacher fixed effects to control for unobserved selection. Generally, the empirical findings on the size of peer effects on student outcomes are mixed. For example, Vigdor and Nechyba (2007) and Burke and Sass (2008) report either small or insignificant effects of peers on student achievements in schools, while Hanushek et al. (2003), Betts and Zau (2004), Hoxby and Weingarth (2007) report sizable positive effects of peers on student achievement. Moreover, estimated peer effects at the classroom level are typically larger than estimates at the grade level or school level.

Natural or controlled experiments are another means of identifying causal effects of peers. The random assignment of students into groups removes correlated effects, since group membership is, by construction, orthogonal to group characteristics. The majority of these studies look at student outcomes as a function of college roommates or classmates (Carrell et al. 2009; De Giorgi et al. forthcoming; Foster 2006; Lyle 2007; Sacerdote 2001; Zimmerman 2003). The effects of classmates are reported to be positive and significant, while the effects of roommates on achievement are relatively weak and insignificant. It is less common to see the random assignment of students into classrooms in K-12 school, with the exception of Kang (2007) and Duflo et al. (2008). Using contemporaneous achievement of a random sample of classmates, Kang (2007) finds strong positive effects of peers on math achievement in South

¹ Lyle (2007) explains this negative correlation in a footnote and De Giorgi et al. (2007) discuss this mechanical relationship in a different context, but they did not investigate its properties.

Korean middle schools. Similarly, Duflo et al. (2008) study the effects of tracking and peers for Kenyan first graders in a large scale controlled experiment and report strong positive effects of peers on math and literature achievement. Moreover, they find that students at all levels of the initial test score distribution benefited from tracking.

Using data generated from the quasi-random assignment of students into classrooms in a large secondary school in Malaysia, I investigate peer effects of a student's classmates on individual math achievement, class absences, and discipline violations in the first year of secondary school (grade 7). The sample school assigns students into classrooms using a computer program that attempts to equalize the average baseline test scores across classrooms. This quasi-random assignment method ensures that the estimated peer effects are unlikely confounded by systematic assignment of teachers of different quality into classrooms of different achievement mixes. Similarly, selection bias is effectively controlled because the school only has limited information about each student's past and the classroom assignment method is strictly enforced.² The data also provide a clear definition of the relevant peer group, since students in the sample attend *all* classes with the same set of classmates during an academic year. As most forms of academic interaction occur at the classroom level, the effects of classmates are likely strongest and the most relevant to policy makers. In contrast to previous studies, this paper's comprehensive coverage of classmates and baseline test scores provides a more complete picture of peer effects.³ Observing standardized baseline test scores is important, as I show that controlling for each student's own baseline test score is essential for removing the mechanical correlation between an individual's outcome and the average outcome of peers. Moreover, because the student body consists of racially homogenous urban students of the same age, potential differential effects of race and age of peers will not confound the estimated effects of peers. This feature is particularly important given recent findings by Angrist and Lang (2004) and Cooley (2008), which show that the effects of peers may differ depending on the racial composition of students in a classroom.

I also examine the mechanical negative relationship between characteristics of individuals and peers and the test of random assignment commonly used in other studies of peer

² This contrasts practices in other countries where parents may request reassignment.

³ For example, Kang (2007) only observed a sample of an individual's classmates and did not have baseline test scores available. Duflo et al. (2008) observed baseline test scores, but these test scores are not comparable across schools.

effects. Guryan et al. (forthcoming) show by simulation that there exists a mechanical negative correlation between characteristics of individuals and their peers in professional golf tournaments even when group assignment is random and they propose a modified test of random assignment that is well-behaved. First, I show both analytically and in simulations that the mechanical negative correlation *increases* as 1) group size (i.e., class size) grows, holding the number of groups constant, and as 2) when the number of groups formed decreases, holding group size fixed. This finding implies that a typical test of random assignment is more likely to reject the null of zero correlation when the number of randomly assigned groups in the sample is small. It also means that past studies using exogenous variation from a small number of classrooms or grade levels within a school might underestimate the size of peer effects. The bias will be present regardless of whether the assignment of peers is random or not. Thus, there is a mechanical explanation for why past studies looking at grade level peer effects tend to find smaller effects than those looking at classroom level peer effects. Second, I show that when the number of observations varies considerably across randomization tracks, the modified test of random assignment proposed by Guryan et al. (forthcoming) may still reject the null hypothesis of zero correlation. This means that neither the typical test of random assignment nor the modified test of random assignment is generally robust to the negative bias. I argue that testing whether group means (i.e., average test scores across classrooms) jointly differ at baseline is sufficient to verify whether group assignment is exogenous. More importantly, I present a robust solution by including an individual's baseline test score as a regressor, yielding a consistent estimate of the peer effect. This also highlights why including a student's own lagged achievement as a regressor is important.

I find that a one standard deviation increase in the average baseline math score of classmates leads to a 0.50 standard deviation increase in a student's own math score. This estimated causal effect of peers on math achievement is the first reported for a classroom setting that is closer to a typical western classroom and free of the mechanical negative correlation. The estimate is larger than estimates by previous studies that use a fixed effects estimator or an instrumental variable estimator to address selection bias and the reflection problem. I argue that the larger magnitude of estimated peer effects is attributed to the control for selection bias and the correction for the mechanical negative correlation. Furthermore, I do not find that the effect of peers varies across different types of students. I also do not find evidence that classroom

heterogeneity – measured by the standard deviation of classroom baseline test scores – matters. The results imply that ability grouping will lead to greater educational inequality, but not affect mean achievement.

Possible mechanisms for peer effects are through effort and conduct. To assess whether having high-achieving peers induces a student to increase the amount of effort and facilitates better behavior, I examine whether the average baseline test scores of classmates has an effect on the number of classes a student missed and on the incidence of discipline violations cited. Class absences and discipline violations are also important measures of outcomes, given the link between schooling and earnings (Angrist and Krueger 1992) as well as the potential association between discipline and crime (Adams 2000) respectively. I find that a one standard deviation increase in the average baseline scores of classmates leads to a 0.52 standard deviation decrease in the number of classes a student missed and a 0.81 standard deviation decrease in the incidence of discipline violations. This finding provides some support that having higher achieving peers can alter a student’s effort and conduct.

2. Identification and Interpretation of Peer Effects

This study is primarily concerned with identifying the effect of classroom peer achievement on a student’s own achievement. Consider the following linear-in-means model:

$$y_{ickt} = \beta_0 + \beta_1 \bar{y}_{-ickt} + \delta_k + \varepsilon_{ickt} \quad (1)$$

The dependent variable, y_{ickt} , is the achievement of student i , in classroom c , cohort-ability group k , in period t .⁴ β_0 is the intercept term and δ_k is a vector of cohort-ability group dummies. The coefficient of interest, β_1 , measures how the average achievement of student i ’s classmates, \bar{y}_{-ickt} , affects student i . ε_{ickt} is the error term.

Equation (1) suffers from Manski’s (1993) reflection problem because researchers cannot effectively isolate the effect of a person’s behavior on the behavior of the group from the effect of the behavior of the group on the person.⁵ Regression equation (1) may yield an estimate of β_1 that is non-zero because: (a) the behavior of peers, as measured by their achievement, influences

⁴ There are 7 cohorts of students and 2 ability groups for each cohort of students. So $k = 13$ (excluding the intercept).

⁵ The reflection problem is analogous to the problem of interpreting the simultaneous movements of a person and her reflection in a mirror, as it may not be clear to a naïve outside observer whether the mirror image causes the person’s movements or simply reflects them.

student i ; (b) similar students are assigned into the same classroom making the achievement of student i and the average achievement peers correlated; (c) student i affects her peers; or (d) good (bad) teachers are assigned to the classroom making the achievement of student i and the average achievement peers positively (negatively) correlated. (b), (c), and (d) are confounding factors that make it difficult to identify the causal effect of peers on student i (i.e., (a)).

If students are randomly assigned into classrooms, an Ordinary Least Squares (OLS) estimate of β_1 in equation (1) will not be biased by unobserved selection of students into classrooms. When students are randomly assigned into classrooms, teachers are also effectively randomly assigned into classrooms.⁶ Hence, there is no systematic relationship between the quality of teachers and the initial composition of a classroom. Random assignment of students into classrooms effectively removes what Manski (1993) terms correlated effects (i.e., factor (b) above).

Because student i and her classmates simultaneously influence each other, we can also express the achievement of peers as a function of student i 's achievement:

$$\bar{y}_{-ickt} = \tilde{\beta}_0 + \tilde{\beta}_1 y_{ickt} + \tilde{\delta}_k + \tilde{\varepsilon}_{ickt} \quad (2)$$

where $\tilde{\beta}_1$ measures the direct effect student i on peers. That is, equation (2) measures the reverse causation of student i on peers or factor (c) listed above. Equation (1) and equation (2) suggest that student i affects her classmates and her classmates affect her at the same time. An OLS estimate of β_1 in equation (1) does not show the direct effects of peers on student i , but the net effect of their repeated influences. This is essentially the endogeneity problem that Moffitt (2001) discussed. The commonly adopted solution to this simultaneity problem is to replace \bar{y}_{-ickt} in equation (1) with the lagged achievement of classmates, $\bar{y}_{-ickt-1}$ (e.g., Betts and Zau, 2004; Hanushek et al. 2003; Hoxby and Weingarth, 2007; Zimmerman, 2003).⁷ Since baseline test scores are determined in elementary school prior to students being randomly assigned into classrooms, this modification effectively removes the feedback effect. Furthermore, because

⁶ It is possible that some classrooms have higher average initial achievement than other classrooms do. If teachers are systematically assigned into classrooms according to these differences, then selection bias may still be present.

