You Are Who You Eat With

Evidence on Academic Peer Effects from School Lunch Lines

Jonathan L. Presler
Sinquefield Center for Applied Economic Research

Saint Louis University, Saint Louis, MO 63103

10/15/2020

Abstract

There is a sizable literature estimating the magnitude and importance of peer effects in education on academic outcomes. Most studies define who is in the peer group, but little work has been done to measure the intensity of connection - how important is each peer? Equally weighting all peers in a reference group assumes that all peers are equally important and may bias estimates towards zero by underweighting important peers and overweighting unimportant peers. I examine classmates using a novel approach to measure the intensity of connection between these students. I use administrative transaction data from the New York City Department of Education to observe the lunch line on a daily basis and use lunch line proximity as a measure of connection strength. The result is a revealed friendship network which I use to identify peer effects. I find that students who eat together are important influencers of one another's academic performance, with stronger effects in math than in reading. Further exploration of the mechanisms supports the claim that these are friendship networks. I also compare the strength of connections from different portions of the school year and find that connections formed at the beginning of the school year are most important, consistent with the story that long-term friendships are more important than short-term friendships.

Keywords: Peer effect, Network, Education.

Acknowledgements

Support for this project comes from the National Institute of Health (R01 grant: 5R01DK097347-02). I would like to thank the NYC Department of Education's Policy Research and Support Group and Office of Food and Nutrition Services for providing data and Stephen O'Brien, and Armando Taddei for their support. For data support, advice, and suggestions, I also thank Meryle Weinstein, Amy Ellen Schwartz, William C. Horrace, Hyunseok Jung, Alfonso Flores-Lagunes, and Joanna Bailey. I thank seminar participants at the Maxwell School at Syracuse University and at the APPAM conference for useful comments on previous drafts. The opinions expressed are those of the author and do not represent the views of the NYC Department of Education or NIH.

1 Introduction

Researchers and policymakers have long thought peer effects to be an important component of education.[1] Current policy discussions around tracking, school choice, and school integration are often based on the assumption that peers play an influential role in education production. A primary focus of the literature is the effect peers have on test scores, but estimates vary widely[2] due to differences in context, methodology, and how the peer group is defined. Social networks are by their nature hierarchical and complex, and each individual is uniquely impacted by a different set of peers. Understanding which peers are relevant is critical to identifying meaningful estimates of peer effects - both for research relevance and tautologically as many models include average group characteristics or outcomes. If the reference group has little impact on a student, or is too broadly defined, we may understate the importance of peers. However, if we define the set of peers too specifically, we may miss other influencers and misstate their importance.

A central question when identifying peer effects is how to define peers. Much of the literature has focused on the scope of the peer group - defining (or assuming) who is and who is not important, but little has been done to measure connection strength - how important each relevant peer is. Heterogeneity of effects may be explored based on race or sex (for example), but intensity of relationship is typically relegated to binary indicators centered around homophily. Bramoullé et al. (2009) points out that if individuals act outside of a group framework,[3] instead having individualized peer groups, then the endogenous and exogenous effects discussed in Manski (1993) can be separately identified. Separate identification of these effects is policy relevant and may imply different prescriptions. For example, Hoxby (2000) shows that elementary students benefit from the presence of a higher percentage of girls in the elementary school classroom. What this does not tell us is whether interaction with girls is important (the endogenous effect), or if it is related to, for example, behavioral differences where young boys demand more teacher time (an exogenous effect). If the former is true, it may be desirable for classrooms to be structured such that student interaction in mixed gender groups is increased. If it is the latter it may be beneficial for teachers to have an aide help manage student behavior.

This paper measures the intensity of connection between students in the same classroom and uses this network structure to determine social spillovers on math and reading test scores. We use administrative point of service (POS) data from the New York City Department of Education (NYCDOE) to observe daily lunch

1. For example, the Coleman Report (Coleman et al. 1966) looked at the achievement gap between whites and blacks and found differences in peer effects to be a contributor.