⁷ Replacing the current average achievement of peers with the lagged average achievement of peers is not necessarily a perfect solution if students are repeatedly assigned into the same classrooms. However, given that students from dozens of different schools were quasi-randomly assigned into different classrooms, using lagged average achievement of peers is appropriate for the current sample.

lagged achievement is not affected by the current classroom environment, common shock effect or factor (d) listed above will also be removed by replacing \bar{y}_{-ickt} in equation (1) with the lagged achievement of classmates, $\bar{y}_{-ickt-1}$. Regression equation (1) is then modified as:

$$y_{ickt} = \beta_0 + \beta_1 \bar{y}_{-ickt-1} + \delta_k + \varepsilon_{ickt} \quad (3)$$

where the estimated β_1 does not capture the reverse causation and common shock effects.

It is important to note that this paper provides a reduced-form estimate of the effect of peers on achievement, as baseline peer achievement is predetermined and may proxy for unobservable peer traits, such as ability, motivation, and effort.⁸ In addition, because the response of a teacher may also vary with the effort and ability of students in the classroom, peer achievement will also capture the indirect effects of peers through teachers.

3. Institutional Background and Data Description

This paper draws data from a large independent Chinese secondary school in Johor, Malaysia. There are generally two types of secondary schools in Malaysia: public schools and independent Chinese schools.⁹ Public secondary schools are run by the government and the language of instruction is primarily Malay.¹⁰ Independent Chinese secondary schools are fee-paying schools run by ethnic Chinese communities. They generally follow curricula set by the United Chinese School Committees Association of Malaysia. The primary language of instruction in these schools is Chinese (Mandarin), although some schools use English textbooks for some subjects. These schools charge relatively low fees, have large class sizes, and rely heavily on donations from local Chinese communities.¹¹ Independent Chinese secondary schools divide grade levels into lower division and upper division. Lower division spans from grade 7 to grade 9 while upper division spans from grade 10 to grade 12.

⁸ Cooley (2009) argues that since peer achievement is a proxy for unobservable peer traits, such as ability, motivation, and effort, equation (1) or (3) does not disentangle what Manski (1993) refers to as endogenous peer effects (e.g., effort) and contextual peer effects (e.g., ability).

⁹ There are also a small number of private schools that are run by private entities. These so called Private Educational Institutions (PEIs) provide education at preschool, elementary, and secondary levels. Some of these schools follow public school curricula and some follow foreign educational systems.

¹⁰ Public secondary schools are also called national secondary schools. Since 2003, the medium of instruction in math and science has been English, but the government recently announced that the medium of instructions in math and science would be switched to Malay.

¹¹ These features are similar to the Catholic schools in the United States and other countries. Ten percent of the students in the sample school receive some form of financial aid or scholarship.

The sample school admits roughly 1000 ethnic Chinese students each year who attended public Chinese elementary schools in the metropolitan area.¹² The Malaysian government halted the opening of new independent Chinese secondary schools in the 1950s, causing excess demand for Chinese secondary education in Malaysia. The school uses standardized admission examinations to ration student slots. A student's aggregate test score in four subjects, Chinese, English, Malay, and Math, determines whether she is admitted into the school or not. Students scoring below a cutoff point are not admitted, but the school accepts transfer students from other independent secondary schools conditional on the students' past achievement.¹³

This paper analyses the effects of peers in grade 7 for academic years 2002 to 2008 using a sample of 6495 students.¹⁴ Grade 7 students in the secondary school are tracked into one of two groups (group A or group B) on the basis of their admission (aggregate) test scores. Generally, the top 300 students are tracked into group A, and the rest, on average about 700, into group B.¹⁵ The average class size at time of assignment is 50 students, but not all students assigned into classrooms subsequently attend the school.¹⁶ Within each track, ability mixing is enforced through a quasi-randomization process. For example, if the male-female ratio is 2:3 in group A, the school will set the number of males per class to be 20, and the number of females per class to be 30. The school will then assign the top female scorer on the admission exam to class 1, the second-best female to class 2, the 6th female to class 6, the 7th female to class 6, the 8th female to class 5, and so forth. Male students are also assigned in the same way.¹⁷ In group B, there are usually 20 to 40 grade repeaters from the previous academic year and the school uses their test scores in grade 7 as the basis for the quasi-random assignment process. Because the school attempts to equalize average test scores across classrooms, selection bias is controlled for. This means that teacher assignment within each track is also effectively free of selection bias.

¹² Less than 1 percent of the student population at this school is non- Chinese.

¹³ It is extremely rare to have transfer students in grade 7.

¹⁴ The school began using admission examinations in 1979, but records of admission examination and test takers before academic year 2002 were lost.

¹⁵ The number of group B classes varies from year to year, ranging from 13 to 19 in the sample.

¹⁶ According to the school administrators, some students who registered for enrollment did not subsequently attend the school because they were admitted into secondary schools in Singapore with generous scholarship awards or were admitted to elite local public secondary schools. As a result, there are generally larger variations in class size than the initial assignment shows.

¹⁷ The assignment is done by a computer program. This classroom assignment method is similar to the one practiced in South Korea as described by Kang (2007).

Students attend all classes with the same set of peers during the school year.¹⁸ This greatly reduces the opportunities for students in different classrooms, especially those that never met prior to secondary school, to interact.¹⁹ This setup is unusual compared to secondary schools in most other countries where students attend different classes with different sets of peers. The quasi-random assignment of racially homogenous students of the same age into classrooms, where students attend all classes with a fixed set of classmates, provides a unique setting for the study of classroom-level peer effects.

Students are regularly given standardized tests during a school year. The school calculates mid-year and end-of-year test scores based on the weighted average of all test scores for the respective semester. The tests are standardized within each track for some subjects and across tracks for others, but are graded by teachers of the respective classes.²⁰ However, in Chinese, English, and Malay classes, because test scores include grades in writing assignments administered and graded by respective teachers, test scores in these subjects are less comparable across classrooms. In this study, I focus on math scores as the outcome measure of achievement since math tests are standardized across classrooms within the same track and there is less room for differences in grading standards across teachers.²¹ I report estimates of peer effects on Chinese, English, and Malay in Appendix 3. The final grade at the end of the school year will determine whether students are promoted to group A, demoted to group B, or forced to repeat the same grade in the following year. Because students are reshuffled across tracks and classrooms in grade 8, I restrict the analysis to grade 7. The school also maintains records of students' attendance and number of school discipline codes violated, which can be used as measures of student outcomes.²² Descriptive statistics of 6,495 grade 7 students in 138 classrooms from academic year 2002 to academic year 2008 are reported in Table 1.

¹⁸ Students may participate in extracurricular activities with students in other classrooms and grade levels. Because extracurricular activities are optional, only approximately 60% of students participate in them.

¹⁹ There are roughly 30 feeder elementary schools in the city.

²⁰ For example, math tests are standardized within each ability group, but not across ability groups. Typically, there are two parts to a test, where the first part involves multiple choice questions and the second part involves short answers.

²¹ The fact that teacher assignment is effectively randomized means that any grading differences will add noise but not bias to estimates.

²² Teachers take attendance throughout the day for every single class. Academic penalty in the form of final grade point deduction and disciplinary penalty in the form of discipline point deduction may be imposed on students who miss classes. Students may be subject to probation or suspension if their cumulative discipline point deductions

4. Evidence of Quasi-Random Assignment into Classrooms

To demonstrate that students were not selectively assigned into classrooms, I run the following regressions using baseline test scores as the dependent variable:

$$y_{ickt-1} = C_c + \delta_k + \varepsilon_{ickt-1} \quad (4)$$

where C_c is a set of classroom fixed effects. If students were assigned into classrooms in the way that the school claimed, all classroom means should be statistically similar. This is equivalent to testing whether C_c are individually and/or jointly different from zero. This test is much stronger than necessary and is not the typical test of random assignment that other studies of peer effects used (e.g., Sacerdote 2001). However, the typical test of random assignment likely suffers from a negative bias when the number of classroom is small and I will discuss this issue in the next section.

Table 2 reports the joint tests of whether at least one of the classroom fixed effects is significantly different from zero, as well as the fraction of classroom fixed effects having individual p-values less than certain thresholds, using different baseline test scores as dependent variables. The top panel in Table 2 shows that we cannot reject the null hypothesis that all classroom fixed effects are jointly equal to one another. The bottom panel in Table 2 demonstrates that the fraction of individual significant classroom effects is close to what the size of the test suggests.