2. See Sacerdote (2011) for an excellent review of the education peer effect estimates. These estimates range from slight negative in Vigdor and Nechyba (2007) near -0.1 to slight positives in Burke and Sass (2007) near 0.05 to Hoxby (2000) with large estimates of 0.3 to over 6. These papers show the range of estimates, but most in the literature fall near the middle range.

3. Bramoullé et al. (2009) uses the term "group interactions" to refer to the idea that people interact and behave in groups, and that all members in the group are equally important. The group may be grade-cohorts or classrooms, for example.

transactions. We use these daily observations to determine which students frequently stand near one another in the lunch line as a measure of friendship. This friendship network has several important features. First, the measure of contact is based on administrative data rather than surveys. This means that friendships are revealed rather than stated, and their revelation means that the friendships need to be reciprocated.[4] Additionally, friendships change and we observe the result of daily decisions students make, rather than a single survey snapshot. Second, our measure of contact between students occurs during lunch, which is an important social space. Lunch is a relatively unstructured environment within school, allowing students more freedom to interact outside direct supervision from teachers and without direct academic consequences. This makes connections observed during the lunch period socially meaningful. Third, connections can have varying strengths,[5] and we allow them to vary on a continuous scale of importance. Next, these networks are constructed fresh during the school year as students do not have control over their classroom assignment, and we can observe how friendship significance evolves over time. Finally, because this network is individual-specific, we do not observe the perfect collinearity between group mean characteristics and mean expected group outcome (the reflection problem).[6]

We find evidence of significant social spillovers on academic outcomes. Effects are stronger in math than in reading, with a one standard deviation improvement in classroom peer performance resulting in an increase in own math performance between 7.5% and 11.1% of a standard deviation. This is on par with our estimates of the black and white achievement gap. Improvement in classroom peer performance in reading by a standard deviation improves own performance by between 4.1% and 6.3% of a standard deviation.

We measure friendship between students based on daily observations of contact over the entire school year. This allows us to treat friendships in different portions of the year as separate networks and observe the evolution of these connections over time. Our results suggest that connections formed at the start of the school year are most important, a finding that is consistent with Patacchini et al. (2017) in which long-term friends are found to be more influential than short-term friends.

There are two relevant literatures to which this paper contributes. The first is the literature around education peer effects on academic outcomes. We see a large variation in the reference group used by researchers. Cohort, or school-grade level, is frequently used as a reference group in order to avoid concerns around sorting into classrooms. Hoxby (2000) examines elementary school students at the cohort level and finds that students perform better in classrooms with higher proportions female, regardless of gender. Burke and Sass (2013) find cohort effects near zero using a student fixed effect model incorporating average peer

4. In survey data like Add Health, students nominate their friends. Lin and Weinberg (2014) use Add Health to show that reciprocated friendships are stronger than unreciprocated friendships on a variety of outcomes, including academics.

5. For example: student A considers student B to be her best friend, student C is a friend, and student D is simply a classmate.

6. As described in Manski (1993), and in section 4.1.

performance. However, they find that that classroom peers are more important than cohort peers and produce meaningful peer effects. This suggests that while choosing the cohort is a convenient way to avoid selection problems, estimates derived from this reference group may understate the overall peer effect. This is intuitive, as peer performance is often measured using mean outcome. Incorporating unimportant peers who contribute little to the outcome should bias estimates towards zero.

It is sometimes possible to zoom in further than the classroom level to examine the effect of students who almost certainly are in one anothers' social network. Sacerdote (2001) does this in the college setting by looking at the effect of roommates on one another. He finds that roommates are important sources of peer effects, but that these are lower bounds on the effects of peers because roommates are typically only a piece of a student's social network. Another way to look beyond the school, cohort, or classroom is to use information on the network structure itself. The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a workhorse in the peer effects literature as a result of its friendship survey of middle and high school students.[7] Lin (2010) uses the friendship network described in Add Health to identify academic peer effects among high schoolers using maximum likelihood estimation of a spatial autoregressive model (SAR). The friendship survey allows researchers to pick a set of relevant peers from the school (networks are within school, not within classroom), but the researcher does not know the relative importance of these students (ex: who is the student's closest and most influential friend).[8]

This paper also contributes to the the partially overlapping networks literature. The key insight from this literature is that identification problems such as reflection are solved when multiple reference groups partially overlap. The intuition is that collinearity issues related to the characteristics and outcome of the reference group are broken up when individuals participate in more than one network, and these networks share some but not all members. To our knowledge, an early draft of Laschever (2013) was the first paper to propose this method. Bramoullé et al. (2009) formalize the partially overlapping reference group approach and shows explicit conditions for overcoming the reflection problem. De Giorgi et al. (2010) provides an empirical example of partially overlapping networks. Students are randomly assigned to nine college courses. Strength of connection between individuals is based on the number of courses students take with one another, and the authors estimate an overall peer effect for this network. W. Horrace et al. (2019) relaxes the assumption that all networks are equally important and estimates the relative importance of networks based on ideas of homophily.

7. A litany of papers have been written using Add Health, including Bifulco et al. (2011) and Patacchini et al. (2017). The appeal of this data, which surveys and follows up with students who were 7-12th graders in US public schools during the 1994-1995 school year, is that it asks students who their friends are and includes survey responses on a variety of outcomes from GPA to smoking and drinking habits.

8. There are some exceptions. Patacchini et al. (2017) divides the networks into students who are friends in both waves and those who are not, finding that students who are long-term friends are more important than those who are friends in just one, both in the short and long term.

The contributions of this paper are as follows. First, we measure opportunities for contact between students in the lunch line to demonstrate a novel approach for constructing a revealed friendship network. Data in which social connections are observed are rare, and this is a large hurdle in empirical peer effects research. Second, our measure of contact allows us to not only refine the classroom environment and weight peers according to a revealed friendship network, but also overcome the binary nature typically seen in friendship networks wherein students are either friends or they are not. We weight friendships on what is essentially a continuous scale of importance, meaning that our results are not predicated on overweighting unimportant students and underweighting important peers as group averages necessarily do. Third, to our knowledge this paper presents an empirical application with the largest set of partially overlapping networks (one for each day of the school year) such that the measure of connection between individuals can be thought of as continuous. Finally, each student has a unique reference group allowing us to decompose the peer effect into its "endogenous" and "exogenous" components[9] using a linear in means model, a distinction which is important for accurate estimation of spillover effects.

The layout of this paper continues as follows. Section 2 describes the data we use and how we construct our sample. Section 3 discusses how we measure contact between students. We discuss identification issues, how we overcome them, and the model we use in Section 4. Section 5 presents our baseline estimates. Section 6 shows that our estimates are robust to several additional specifications and robustness checks. Section 7 concludes.

2 Data and Sample Construction

2.1 Data

Data used in this paper are student level and come from the New York City Department of Education (NYCDOE) administrative database for the 2018 academic year, with lag outcomes coming from the 2017 academic year. We observe student characteristics such as sex, race, grade, zip code, poverty status, and whether a student is an English language learner. We also observe homeroom classroom assignment and test score outcomes for reading and math. For elementary school students, homeroom class assignment indicates the student's primary classroom. In addition, we observe student lunch transactions at the point of sale (POS). The data indicate the exact timing of lunch purchase transactions for students (to the second) for every day during the school year. We use this transaction information to observe the order of students in the lunch line and measure social connections in the classroom by observing which students are often in

9. As described in Manski (1993), and in section 4.1.

proximity to one another in the lunch line.

2.2 The point of sale system

The New York City Department of Education (NYCDOE) began implementing a point of sale (POS) system in their school cafeterias in 2010 . By academic year 2018, 88.0% of schools had the system installed at the start of the school year. These schools served 90.2% percent of the over one million students in the school district. Implementation started in large schools first, with a focus on middle and high schools where the district felt these systems would do the most good. However, by academic year 2018 the system was in 93.4% of elementary schools, and these schools served 95.5% of grade 1-5 students.

Table 1 shows that the makeup of the schools with POS systems is slightly different than the full school district. Students in schools with a POS system are more likely to be Hispanic or Asian/other and less likely to be black. This is likely because schools with POS systems are more likely to be located in Staten Island and Queens, and less likely to be in the Bronx or Brooklyn. These differences, while statistically significant, are not large.