Given that baseline test scores are statistically similar across classrooms, one may be worried that there is not enough cross-classroom variation in baseline test scores for the identification of peer effects. Table 3 reports the summary statistics for classroom-level average test scores after conditioning on the set of cohort-ability group fixed effects. That is, each observation is a classroom's average of regression residuals across individuals based on the equation $y_{ickt-1} = \delta_k + \varepsilon_{ickt-1}$. The standard deviations here provide some assurance that there is still a reasonable amount of variation in the data.

5. A Typical Test of Random Assignment and Mechanical Negative Correlation

reach a certain threshold. Students who never miss a single class during the school year are given honors, which can offset any disciplinary penalties they received.

5.1 A Typical Test of Random Assignment

Guryan et al. (forthcoming) point out that when the sample size N is small a typical test of random assignment of the following form will likely falsely reject the null hypothesis that $\alpha_1 = 0$:

$$y_{ickt-1} = \alpha_0 + \alpha_1 \bar{y}_{-ickt-1} + \delta_{kt} + \varepsilon_{ickt} \quad (5)$$

Note that baseline achievement is measured at the end of grade 6, while cohort track fixed effects are specific to grade 7. Guryan et al. demonstrate in Monte Carlo simulations that the OLS estimate of α_1 in equation (5) will be negative even when peers and outcomes are randomly generated.²³ To illustrate this problem, consider a simple case when test scores are randomly generated, only one classroom is formed, and we estimate the following model:

$$y_{ic} = \alpha_0 + \alpha_1 \bar{y}_{-i,c} + \varepsilon_{ic} \quad (6)$$

In this case, OLS estimator always yields $\hat{\alpha}_1 = -(N-1)$ and $\hat{\alpha}_0 = N\bar{y}$, because by definition:

$$y_{ic} = N\bar{y} - (N-1)\bar{y}_{-i,c} \quad (7)$$

The correlation between y_{ic} and $\bar{y}_{-i,c}$ is negative because each y_{ic} in the sample is drawn without replacement. We can also see that the bias does not diminish, but increases with N , in this one-classroom problem.

If C classrooms of size n are formed in the sample of N students, the OLS estimator will yield:

$$\hat{\alpha}_1 = \frac{\sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y})(\bar{y}_{-i,c} - \bar{y})}{\sum_{c=1}^C \sum_{i=1}^n (\bar{y}_{-i,c} - \bar{y})^2} \quad (8)$$

Note that $\bar{y}_{-i,c} = \frac{n\bar{y}_c - y_{ic}}{n-1} = \frac{n}{n-1}\bar{y}_c - \frac{1}{n-1}y_{ic}$, where \bar{y}_c is the mean outcomes of classroom c that individual i is in, and $N = C \cdot n$.

Table 4 illustrates how the mean of $\hat{\alpha}_1$ from regression equation (5) is related to various sample sizes (N) and class sizes (n) based on 10,000 repetitions of Monte Carlo simulations

²³ Lyle (2007) observed this mechanical negative correlation but did not examine its property. De Giorgi et al. (forthcoming) implement an instrumental variable strategy that relies on this mechanical negative correlation in the first stage of the regression to avoid the bias associated with weak instrumental variable.

where $y_{ic} \sim iid N(0,1)$. For example, when n is fixed at 2, and N increases from 4 to 1024, the average of the estimated slope coefficients decreases from -0.33 to -0.0004 (the first column). When we hold N fixed, and increase n , the negative relationship gets stronger. When we hold C fixed and increase both N and n , the negative bias also gets more severe. This highlights why in studies that look at college roommates, such as that by Sacerdote (2001), where n is 2 and N is 1589 students, the test of random assignment fails to reject the null hypothesis that $\alpha_1 = 0$. On the other hand, if a peer effect study uses variation of student characteristics across classrooms within a school that has about 512 to 1024 students and an average class size of 32, the negative bias can be as large as 0.1 to 0.5 standard deviations. This negative bias is comparable to some of the peer effect estimates previously reported.

The negative correlation is present because sampling of individuals is done without replacement. For a given sample, the sample mean is known. If an individual's peers have above-average characteristics, the individual must have below-average characteristics. When the number of classrooms in a sample is small, the weight of the average characteristics of an individual's peers on the sample mean is large. For a given sample average, if the peers have above-average characteristics, then the individual's characteristics must be way below the average, leading to a strong negative correlation between the individual's characteristics and the average characteristics of peers. However, as the number of classrooms formed within the sample increases, the weight of the average characteristics of the peers on the sample mean becomes smaller. Indeed, as the number of classrooms approaches infinity, an individual's classmates are similar to a randomly drawn individual in an extremely large sample. Thus, as the number of classrooms increases, the negative correlation between the individual's characteristics and the peers' characteristics becomes weaker and the negative bias diminishes. Appendix 1 shows that $\hat{\alpha}_1$ converges in distribution to a random variable with a non-positive mean, as class size $n \rightarrow \infty$, while holding the number of classrooms, C , fixed. On the other hand, $\hat{\alpha}_1$ converges in probability to zero, as $C \rightarrow \infty$, while holding n fixed.

The presence of this mechanical negative correlation means that: (i) researchers may find non-positive peer effects when the data used have a small number of groups (classrooms), even if the causal effect of peers is positive;²⁴ (ii) a typical test of random assignment using baseline test

²⁴ That is, the null hypothesis that a typical study of peer effects tests against is incorrect.

scores will suffer from a negative bias, especially in cases where class size is relatively large or the number of classrooms is small; and (iii) there is a mechanical explanation for why estimated peer effects at the grade level tend to be smaller than estimated peer effects at the classroom level.

5.2 A Modified Test of Random Assignment

Guryan et al. (forthcoming) (GKN) propose a modified test of random assignment that does not suffer from this mechanical bias. They argue that the typical test of random assignment used in the study of peer effects can be modified as:

$$y_{ickt-1} = \alpha_0 + \alpha_1 \bar{y}_{-ickt-1} + \alpha_2 \bar{y}_{-ikt-1} + \delta_k + \varepsilon_{ickt-1} \quad (9)$$

where the additional variable \bar{y}_{-ikt-1} is the average test score of all individuals in the same randomization track k except individual i . They essentially propose including the “one-classroom” measure of peers shown in equation (7). I will refer to \bar{y}_{-ikt-1} as “GKN’s correction term”. GKN’s correction method requires that there be multiple tracks of randomization and that the population size differs across tracks (otherwise, OLS estimators will always yield $\hat{\alpha}_1 = 0$ and $\hat{\alpha}_2 = -(N-1)$). They show that by including their correction term in the test of random assignment, the test becomes well-behaved in Monte Carlo simulations, whether the sample size is small or large.

However, when randomization tracks differ considerably in size, GKN’s correction term will not fully correct for the negative bias. To illustrate this problem, consider two hypothetical schools, A and B, each of which has two tracks and students are randomly assigned into classrooms (of size 50) within each track. In school A, track 1 has 1000 students and track 2 has 950 students. In school B, track 1 has 1000 students and track 2 has 500 students. Furthermore, assume that the baseline test score of each student y_{i0} is *iid.* $N(0,1)$. I run the modified test of random assignment, i.e., equation (9), in 10000 Monte Carlo repetitions to evaluate whether the average of $\hat{\alpha}_1$ centers at zero. The averages of $\hat{\alpha}_1$ for school A and school B are reported in column (1) and column (2) of Table 5 respectively. Column (1) shows that GKN’s correction term successfully removes the negative bias for school A in the test of random assignment and column (2) shows that GKN’s correction term fails to remove the negative bias in the test of random assignment for school B. Therefore, the modified test of random assignment proposed by

Guryan et al. (forthcoming) is not generally reliable when the randomization tracks do not have similar size.

5.3 Correction for The Mechanical Negative Bias

Although we may not use the modified test of random assignment to verify whether students were randomly assigned into classrooms, we must still control for the negative bias. Expressing regression equation (3) in the following way will help elucidate why the negative correlation must be controlled for:

$$y_{ickt} = \beta_0 + \beta_1 \bar{y}_{-ickt-1} + \delta_{kt} + \beta_2 y_{ickt-1} + v_{ickt} \quad (10)$$

where $\varepsilon_{ickt} = \beta_2 y_{ickt-1} + v_{ickt}$.

Since $\text{cov}(y_{ickt}, y_{ickt-1} | \delta_{kt}) \neq 0$ according to numerous studies, and when n and N are fixed and finite it is likely that $\text{cov}(y_{ickt-1}, \bar{y}_{-ickt-1} | \delta_{kt}) < 0$, β_1 estimated using regression equation (3) will be biased downward when the individual's own baseline test score is omitted. This means that once y_{ickt-1} is controlled for, v_{ickt} will not be correlated with $\bar{y}_{-ickt-1}$.

The following simulation helps demonstrate the effectiveness of an individual's own baseline test score in absorbing the negative bias. Assume that peer effects are absent and that each student's outcome is solely determined by her own baseline test score in the following form²⁵:

$$y_{it} = 0.8y_{it-1} + 0.6u_{it} \quad (10)$$

where $u_i \sim iid. N(0,1)$. Note that parameter values are arbitrarily chosen to make $y_{it} \sim N(0,1)$. If an individual's own baseline test score is an effective control variable for the mechanical negative bias, we would expect the following regression:

$$y_{ickt} = \alpha_0 + \alpha_1 \bar{y}_{-ickt,t-1} + \alpha_2 y_{ickt-1} + \delta_k + v_{ickt} \quad (11)$$

to yield $\hat{\alpha}_1 = 0$ on average. On the other hand, if peer effects are present such that:

$$y_{it} = 0.5\bar{y}_{-i,t-1} + v_{it}, \quad (12)$$

then we would expect the OLS estimator to yield $\hat{\alpha}_1 = 0.5$ on average.

²⁵ This can be viewed as outcome being solely determined by ability.

Table 6 columns (1) and (2) report the average of $\hat{\alpha}_i$ as assumed in equation (10) and equation (12) respectively based on 10000 Monte Carlo repetitions of regression equations (11) for a hypothetical school with 1500 students and 30 classrooms. Columns (1) and (2) show that the average of $\hat{\alpha}_i$ corresponds to the values specified in equation (10) and equation (12), respectively. The Monte Carlo simulation results provide support for using own baseline test scores of student i to control for the mechanical negative correlation.

To summarize, this section examined the mechanical negative correlation inherent in a typical linear regression framework used to estimate peer effects. I also show that the typical test of random assignment and the modified test of random assignment used by previous peer effect studies are not always reliable. Because of this mechanical negative correlation, it is important to include a student's own baseline test score as a regressor. On the other hand, to verify random assignment, peer effect studies can test whether baseline test scores are similar across groups as illustrated in table 2.

6. Results

6.1 Peer Effects on Math Achievement: Linear-in-means Specification

Columns (1) to (8) in Table 7 presents OLS estimates of the effect of peers on a student's own math score based on equation (10). Columns (1) to (4) report estimates using math score in semester 1 as the dependent variable, while columns (5) to (8) report estimates using math score in semester 2 as the dependent variable. Columns (9) to (12) report first-differenced estimates of peer effects that exclude track fixed effects.

Columns (1) and (5) show that without controlling for the mechanical negative correlation, the effect of peers on math achievement is not statistically different from zero, even at the 10% level. Columns (3) and (7) show that if we ignore that two ability groups that exist for each cohort of students and rely on lagged achievement to control for selection bias, the effect of peers will be severely underestimated. In particular, column (3) shows that we may even estimate a significant negative effect of peers when it is in fact positive and insignificant. Columns (2), (4), (6) and (8) show that using a student's own baseline test scores as control variables leads to significant larger estimates of peer effects. The results indicate that peer effects are statistically significant in semester 2, but not in semester 1, suggesting that it takes time for peer effects to

result in observable changes in own outcomes. According to the preferred specifications, i.e., column (4) and column (8), a student's own math score is predicted to increase by 0.32 points in semester 1 and 0.89 points in semester 2 for every one point increase in the average baseline math scores of peers. The effect of peers on a student's own math score in semester 2 is equivalent to a change of 0.50 standard deviations.²⁶

Columns (9) to (12) show that the effect of peers will be underestimated when researchers do not know how students are assigned into classrooms and attempt to control for unobserved selection by using a first-difference (or fixed-effects) estimator. Estimates reported in columns (9) to (2) are based on regression equations that use the difference between an individual's semester 2 test score and semester 1 test score as the dependent variable and the difference between average peers' test score in semester 1 and average peers' test score at baseline as the regressor of interest, and includes a set of cohort fixed effects. Since there are only two periods for each student, this first-difference estimator is equivalent to a fixed-effects estimator. Although this set up is slightly different to other studies, such as Hanushek et al. (2003) and Betts and Zau (2004), where different peer groups are observed over periods, it serves to illustrate how first-difference estimator will still yield much smaller estimates. Column (9) to column (11) shows that the estimates of peer effects are negative. Even when a set of individual's baseline test scores is included as control variables, the estimate of peer effect is still close to zero.

In summary, I show that when we effectively control for unobserved selection and the mechanical negative correlation, the estimated effect of peers on an individual's math achievement is significantly positive and large. The size of the estimate is larger than most previous estimates, especially those based on fixed effects estimators. Indeed, when we ignore the presence of two ability groups and rely on student fixed effects to address unobserved selection in the current sample, the estimated effect of peers on math achievement is considerably lower than those that exploit the within-ability-group quasi-random assignment mechanism and include own baseline test scores as control variables.

²⁶ This figure is calculated as follow: $(10.03) \cdot (0.89) / 17.96 = 0.5$.

6.2 Differential Effects of Peers and Classroom Heterogeneity

The linear-in-means results presented in the previous section indicates that the way in which students are grouped can have distributional effects. Specifically, the strong positive effect of peers suggests that grouping high-achieving students with low-achieving students will benefit the latter at the cost of the former. However, the estimate does not indicate whether the grouping high-achieving students with low-achieving students will increase or decrease the average achievement of a school. For instance, if the peer effect is stronger for low-achieving students than is for high-achieving students, then ability grouping will raise average achievement of a school. Similarly, if classroom heterogeneity hurts student achievement, then ability grouping may raise average achievement. I consider two specifications to assess whether peer effects vary for different types of students and whether classroom heterogeneity matters to student achievement.

First, I consider the following non-linear specification, which allows the effect of peers to vary for students at different ranges of the initial test score distribution:

$$y_{ickt} = \beta_1 I_{ickt-1}^{Bottom_1} + \beta_2 I_{ickt-1}^{Middle} + \beta_3 I_{ickt-1}^{Top} + \gamma_1 I_{ickt-1}^{Bottom} \bar{y}_{-ickt-1} + \gamma_2 I_{ickt-1}^{Middle} \bar{y}_{-ickt-1} + \gamma_3 I_{ickt-1}^{Top} \bar{y}_{-ickt-1} + \pi_1 I_{ickt-1}^{Bottom} y_{ickt-1} + \pi_2 I_{ickt-1}^{Middle} y_{ickt-1} + \pi_3 I_{ickt-1}^{Top} y_{ickt-1} + \delta_k + \varepsilon_{ickt} \quad (14)$$

where I_{ickt-1}^P is an indicator of whether student i 's baseline math score is in the p range of the test score distribution and p is either bottom 25 percentile, middle 50 percentile, or top 25 percentile. Regression equation (14) allows the effect of average past achievement of peers to vary for students at different ranges of the initial test score distribution. Unlike past studies of peer effects using this particular type of non-linear specification, a set of interaction terms $I_{ickt-1}^P y_{ickt-1}$ is included to correct for mechanical correlations (Carrell et al. 2009). Appendix 2 shows Monte Carlo simulation results that compare the bias in cases where these correction terms are excluded and included, respectively.²⁷

Table 8 reports non-linear estimates of peer effects. For students whose baseline math scores were in the bottom 25 percentile of the test score distribution, for every one point increase in the average baseline math score of their classmates, their own math score is predicted to increase by 0.90 points. For students whose initial math scores were in the middle 50 percentile,

²⁷ I also assess some other widely used non-linear specifications and find significant relationships between individuals and peers when outcomes are *iid*. $N(0, 1)$ random variables. However, there is no easily implementable correction method available. Simulation results are available upon request.

the effect of peers is 0.86 points. For students in the top 25 percentile, the effect of peers is estimated to be 0.85 points. The effects of peers are not statistically different for students at different ranges of the baseline math score distribution. Thus, there is no evidence that peer effects are nonlinear for this sample of students.

The second specification adds the classroom standard deviation of baseline math score, s_{ct-1} , as a measure of classroom heterogeneity, in the linear-in-means model of peer effects:

$$y_{ickt} = \beta_0 + \beta_1 \bar{y}_{-ickt-1} + \beta_2 y_{ickt-1} + \beta_3 s_{ct-1} + \delta_k + \varepsilon_{ickt} \quad (15)$$

If classroom heterogeneity impedes achievement, then we would expect β_3 to be negative. Furthermore, I also estimate a version of equation (14) that adds the interaction terms of classroom standard deviation of baseline math score with a set of indicators of the range of a student's initial math score.

Table 9 reports whether the standard deviation of baseline math scores at the classroom has any effect on a student's current math score. Column (1) and Column (2) show that larger variation in the initial math scores of students in the same classroom has a negative effect on math achievement, although the effect is not significantly different from zero. Thus, there is not enough evidence to suggest that classroom heterogeneity impedes math achievement.

Table 10 reports whether the effect of classroom heterogeneity on math achievement differs across different types of students in semester 2. The point estimates show classroom heterogeneity is more likely to hurt students in the bottom of the initial test score distribution than students in the middle and top of the initial test score distribution. However, the estimates are not statistically different from one another.

In summary, the results presented in this section do not provide evidence that the effects of peers vary across different types of students. The results also do not provide significant evidence that students assigned into classrooms with greater variance in initial achievement are hurt.