The primary way students interact with the POS system is either by entering a PIN in a keypad or a cafeteria worker uses a list of names and faces to enter the transaction as students move through the line. This is not standardized over the district, can vary by school, and is not observed.

2.3 Sample

The sample is taken from the universe of students in the NYC public schools for the year 2018. We examine elementary school students for two main reasons. First, homeroom assignment corresponds to the student's primary classroom in elementary school. Second, elementary school students are more likely to participate in the school lunch program than middle or high school studentsbecause they have less autonomy and do not have the same outside options as older students who may be allowed to go off campus during lunch. School lunch participation for our sample is 65.8%. Additionally, we limit our analysis to students in general education classrooms.

Table 1 illustrates our sample selection process. We begin with all general education fourth and fifth grade students in schools with a POS system in place for the entire academic year. The POS system is important because we measure social connection through repeated observation of lunch transactions, such that students who are observed together frequently are considered friends. We restrict to fourth and fifth grade students because standardized tests begin in the third grade, and we include a lag test score in our model. We are unable to measure connection to students who never participate in school lunch, and 3.7% of students fall into this category. We lose 7.5% of students because they are missing either a current year score or a lag test score for both math and reading. Some students do not participate in standardized tests, so we lose another 2.3% of students from test non-participation in both math and reading. We lose less than half a percent of students from the following three reasons. First, we exclude lunch transactions occurring before 10am and after 2pm. Transactions occurring outside this window are rare and may be improperly coded breakfast transactions, transactions entered after the fact (such that timing is not indicative of the lunch line order), or simply an unreasonable assigned lunch time.[10] Second, we remove transactions occurring more than an hour earlier or later than the mean transaction time for a classroom. These students appear to be "out of line", and as a result are not relevant for determining who is next to whom in the lunch line. Including them would simply add noise to our estimates, so we remove them. Third, we remove transactions which occur simultaneously for the entire classroom. This is indicative of an unusual event, such as a field trip, and gives no information relevant to the lunch line order. The final exclusion we make is excluding students in classrooms with less than 20 students, resulting in the largest loss of students (15.78%). We choose to look only at classrooms that are larger than twenty students because we are concerned classrooms that appear smaller may be integrated co-teaching (ICT) classrooms, and we do not want unobserved peers.[11]

Table 3 gives some summary statistics regarding our sample. Our sample includes fewer black students and more Hispanic and Asian/other students. This is likely the result of where the POS systems have been implemented, as the Bronx and Brooklyn are underrepresented while Queens and Staten Island are overrepresented in locations having received POS systems. Because implementation has occurred in a large proportion of schools, discrepancies are small. Test scores are normalized z scores across grade level in the school district, so our sample is slightly higher performing than average. Average class size in our sample is 25.6 students, and the lunch participation rate is 65.8%. Math and Reading scores are z-scores standardized to zero for the entire NYC public school student population.

3 Network Construction

3.1 Defining Social Distance

This paper uses a novel approach to measure contact between students and reveal the classroom friendship network. We observe the timing of every lunch transaction in the POS system every day during the school

10. Some lunch times are even more unreasonable than these bounds we place on lunch times, as in Brand's (2019) article "Why do some NYC school kids still eat lunch before some of us have had breakfast?" However, times like these are even more of an anomaly for elementary students than the high schoolers discussed in the article.

11. ICT classrooms combine general education students and students with disabilities together. Students learn from the general education curriculum and are taught by a team of two teachers: one general education teacher and one special education teacher. ICT classrooms typically have a ratio of 40% students with disabilities and 60% general education students.

year, and this timing is precise to the second. This allows us to observe the lunch line both in terms of physical order and in the timing of movement through the line. We use this information to construct a peer network, but first discuss how to extract a meaningful social distance from this information. There are two ways we might consider using this information. The first is to use the actual timing of the transactions to measure distance between students. However, this is not our preferred method, as we believe this to be a noisy estimate of social proximity between students. We discuss this measure further in section 6.4.

Our preferred method is to transform the near-continuous timing data into ordinal data. This allows us to think about distance as the physical proximity of students to one another rather than temporal proximity. We argue that because lunch is a relatively unstructured and social time, students' primary concern is who they are able to socialize with in the line and then during lunch. The simplest way to transform the observed order into a social distance is to look at whether any two students i and j are within some threshold distance (number of students) of one another. Our baseline model uses a threshold distance of one - whether two students are next to one another in line. For robustness, we also look at larger threshold distances in Section 6.2.

It is worth discussing the implication of the observed lunch line order, as the ordering process is a black box to us as researchers, and the method of ordering likely varies by classroom. We discuss some possibilities for how students are ordered, fitting them into three categories: students have agency over their choice of line position, students are ordered by someone else (such as the teacher), and students have agency within a constraint. We then provide evidence that in the majority of classrooms, students either have at least some agency over their position in line or the order they are given changes frequently.

First we discuss situations in which students have agency over their position in line. Students must balance a choice between being in line with their friends and their preference for being towards the front or back of the line. For most students, we believe the choice of being in line with friends is more important than their line position. If this is true, then it is clear that the line order contains information relevant to the social network in the classroom. However, it is possible that many students' preference for being at the front of the line (for example) dominates their desire to be near friends. A classroom in which all students wish to be first would see a race to the front of the line. Thus line order depends upon classroom geography and where students sit in relation to the door (start of the line), with students sitting near one another tending to line up near one another. If students sitting near one another are more likely to talk to one another or work together during class time, then this gives us another reason physical proximity in the lunch line would be socially important. In both of these situations, students who are near one another in the lunch line would be expected to be more influential in one another's social network - at least as it relates to academics within the classroom - than a randomly selected classmate. The truth is likely some combination of these two situations.

For students geographically near the door, they have the option to be first or wait for their friends. Those further from the door do not have this choice. Thus in a classroom in which all students wish to be first, the benefits to rushing decrease in distance from the door. A tipping point could occur at which point students switch from racing to the front of the line to waiting for their friends.

Second, it is possible that students could be ordered by their teacher according to some metric - perhaps alphabetical. We do not observe names, and so we cannot test this hypothesis directly (although we do look at how much strict ordering exists in our sample). If students are ordered based on name, we expect little reason for these students to be socially more important than other students.[12] The teacher could order students by some other method - perhaps according to student characteristics (demographic, performance, or behavioral). If we believe that students with similar characteristics are more likely to be friends with one another (homophily), then observing similar characteristics in students near one another may be indistinguishable from an external ordering placed upon the students according to this same set of criteria. These students may also be more socially important to one another than a randomly selected classmate, as W. Horrace et al. (2019) shows.

Finally, there is the possibility that students sort into the lunch line based on some combination of autonomy and rules. For example, the teacher may dismiss students from their classroom tables, so that students form a line within a subset of the classroom - they have autonomy within a constraint. Students face a similar decision whether to line up next to friends (within the constraint group) or in terms of optimal position. Notice that both physical and social positioning are constrained, as a student with preference for the front of the line may not have a choice over line position until the first half of the line is filled. Similar to when students have full autonomy, line order likely reflects some level of student importance - either through selecting friend groups or the importance of the constraint group (such as classroom geography). The result is similar to that of full autonomy, but the effect of these peers is likely smaller than under full autonomy, as this is a group of "next-best" friends.

While the line-up process is itself unobserved, we provide evidence that students have at least some agency over this decision by considering whether students are ordered into roughly the same order each day. To do this, we construct a measure of within-classroom noise as detailed in Appendix B. The measure M is based on the number of order inversions (swaps in the order of students i and j) observed in the order over the year, and it is normalized such that it is invariant to classroom size and participation rate. Figure 2 shows the distribution of that measure and that the bulk of classrooms (average measure value is 0.203) are closer

12. Outside the notion that students of similar cultural or ethnic backgrounds might have similar names and thereby be grouped together. While some work looks at the ability to predict ethnicity based on names, such as Elliott et al. (2009) and Ryan et al. (2012), the success of these algorithms is still limited. Predicting ethnicity based on alphabetical ranking within an average group size of around 25.6 would be unsuccessful.

to a uniformly random distribution (value of 0.25) than fully ordered (value of zero), but that there is more order than complete randomness.[13] This is consistent with the idea that students in most classrooms have agency over their position in line, and choose positions in ways that are varied but less than random (ex: in order to be with their friends). It is important to note that the distribution of this measure has a small tail with what may be considered abnormally low noise (where we may think classrooms are ordered). If we let 0.1 be the threshold below which classrooms are ordered, about 3% of classrooms are ordered.[14] In Section 6.3 we remove these as a robustness check.