6.3 Peer Effects on Class Absences and Discipline Violations

The estimates of peer effects reported in this paper indicate that the effects of peers are strong. Besides learning from peers, possible mechanisms for peer effects are through effort and conduct. To assess whether having better achieving peers induces a student to increase her effort, I investigate whether the average baseline test scores of peers predicts effort, using student's

absences as a proxy for own effort. Similarly, I examine whether having high achieving peers lower a student's total number of discipline violations.²⁸ Measured discipline violations may range from minor incidents, such as classroom disruption and dress code violations, to serious incidents, such as academic dishonesty, fighting, and theft. Specifically, we would expect that having better achieving peers will lower absence rates and discipline violation rates.

Table 11 reports the estimated effects of peers on class absences and discipline violations. Column (1) reports the effect of peers on the number of classes missed in semester 1; column (2) reports the effect of peers on the number of classes missed in semester 2; column (3) reports the effect of peers on the total number of classes missed during the academic year; column (4) reports the number of discipline violations cited during the academic year. The estimates indicate that high achieving peers have strong negative effects on class absences and discipline violations. For every one standard deviation increase in the average baseline scores of peers, the number of classes missed by a student during a school year is predicted to fall by 0.52 standard deviations. The effect of peers on discipline violation is even stronger. A one standard deviation increase in the average baseline scores of peers is associated with a 0.84 standard deviation decrease in the number of discipline violations cited.

7. Conclusion

This paper exploits the quasi-random assignment of students into classrooms in a large secondary school in Malaysia to estimate the causal effects of peers on math score, class absences, and discipline violations. Because the secondary school attempts to equalize average baseline test scores across classrooms and students attend all classes with the same set of peers, the data provide an unusual setting to estimate classroom peer effects that minimizes biases resulting from non-random selection.

I examine the mechanical negative correlation between the characteristic of an individual and the average characteristics of the individual's peers. Specifically, I show that the size of the negative bias is large when class size is large or when the number of classrooms is small. This mechanical negative correlation arises because individuals are sampled without replacement. For a given sample average, if an individual's peers have above-average ability, the individual's

²⁸ Figlio (2005) examines the impact of peer disruptive behavior on student behavior and achievement using the fraction of boys with female-sounding names in a classroom as an instrumental variable for peer behavior. My analysis looks at the "reverse causality" of peer achievement on discipline violations that include disruptive behavior.

ability must be significantly below the sample average by construction. When only a few classrooms are formed in the sample, the individual's peers are weighted heavily in the calculation of the sample average, which means that the individual must deviate more strongly away from the peers. The presence of this mechanical negative correlation means that a typical regression for estimating peer effect and the typical test of random assignment of peers may suffer from a negative bias. I also illustrate that the modified test of random assignment proposed by Guryan et al. (forthcoming) is not reliable when randomization tracks differ considerably in size. Since a typical test of random assignment is not generally reliable, researchers may test whether pre-determined characteristics of individuals significantly differ across classrooms to verify whether students are randomly assigned into classrooms. Most importantly, I demonstrate that including an individual's own baseline test score as a regressor is sufficient to control for the mechanical negative bias. This means that even when students are randomly assigned into classrooms, omitting lagged achievement can lead to the underestimation of peer effects.

I find that a one standard deviation increase in the average baseline math score of classmates leads to a 0.50 standard deviation increase in a student's current math score. The estimated effect of peers on math score is similar to what Duflo et al. (2008) estimated for Kenyan first graders. The current estimates are the first reported for middle income countries and are also larger than those estimated in most other studies in developed countries. For example, Kang (2007) reported that a one standard deviation increase in average peers' current math score is associated with a 0.26 to 0.27 standard deviation increase in a student's own math score in South Korean middle schools. Hanushek et al. (2003) report estimates of grade-level peer effects that range between 0.15 and 0.24 in Texas elementary schools.

One potential explanation for the strong effect estimated in this paper is the amount of classroom time that students in the current sample spent together. These students attended all classes with the same set of peers for 5 to 6 hours per day, 6 days a week, for about 35 weeks a year. Math classes account for roughly 1/6 of all classes. Second, because students in the sample are racially homogenous and about the same age, estimates are less likely to be confounded by race and age differences. This is consistent with Cooley's (2008) findings that heterogeneity in the racial mix of a classroom can potentially confound the estimated effect of average peers' achievement on a student's own achievement.²⁹ Third, the use of baseline test scores to control

²⁹ Angrist and Lang (2004) also find evidence of differential peer effects depending on the race of students.

for the mechanical negative correlation and the observation of all individuals in a class may perhaps explain why the estimate is greater than what Kang (2007) found in South Korean middle schools, where the classroom assignment method and institutional setting are similar. Fourth, I show that using a first-difference (or fixed-effects) estimator and/or including individual's lagged achievement as a regressor to address for selection bias (due to ability grouping) will likely underestimate the effect of peers.

The estimated peer effect is large. Having high-achieving peers may lead to better classroom discussions, make teachers better engage in teaching, induce students to increase effort, and discourage students from mischief. To assess whether having high-achieving peers induces a student to increase effort and facilitates better behavior, I examine whether the average baseline test scores of classmates has an effect on both the number of classes a student missed and on the incidence of discipline violations cited. I find that a one standard deviation increase in the average baseline scores of classmates leads to a 0.52 standard deviation decrease in the number of classes a student missed and a 0.81 standard deviation decrease in the incidence of discipline violations. This finding demonstrates having higher achieving peers can alter a student's effort and conduct and provides information on the effects of peers on measures that are related to earnings and criminal behavior.

Unlike previous findings, such as those by Burke and Sass (2008), and Hoxby and Weingarth (2007), I do not find evidence that peer effects vary across different types of students or that dispersion of peer quality matters to a student's own math achievement. The finding that dispersion of peer quality does not lower achievement is similar to findings by Duflo et al. (2008) in non-tracking Kenyan elementary schools, Kang (2007) in non-tracking Korean middle schools, and Hanushek et al. (2003) in Texas elementary schools. Overall, the current findings imply that whether ability grouping or mixing is used, it will have distributional effects on student achievement but no effect on efficiency. This means that if a school mixes high-achieving students with low-achieving students in the same classroom, average achievement of the school will not change, but low-achieving students will gain at the cost of high-achieving students.

A number of plausible factors may explain why this study does not find evidence that peer effects vary across different types of students. First, institutional differences in class size, classroom set up, teacher quality, etc., may explain the different results. Second, I illustrated in this paper that even with random assignment, linear-in-means and non-linear specifications are

still vulnerable to mechanical correlations. Although the specifications that I employed have easily implementable bias correction methods, other non-linear specifications may suffer from bias that arises from mechanical correlations. Second, this paper does not detect a significant effect of dispersion of peer quality – measured by the standard deviation of classroom baseline test scores – on student achievement. Because the school enforces ability mixing, the variation in dispersion of peer quality across classrooms may be too small for identifying the effect of classroom heterogeneity on achievement. It is also possible that “local” variations in the composition of the student population do not provide enough power for identifying the differential treatment effects of peers and the effect of classroom heterogeneity (Duflo et al., 2008). Specifically, a teacher may target instructions better when the variation in student quality within a classroom is small. The variation in the dispersion of peer quality across classrooms in an ability mixing environment may not be large enough for teachers to modify instruction style. Therefore, the strength of peer effect specifications in informing optimal classroom assignment policy may need to be reconsidered.

References

Adams, A. Troy (2000) "The Status of School Discipline and Violence" *Annals of the American Academy of Political and Social Science*, vol. 567, pp.140-156.

Angrist, Joshua D. and Alan B. Krueger (1991) "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, vol.106 (4), pp.979-1014.

Angrist, Joshua D. and Kevin Lang (2004) "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program," *American Economic Review*, vol. 94(5), pp.1613-1634.

Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster, and Josh Kinsler (2005) "Estimating Spillovers in the Classroom with Panel Data," Unpublished Manuscript.

Betts, Julian and Andrew Zau (2004) "Peer Groups and Academic Achievement: Panel Evidence from Administrative Data", Unpublished manuscript, UCSD.

Bishop, John (1989) "Is the Test Score Decline Responsible for the Productivity Growth Decline?" *American Economic Review*, vol.79, pp.178-197.

Bound, John and George Johnson (1992) "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations," *American Economic Review*, vol.82, pp.371-392.

Burke, Mary A. and Tim R. Sass (2008) "Classroom Peer Effects and Student Achievement," Federal Reserve Bank of Boston Working Paper No. 08-5.

Carrell, Scott E., Richard L. Fullerton, and James E. West (2009) "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, vol.27 (3), pp.439-464.

Cooley, Jane (2008) "Desegregation and the Achievement Gap: Do Diverse Peers Help?" Unpublished manuscript.

Cooley, Jane (2009) "Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box," WCER Working Paper No. 2009-8, September.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli (forthcoming) "Be as Careful of the Books You Read as of the Company You Keep: Evidence on Peer Effects in Educational Choices," *American Economic Journal – Applied Economics*.

Duflo, Esther, Pascaline Dupas, and Michael Kremer (2008) "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," Unpublished manuscript.

Figlio, David (2007) "Boys named Sue: Disruptive children and their peers," *Education Finance and Policy*, vol.2 (4), pp.376-94.

Foster, Gigi (2006) "It's Not Your Peers, and It's Not Your Friends: Some Progress toward Understanding the Educational Peer Effect Mechanism," *Journal of Public Economics*, vol.90 (8-9), pp.1455-1475.

Gamoran, Adam and Robert D. Mare (1989) "Secondary School Tracking and Educational Inequality: Compensation, Reinforcement, or Neutrality?" *American Journal of Sociology*, vol.94 (5), pp.1146-1193.

Guryan, Jonathan, Kory Kroft, and Matt Notowidigdo (forthcoming) "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments," *American Economic Journal – Applied Economics*.

Hanushek, Eric, John F. Kain, John Markman, and Steven G. Rivkin (2003) "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics*, vol.18 (5), pp.527-544.

Hoxby, Caroline and Gretchen Weingarth (2007) "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects," Unpublished manuscript, Harvard University.

Kang, Changhui (2007) "Classroom Peer Effects and Academic Achievement: Quasi-randomization Evidence from South Korea," *Journal of Urban Economics*, vol.61, pp.458-495.

Lyle, David (2007) "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point," *Review of Economics and Statistics*, vol.89 (2), pp.289-299.

Manski, Charles (1993) "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies*, vol.60 (3), pp.531-542.

Moffitt, Robert (2001) "Policy Interventions, Low-Level Equilibria, and Social Interactions," in *Social Dynamics*, S. Durlauf and H. P. Young eds., Cambridge: MIT Press.

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

Vigdor, Jacob and Thomas Nechyba (2007) "Peer Effects in North Carolina Public Schools," in *Schools and the Equal Opportunity Problem*, P. Peterson and L. Woessmann eds., MIT Press.

Zimmerman, David J. (2003) "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics*, vol. 85 (1), pp.9-23.

Table 1 Descriptive Statistics

Variables		Mean	Std. Dev.	Min	Max
<u>Own scores</u>					
Math score:	Semester 2	63.18	17.96	4.00	100.00
	Semester 1	66.93	14.41	11.00	99.00
	Baseline	61.79	14.14	19.00	100.00
Chinese score:	Semester 2	76.78	8.76	12.00	96.00
	Semester 1	75.41	8.40	31.00	94.00
	Baseline	66.45	11.50	16.00	99.00
English score:	Semester 2	67.75	12.17	14.00	98.00
	Semester 1	66.85	11.07	18.00	96.00
	Baseline	63.88	14.75	17.00	100.00
Malay score:	Semester 2	71.21	12.72	9.00	98.00
	Semester 1	72.57	11.86	20.00	98.00
	Baseline	59.74	15.99	6.00	97.00
Average test score:	Baseline	62.97	9.53	34.00	92.00
Class absences:	Full year	9.81	16.20	0.00	514.00
	Semester 2	6.07	11.42	0.00	471.00
	Semester 1	3.75	9.47	0.00	217.00
Discipline violations:	Full year	0.15	0.58	0.00	9.00
<u>Peers' scores</u>					
Math score:	Baseline	61.76	10.02	41.53	84.90
Chinese score:	Baseline	66.43	7.91	53.91	85.83
English score:	Baseline	63.90	10.64	46.96	86.42
Malay score:	Baseline	59.75	11.37	42.37	81.63
Average test score:	Baseline	62.96	7.87	52.44	78.83
<u>Other characteristics</u>					
Male (= 1)		0.48	0.50	0.00	1.00
Class size		48.79	1.84	45.00	53.00

Note: Sample includes 6,495 grade 7 students in academic years 2002 to 2008. There are 138 classrooms in the sample. Repeaters and students who dropped out are excluded from the sample.

Table 2 Tests of Quasi-Random Classroom Assignment

	<u>Baseline Test Scores</u>			
	Math	Chinese	English	Malay
F-statistics of classroom fixed effects (p-value)	0.83 (0.91)	0.85 (0.89)	0.79 (0.96)	0.74 (0.99)
<u>Fraction of fixed effects individually having:</u>				
p-value < 0.01	0.014	0.000	0.007	0.000
p-value < 0.05	0.072	0.043	0.022	0.022
p-value < 0.10	0.101	0.051	0.051	0.036
Total number of classrooms	138	138	138	138

Note: Sample includes 6495 grade 7 students and all specifications include a set of cohort x track fixed effects. Robust standard errors are used.

Table 3 Summary Statistics of Average Baseline Test Scores across Classrooms

Variables	Mean	Std. Dev.	Min	Max
Math score at baseline	-0.0008	1.296	-3.651	3.539
Malay score at baseline	-0.0016	1.346	-2.946	4.532
Chinese score at baseline	-0.0024	1.028	-3.069	2.427
English score at baseline	-0.0001	1.230	-3.363	3.225

Note: Sample size is 138 classrooms. These statistics were obtained in a few steps. First, regression residuals were calculated using estimates from equation: $y_{iikt-1} = \delta_k + \varepsilon_{iikt-1}$. Then, classroom-level averages were calculated and used to produce statistics in Table 3.

Table 4: The Mean of $\hat{\alpha}_1$ by N and n

Sample size (N)	Class size (n)									
	2	4	8	16	32	64	128	256	512	1024
2	-1									
4	-0.33 (0.006)	-3								
8	-0.146 (0.005)	-1.253 (0.012)	-7							
16	-0.065 (0.003)	-0.553 (0.008)	-2.587 (0.025)	-15						
32	-0.032 (0.002)	-0.258 (0.005)	-0.934 (0.014)	-4.505 (0.049)	-31					
64	-0.015 (0.002)	-0.111 (0.003)	-0.372 (0.007)	-1.319 (0.022)	-7.456 (0.094)	-63				
128	-0.010 (0.001)	-0.062 (0.002)	-0.168 (0.004)	-0.468 (0.009)	-1.712 (0.031)	-11.622 (0.174)	-127			
256	-0.003 (0.001)	-0.028 (0.002)	-0.075 (0.003)	-0.196 (0.005)	-0.523 (0.010)	-1.907 (0.041)	-16.995 (0.305)	-255		
512	-0.003 (0.001)	-0.013 (0.001)	-0.040 (0.002)	-0.090 (0.003)	-0.213 (0.005)	-0.565 (0.011)	-2.395 (0.063)	-25.704 (0.540)	-511	
1024	-0.0004 (0.0004)	-0.0068 (0.0008)	-0.020 (0.001)	-0.042 (0.002)	-0.099 (0.003)	-0.232 (0.005)	-0.585 (0.013)	-2.273 (0.065)	-37.029 (0.933)	-1023

Note: $y_{ic} \sim iid N(0,1)$. Bold coefficients are the averages of $\hat{\alpha}_1$ over 10,000 Monte Carlo repetitions. Standard errors are reported in parentheses.

Table 5: Bias when GKN's Correction is used as a Control

	School A	School B
	y_{it-1}	y_{it-1}
Average of estimated α_1 with GKN's correction as a regressor	0.00004 (0.00006)	-0.007 (0.001)***
Track 1 size (N_1)	1000	1000
Track 2 size (N_2)	950	500

Note: Regressors include an intercept, average y_{t-1} of peers, a track indicator, and the correction term. Class size is 50 students. Coefficients are the average of slope estimates of regression equation (9) based on 10000 Monte Carlo repetitions. Standard errors reported in parentheses.

Table 6: Average of Estimated Slope Coefficients When Own Baseline Score is a Control

	(1)	(2)
	y_{it}	y_{it}
Average of estimated slope coefficients	0.0003 (0.001)	0.50 (0.002)
The true parameter value	0.00	0.50

Note: Regressors include an intercept term, average $y_{i,t-1}$ of peers, and baseline test score $y_{i,t-1}$. Sample size of each regression is 1500 students and class size is 50 students. Coefficients are the average of slope estimates based on 10000 Monte Carlo repetitions. Standard errors reported in parentheses.

Table 7 OLS Estimates of Peer Effects on Math Score

	Semester 1				Semester 2				First Difference			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Peers' Average Baseline Math	0.09 (0.28)	0.29 (0.24)	-0.11 (0.05)**	0.32 (0.24)	0.68 (0.44)	0.86 (0.41)**	0.34 (0.07)***	0.89 (0.41)**				
Peers' First-Difference Average Lagged Math									-0.16 (0.09)*	-0.12 (0.09)	-0.15 (0.09)*	0.02 (0.09)
Own Baseline Math	-	0.74 (0.02)***	0.73 (0.02)***	0.74 (0.02)***	-	0.68 (0.02)***	0.67 (0.02)***	0.69 (0.02)***		0.06 (0.02)***		0.12 (0.01)
Own Baseline Chinese	-	-		0.20 (0.02)***	-	-		0.18 (0.02)***				0.03 (0.02)
Own Baseline English	-	-		0.09 (0.02)***	-	-		0.09 (0.02)***				0.05 (0.01)***
Own Baseline Malay	-	-		0.13 (0.02)***	-	-		0.16 (0.02)***				0.10 (0.01)***
Own Semester 1 Math											-0.003 (0.01)	
Cohort x track fixed effects	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No	No	No	No
Cohort fixed effects	-	-	Yes	-	-	-	Yes	-	Yes	Yes	Yes	Yes
R-squared	0.138	0.403	0.3865	0.436	0.179	0.325	0.317	0.350	0.061	0.064	0.061	0.0916
F-stat	39.45	200.19	374.69	183.35	26.27	112.61	178.94	105.95	8.80	10.34	7.78	16.43
<u>Standard deviation</u>												
Main regressor (peer characteristic)	10.03	10.03	10.03	10.03	10.03	10.03	10.03	10.03	9.08	9.08	9.08	9.08
Dependent variable	14.41	14.41	14.41	14.41	17.96	17.96	17.96	17.96	10.92	10.92	10.92	10.92
<u>Standard deviation change</u>												
Peer effect	0.06	0.20	-0.08	0.22	0.38	0.48	0.19	0.50	-0.13	-0.10	-0.13	0.02

Note: Sample includes 6495 grade 7 students. The bottom panel presents standard deviations of key variables and peer effect measured in standard deviation of change. Columns (9)-(12) use first-difference test score between semester 2 and semester 1 as the dependent variable and first-difference average lagged test score of peers as the measure of peer characteristic. Robust standard errors clustered at the classroom level are reported in parentheses. *** Significant at 1%; ** significant at 5%; * significant at 10%.