3.2 Scaling from daily observations to the friendship network

We observe daily lunch transaction timings over the entire school year, which we translate into the lunch line order for each day. The next step is to zoom out to the full year, such that students observed in frequent close proximity to one another on individual days are considered friends. Because some students do not participate in school lunch every day (or are absent from school), this process is less straightforward than we might like. On a day that a student does not attend school, we miss their signal of who they would choose to stand in line next to on that day, and they also limit the choice set of the students who remain (by removing themselves from the candidate pool).

We start to think about constructing the network by averaging daily observations together, akin to what De Giorgi et al. (2010) do with classes. This proximity matrix gives us the percent of days each student is near each other student. We may be concerned that students with low participation will appear to have artificially low connections measured by this proximity matrix. There are two ways we can address this. First, we can simply row-normalize the average proximity matrix such that each row sums to one. Row normalization is common in the literature to transform a proximity matrix to the weighting matrix used in estimation, because it improves interpretability of results by appropriately weighting influential peers for the given student such that we have the weighted average (characteristics or outcome) of the peer group. Each row i indicates student i's relevant peer group, appropriately weighted. We plan to row normalize for the interpretation benefits, but it is important to notice that row-normalization changes the interpretation of the proximity matrix from the percent of school days both students are near one another to be the percent of days student i is present that student i was near every other student j. For student i with low participation, this moves their average connections with students from near zero to the percentage of times i participated and was near each other student j. By increasing the weight on the days a student does participate, we have addressed the issue of not observing who a student would choose to be near if they did attend. However, it

13. Appendix B outlines how the measure behaves under changes in class size, participation rate, and levels of randomness.
14. 134 of 4,077 classrooms are below the 0.1 threshold.

is not clear that we have addressed the second problem in which student i is removed from the choice set of other students.

In order to address this second concern, we construct a proximity matrix for the percent of times we observe students near one another when both are present:

$$p_{ij} = \frac{\sum_{d=1}^{D} S_d(i,j)}{\sum_{d=1}^{D} \delta_d(i,j)} \text{ for } i \neq j \text{ ; and } p_{ij} = 0 \text{ for } i = j \tag{3.1}$$

where $S_d(i,j)$ indicates that students i and j are next to one another on day d and $\delta_d(i,j)$ indicates that both i and j are participating in lunch on day d. The proximity matrix is then $P = \{p_{ij}\}$. As mentioned, for estimation we create our weighting matrix W by row-normalizing the proximity matrix such that each row sums to one. While averaging the daily observations as done in De Giorgi et al. (2010) and the method described in equation 3.1 lead to different proximity matrices, row normalization makes the resulting weighting matrices identical. Figure 1 provides a simple example to illustrate how we convert daily observations into a proximity matrix (according to equation 3.1) and corresponding weighting matrix. In the example, we observe five students over six days, and one student is absent or not participating each day. Figure 1a shows the daily proximity matrix for each individual, where a dark square illustrates connection and a white square represents no connection. Notice that students who are at the front or the back of the line have only one connection, while every other student has two. Part 1b applies equation 3.1 to the daily observations, calculating the percent of days both students are present for which they are next to one another. We now have a continuum of connection strengths and the darker the square the stronger the connection. Figure 1c shows the result of row normalization. Graphically, it appears that row normalization has dampened the effect, but this is not the case. Instead, it proportionally reweights the proximity matrix so that each row, when multiplied by Y or X, creates a weighted average of the relevant peer outcome or characteristics.

Admittedly there are other ways we could construct the proximity matrix, and there are potential concerns with the way we have constructed ours. Perhaps most concerning is that low-participation students could appear overly important for those they stand near when they do participate. We address this by looking at an alternate proximity matrices for robustness in Section 6.

It is also important to distinguish between absence and non-participation. The previous discussion dealt with absence from school, but non-participation adds an additional complication. The majority of non-participation in elementary school lunch is because students brought their own lunch from home. Thus if a student is present at school, but not participating in lunch, they are likely present in the lunch line - at least during travel from the classroom to the cafeteria - and importantly they are part of the decision process when students decide where to stand in line. Thus two students we observe as being next to one another may actually have another student between them (or more than one) during the decision process for who to stand near. This is an issue of truncated data, and likely a significant source of noise in the model. The result is that an observed distance of one between two students is actually a distance of at least one. This means connections we observe are weaker than actual connections in the classroom, and results obtained from this data are likely a lower bound on the peer effect from lunch-mates.

4 Methods

4.1 Identification

Identification of peer effects is notoriously difficult, and in this section we discuss some of the common issues and how we address them in this paper.

In his seminal paper, Manski (1993) discusses the different effects which may be captured in a naive model of peer effects. The first is the endogenous effect, which is often the effect of interest to researchers and policymakers. This is the effect of one individual's performance on the performance of another. For example, we observe an endogenous effect for two students working together on a group project if the performance of one student varies based on skill level (performance) of that student's partner. This is of interest to policymakers because the endogenous effect is a multiplier, causing spillovers to other students. If an intervention is applied to some students and improves their performance, all students will benefit because of the dependence of all students' performance on that of their peers. The endogenous effect is named because it directly places the outcome on the right hand side of our model. Use of the linear in means model structure (assuming the peer effect is a weighted average of peer performance) and maximum likelihood estimation allows us to solve for this endogeneity in our results. More details about the model follow in Section 4.2 and about the estimation procedure in Appendix A. The second effect is the exogenous effect, sometimes called a contextual effect. Exogenous effects control for student characteristics in the peer group. For example, we might expect that wealthier students perform better on tests, all else equal. As a result classroom performance may increase with wealth, but this is due to characteristics of the student rather than interactions with them. The final effect Manski discusses is the correlated effect. This is often not a social effect at all, but is related to common exposure by students to the same treatment. For example, a lack of adequate facilities or a good teacher are felt by all students in the classroom, but they are not related to the students or any interaction with them.

In many reduced form models of peer effects, we cannot distinguish between exogenous and endogenous effects because the performance of the reference group is collinear with the characteristics of this group. This is known as the reflection problem (Manski 1993). However, when individuals have unique reference groups, this is sufficient to separately identify endogenous and exogenous effects. This is because the collinearity issue arises when individuals share a reference group, but when this does not exist, there is no collinearity issue to worry about here. The individual level reference groups arise because each student is next to a unique set of students each day (another student will likely be next to one of the students, but there cannot be more than one student next to both students). We may be concerned that averaging over the days could cause different students to have the same reference group. However, with 180 days and differing levels of participation among students this does not occur in our sample.

Correlated effects are commonly addressed using fixed effects (ex: Ajilore et al. 2014, Bifulco et al. 2011, W. C. Horrace et al. 2016, Lin 2015), and we follow this trend with the inclusion of classroom fixed effects. Intuitively, we can think of this as controlling for a teacher effect, although it also controls for other group treatments such as quality of the built environment, scheduled lunch time, and principle quality to name a few.

Selection can be a problem if students are sorted into reference groups based on shared characteristics. Because we are looking at the revealed friendship network within a classroom, we are not concerned with selection into the within-classroom network - this is in fact what we are interested in measuring. A potential concern is if students are sorted into classrooms based on shared characteristics such that the strength of the social spillovers within a classroom are correlated with these characteristics. We test whether student characteristics explain classroom assignment and show that class assignment based on observed student characteristics is consistent with randomness in Appendix C.