Table 8 Differential Effects of Peers

	Bottom 25%	Middle 50%	Top 25%
Peers' baseline math score	0.90 (0.42)**	0.86 (0.42)**	0.85 (0.41)**
P-value of joint test of differential peer effects	0.57	-	0.83
R-squared		0.33	
F statistic		87.08	

Note: Sample includes 6495 grade 7 students and the specification includes a set of cohort x track fixed effects, a set of dummies indicating whether the student is in the top, middle, or bottom of the initial distribution, as well as a set of interaction terms for baseline math score and whether the student is in the top, middle, or bottom of the initial distribution. Separate coefficients are estimated for students in each of the category defined by the baseline math score distribution. The p-value reported in the bottom row is related to the test of whether the effect of peers is different to students in the middle 50 percentile. Robust standard errors clustered at the classroom level are reported in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%.

Table 9 Effects of Classroom Heterogeneity on Math Score

	Semester 1	Semester 2
Peers' baseline math score	0.35 (0.24)	0.89 (0.40)**
Standard deviation of baseline math score of class	-0.23 (0.32)	-0.01 (0.51)
R-squared	0.44	0.35
F statistic	174.17	100.37

Note: Sample includes 6495 grade 7 students and all specifications include a set of cohort X track fixed effects and a student's own baseline test scores. Robust standard errors clustered at the classroom level are reported in parentheses. *** Significant at 1%; ** significant at 5%; * significant at 10%.

Table 10 Differential Effects of Classroom Heterogeneity on Math Score in Semester 2

	Bottom 25%	Middle 50%	Top 25%
Peers' baseline math score	0.91 (0.42)**	0.85 (0.42)**	0.85 (0.41)**
SD of baseline math score of class	-0.219 (0.55)	-0.061 (0.56)	0.177 (0.55)
P-value of joint test of differential peer effects	0.49	-	0.999
R-squared		0.33	
F statistic		76.99	

Note: Sample includes 6495 grade 7 students and all specifications include a set of cohort X track fixed effects, a set of dummies indicating whether the student is in the top, middle, or bottom of the initial distribution, as well as a set of interaction terms for baseline math score and whether the student is in the top, middle, or bottom of the initial distribution. Separate coefficients are estimated for students in each of the category defined by the baseline math score distribution. The p-value reported in the bottom row is related to the test of whether the effect of classroom heterogeneity is different to students in the middle 50 percentile. Robust standard errors clustered at the classroom level are reported in parentheses. *** Significant at 1%; ** significant at 5%; * significant at 10%.

Table 11 OLS Estimates of Peer Effects on Class Absences and Discipline Violations

	(1) Absence Semester 1	(2) Absence Semester 2	(3) Absence Full Year	(4) Discipline Violation
Baseline Peers' Average Scores	-0.28 (0.22)	-0.80 (0.28)***	-1.08 (0.37)***	-0.06 (0.02)**
R-squared	0.007	0.026	0.022	0.050
F-stat	2.31	11.02	7.49	6.55
<u>Standard deviation</u>				
Baseline peers' average	7.87	7.87	7.87	7.87
Current own absences or discipline violation	9.47	11.42	16.20	0.58
<u>Standard deviation change</u>				
Peer effect	-0.23	-0.55	-0.52	-0.84

Note: Sample includes 6495 grade 7 students and all specifications include a set of cohort X track fixed effects and own baseline test scores. The bottom panel presents standard deviations of key variables and peer effect measured in standard deviation of change. Robust standard errors clustered at the classroom level are reported in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%.

Appendix 1 Mechanical Negative Correlation

I show mathematically how the mechanical negative correlation exists in a typical study of peer effects, as well as its relationship with class size (n) and the number of classrooms (C) in this Appendix.

Consider the model, $y_{ic} = \alpha_0 + \alpha_1 \bar{y}_{-i,c} + \varepsilon_{ic}$, where C classrooms of size n are formed.

The OLS estimator is:

$$\hat{\alpha}_1 = \frac{\sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y})(\bar{y}_{-i,c} - \bar{y})}{\sum_{c=1}^C \sum_{i=1}^n (\bar{y}_{-i,c} - \bar{y})^2} \quad (\text{A1})$$

Note that $\bar{y}_{-i,c} = \frac{n\bar{y}_c - y_{ic}}{n-1} = \frac{n}{n-1}\bar{y}_c - \frac{1}{n-1}y_{ic}$, where \bar{y}_c is the mean outcome of classroom c that individual i is in.

We can express equation (A1) as:

$$\begin{aligned} \hat{\alpha}_1 &= (n-1) \frac{\sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y}) [n\bar{y}_c - (n-1)\bar{y} - y_{ic}]}{\sum_{c=1}^C \sum_{i=1}^n [n\bar{y}_c - (n-1)\bar{y} - y_{ic}]^2} \\ &= -(n-1) \frac{\sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y}) [y_{ic} - n\bar{y}_c + (n-1)\bar{y}]}{\sum_{c=1}^C \sum_{i=1}^n [y_{ic} - n\bar{y}_c + (n-1)\bar{y}]^2} \\ &= -(n-1) \frac{\sum_{c=1}^C \sum_{i=1}^n [y_{ic} - n\bar{y}_c + (n-1)\bar{y} + n(\bar{y}_c - \bar{y})] [y_{ic} - n\bar{y}_c + (n-1)\bar{y}]}{\sum_{c=1}^C \sum_{i=1}^n [y_{ic} - n\bar{y}_c + (n-1)\bar{y}]^2} \\ &= -(n-1) - (n-1) \frac{\sum_{c=1}^C \sum_{i=1}^n n(\bar{y}_c - \bar{y}) [y_{ic} - n\bar{y}_c + (n-1)\bar{y}]}{\sum_{c=1}^C \sum_{i=1}^n [y_{ic} - n\bar{y}_c + (n-1)\bar{y}]^2} \end{aligned}$$

$$\begin{aligned}
&= -(n-1) - (n-1) \frac{\sum_{c=1}^C n(\bar{y}_c - \bar{y})[(n^2 - n)\bar{y}_c + n(n-1)\bar{y}]}{\sum_{c=1}^C \sum_{i=1}^n [y_{ic} - n\bar{y}_c + (n-1)\bar{y}]^2} \\
&= -(n-1) + (n-1) \frac{n^2(n-1) \sum_{c=1}^C (\bar{y}_c - \bar{y})^2}{\sum_{c=1}^C \sum_{i=1}^n [y_{ic} - n\bar{y}_c + (n-1)\bar{y}]^2} \tag{A2}
\end{aligned}$$

The numerator of (A2) can be expressed as:

$$\begin{aligned}
\sum_{c=1}^C \sum_{i=1}^n [y_{ic} - n\bar{y}_c + (n-1)\bar{y}]^2 &= \sum_{c=1}^C \sum_{i=1}^n [y_{ic} - \bar{y} - n(\bar{y}_c - \bar{y})]^2 \\
&= \sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y})^2 - 2n \sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y})(\bar{y}_c - \bar{y}) + n^3 \sum_{c=1}^C (\bar{y}_c - \bar{y})^2 \\
&= \sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y})^2 - 2n^2 \sum_{c=1}^C (y_{ic} - \bar{y})^2 + n^3 \sum_{c=1}^C (\bar{y}_c - \bar{y})^2 \\
&= \sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y})^2 + (n^3 - 2n^2) \sum_{c=1}^C (\bar{y}_c - \bar{y})^2
\end{aligned}$$

Thus,

$$\begin{aligned}
\hat{\alpha}_1 &= -(n-1) + (n-1) \frac{n^2(n-1) \sum_{c=1}^C (\bar{y}_c - \bar{y})^2}{\sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y})^2 + (n^3 - 2n^2) \sum_{c=1}^C (\bar{y}_c - \bar{y})^2} \\
&= -(n-1) + (n-1) \frac{Cn(n-1) \frac{n}{C} \sum_{c=1}^C (\bar{y}_c - \bar{y})^2}{nC \left[\frac{1}{nC} \sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y})^2 \right] + n(n-2)C \left[\frac{n}{C} \sum_{c=1}^C (\bar{y}_c - \bar{y})^2 \right]} \\
&= -(n-1) + (n-1) \frac{n(n-1) \frac{n}{C} \sum_{c=1}^C (\bar{y}_c - \bar{y})^2}{n \left[\frac{1}{nC} \sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y})^2 \right] + n(n-2) \left[\frac{n}{C} \sum_{c=1}^C (\bar{y}_c - \bar{y})^2 \right]} \\
&= -(n-1) + (n-1) \frac{n(n-1)s_c^2}{ns^2 + n(n-2)s_c^2}
\end{aligned}$$

$$\begin{aligned}
\hat{\alpha}_1 &= (n-1) \left[\frac{n(n-1)s_c^2}{ns^2 + n(n-2)s_c^2} - 1 \right] \\
&= (n-1) \left[\frac{(n-1)s_c^2 - s^2 - (n-2)s_c^2}{s^2 + (n-2)s_c^2} \right] \\
&= (n-1) \frac{s_c^2 - s^2}{s^2 + (n-2)s_c^2} \\
&= (s_c^2 - s^2) \frac{n-1}{s^2 + (n-2)s_c^2} \tag{A3}
\end{aligned}$$

where,

$$s_c^2 = \frac{n}{C} \sum_{c=1}^C (\bar{y}_c - \bar{y})^2 \quad (\text{The between class variance of mean})$$

and,

$$\begin{aligned}
s^2 &= \frac{1}{nC} \sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y})^2 \\
&= \frac{1}{nC} \sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y}_c + \bar{y}_c - \bar{y})^2 \\
&= \frac{1}{nC} \sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y}_c)^2 + \frac{1}{nC} \sum_{c=1}^C \sum_{i=1}^n (\bar{y}_c - \bar{y})^2 \\
&= \frac{1}{nC} \sum_{c=1}^C \sum_{i=1}^n (y_{ic} - \bar{y}_c)^2 + \frac{1}{C} \sum_{c=1}^C (\bar{y}_c - \bar{y})^2 \\
&= s_u^2 + \frac{1}{n} s_c^2 \quad (\text{The overall variance})
\end{aligned}$$

Case 1: $n \rightarrow \infty$ holding C fixed

$$\hat{\alpha}_1 \rightarrow \underset{n \rightarrow \infty}{plim} (s_c^2 - s^2) \neq 0$$

If we assume that $y_{ic} \sim iid N(\mu, \sigma_y^2)$, then

$$s_c^2 \xrightarrow{d} \frac{1}{C} \sum_{c=1}^C (\sqrt{n}\bar{y}_c - \sqrt{n}\bar{y})^2 \stackrel{d}{=} \frac{1}{C} \sum_{c=1}^C (\xi_c - \bar{\xi})^2$$

where the distribution of ξ_c is $iid N(\mu, \sigma_y^2)$. In addition,

$$\begin{aligned}
s^2 &\rightarrow \underset{n \rightarrow \infty}{plim}(s_u^2) \\
&= \frac{1}{C} \sum_{c=1}^C \underset{n \rightarrow \infty}{plim} \frac{1}{n} \sum_{i=1}^n (y_{ic} - \bar{y}_c)^2 \\
&= \frac{1}{C} \sum_{c=1}^C \sigma_y^2 \\
&= \sigma_y^2
\end{aligned}$$

Thus,

$$\hat{\alpha}_1 \xrightarrow{d} \frac{1}{C} \sum_{c=1}^C (\xi_c - \bar{\xi})^2 - \sigma_y^2 \equiv \Omega$$

That is, $\hat{\alpha}_1$ converges in distribution to Ω , and:

$$E(\Omega) = \left(\frac{C-1}{C} - 1 \right) \sigma_y^2 = -\frac{\sigma_y^2}{C} < 0$$

Case 2: $C \rightarrow \infty$, holding n fixed.

Under the assumption that $y_{ic} \sim iid N(\mu, \sigma_y^2)$, $s_c^2 \xrightarrow{p} \sigma_y^2$ and $s^2 \xrightarrow{p} \sigma_y^2$, and hence,

$$\hat{\alpha}_1 \xrightarrow{p} 0$$

The inverse relationship between y_i and i 's classmates diminishes as C gets larger. As C gets larger, the between class variance of mean approaches the overall variance. However, holding C fixed, as n gets larger, $\hat{\alpha}_1$ converges in distribution to a distribution with a negative mean. In other words, the driving factor for the negative correlation is not the size of sample population (N) per se, but the number of classrooms (C) formed.

Appendix 2 Bias in a Non-linear Specification and a Correction Method

I present Monte Carlo simulation results showing the bias suffered from a commonly used non-linear model of peer effects in this appendix. In addition, I also show in Monte Carlo simulations that using baseline test score as a control variable can effectively control for the bias.

Consider using the following regression equation to examine whether the effects of peers vary across different types of students:

$$y_{ict} = \beta_1 I_{ict-1}^{Bottom_1} + \beta_2 I_{ict-1}^{Middle} + \beta_3 I_{ict-1}^{Top} + \gamma_1 I_{ict-1}^{Bottom} \bar{y}_{-ict-1} + \gamma_2 I_{ict-1}^{Middle} \bar{y}_{-ict-1} + \gamma_3 I_{ict-1}^{Top} \bar{y}_{-ict-1} + \varepsilon_{ict} \quad (A4)$$

where y_{ict} is the current achievement of student i , \bar{y}_{-ict-1} is the baseline achievement of student i 's peers, $I_{ict-1}^{Bottom_1}$ is an indicator for whether y_{ict-1} is in the bottom 25 percentile of the initial test score distribution, I_{ict-1}^{Middle} is an indicator for whether y_{ict-1} is in the middle 50 percentile of the initial test score distribution, and I_{ict-1}^{Top} is an indicator for whether y_{ict-1} is in the top 25 percentile of the initial test score distribution. Assume that students are randomly assigned into classrooms, peer effects are absent, and student i 's achievement is solely determined by her past achievement:

$$y_{ict} = 0.8y_{ict-1} + 0.6u_{ict} \quad (A5)$$

where $y_{ict-1} \sim iid. N(0,1)$ and $u_{ict} \sim iid. N(0,1)$. Simulations show that OLS estimates based on regression equation (A4) will produce non-zero $\hat{\gamma}_1$, $\hat{\gamma}_2$, and $\hat{\gamma}_3$ on average.

To illustrate the extent of the bias, consider an example where 1500 students are randomly assigned into 30 classrooms, each with 50 students, and achievement is determined by equation (A5). Table A1 column (1) reports the average of $\hat{\gamma}_1$, $\hat{\gamma}_2$, and $\hat{\gamma}_3$, which are statistically different from zero, based on 10,000 Monte Carlo regressions of (A4). Column (2) shows that even when the sample size is 5000 students and 100 classrooms are formed, the mechanical correlations remain significant.

I propose including a set of interaction terms of baseline test score with the indicator of where a student sat on the initial test score as control variables in equation (A4):

$$y_{ict} = \beta_1 I_{ict-1}^{Bottom_1} + \beta_2 I_{ict-1}^{Middle} + \beta_3 I_{ict-1}^{Top} + \gamma_1 I_{ict-1}^{Bottom} \bar{y}_{-ict-1} + \gamma_2 I_{ict-1}^{Middle} \bar{y}_{-ict-1} + \gamma_3 I_{ict-1}^{Top} \bar{y}_{-ict-1} + \pi_1 I_{ict-1}^{Bottom} y_{ict-1} + \pi_2 I_{ict-1}^{Middle} y_{ict-1} + \pi_3 I_{ict-1}^{Top} y_{ict-1} + \varepsilon_{ict} \quad (A6)$$

Column (3) and column (4) in table A1 show that adding these control variables is effective in absorbing the mechanical correlations. Comparing column (3) to column (1) and

column (4) to column (2), it is clear that adding the control variables can effectively reduce the extent of the bias.

Table A1 Bias with and without Individual Baseline Control Variables

	(1)	(2)	(3)	(4)
Average of γ_1 estimates	-0.0048 (0.0033)	-0.0031 (0.0017)*	0.0022 (0.0028)	0.0011 (0.0015)
Average of γ_2 estimates	-0.0103 (0.0018)***	-0.0025 (0.0010)***	0.0019 (0.0016)	0.0005 (0.0008)
Average of γ_3 estimates	-0.0135 (0.0033)***	-0.0031 (0.0017)*	-0.0031 (0.0028)	-0.0002 (0.0015)
Sample size (N)	1500	5000	1500	5000
Class size (n)	50	50	50	50
Number of classrooms (C)	30	100	30	100
Control variables	No	No	Yes	Yes

Note: Column (1) and column (2) are based on equation (A4). Column (3) and column (4) are based on equation (A6). Coefficients are the average of estimated effects of peers when peer effects are assumed zero using 10000 Monte Carlo repetitions. Standard errors reported in parentheses. *** 1% significant; ** 5% significant; * 10% significant.