4.2 Baseline Model

This paper uses a revealed friendship network to measure academic spillovers in the classroom. For our baseline model, we construct a within-classroom network according to equation 3.1. In the linear in means model we use, this network is multiplied by both the outcome Y and student characteristics X so that we can separately identify endogenous and exogenous effects. Below is the basic format of the linear in means model we estimate:

$$Y = \alpha + WY + WX + X\beta + \theta + U \tag{4.1}$$

where Y is the outcome of interest, α is a constant, W is the weighting matrix as defined at the end of section 3.2, θ is the classroom fixed effect, β is the estimate of own characteristics X, and U is the error term. Controls in X include lag test scores and indicators of sex, ethnicity, zip code, English language learning, and poverty status.

Modeling exogenous effects allows us to control for the characteristics of students in the reference group, thereby isolating the endogenous effect of interest - the effect of one student's performance on another's. The endogenous effect is important to distinguish from the exogenous effect because it captures the spillover effects resulting from social interaction, whereas the exogenous component controls for student characteristics. References to estimates of the peer effect refer to this endogenous effect.

We estimate our model using Maximum Likelihood Estimation (MLE) and follow W. Horrace et al. (2019) and Lee and Yu (2010). Details of the estimation procedure are found in Appendix A.

4.3 Interpretation

It is important to note that estimates of the endogenous effect from model 4.1 are multiplier effects. This means that interpretation of the estimated structural parameter $\hat{\lambda}$ is done by converting the result as below:

$$\hat{\gamma} = \frac{1}{1-\hat{\lambda}} \tag{4.2}$$

Thus an estimate of $\hat{\lambda} = 0.05$ is interpreted as a multiplier of 1.053. This means that a ten percent improvement in test scores for a student's reference group results in a 0.53 percent improvement in the student's own test score. Notice that for small λ, the multiplier γ is comparable in magnitude.

5 Results

5.1 Baseline Results

Table 4 shows our baseline results for math and reading scores of fourth and fifth graders using a proximity measure in which students are next to one another in the lunch line. Our outcome of interest is test scores, and these are z-scores normalized citywide among students in the same grade. In addition to the controls shown, the model also includes fixed effects for zip code of residence. We also include these zip codes in the exogenous effect.The first line of Table 4 shows a math peer effect for students in the lunch line together of 0.089, which is statistically significant. As discussed in section 4.3 this is a multiplier effect, and so we interpret this as a multiplier of 1.098 or an increase of 0.098 units. The endogenous effect for reading is also significant, but smaller at 0.053, or a multiplier of 1.056. The fact that both estimates of the endogenous effect λ are positive is consistent with our intuition and the general findings of the literature, which is that improvements in the reference group should lead to improvements in own outcomes. We can interpret these results by saying that if a student's relevant peers exogenously improve their performance by one standard deviation, we expect to see improvements in own performance in math by 9.8% of a standard deviation. This is equivalent to the black-white test score gap in math. The gap is larger and the spillover effect is smaller in reading, so the equivalent improvement is equivalent in magnitude to about 40% of this gap.

It is difficult to associate meaning to a comparison of these estimates to static estimates because they are multipliers and therefore amplify all other elements of the education production function. We can think about the interpretation of these multiplier estimates when combined with additional external information and compute an average effect. The average classroom in our data has substantial variation in student ability, which we see manifest itself in student performance. If we collect the top performer in all classrooms, we find that the average classroom has a student performing 1.5 standard deviations above the mean.[15] This is mirrored in low performers.[16] We then collect the strongest connection we observe in each classroom, which we can think of as a student's best friend. The average student's largest connection is 0.357, meaning that over one third of the time we see both students present, they are next to one another in the lunch line. When we conduct our row normalization, the meaning is preserved, but the value of the matrix cell for the strongest connection reduces to 0.202. This means that the benefit to a student of connecting with the best student in the classroom, rather than an average student, is 0.030 in math and 0.018 in reading. This is equivalent to half the effect of poverty and nearly one third of the black-white test score gap in math, and it stems from only one peer connection. In reading the effects are smaller than math, being one third the effect of poverty and 14% of the black-white test score gap (which is larger in reading).

The fact that the spillover is larger in math than reading is consistent with the idea that students learn verbal and reading skills at home, but primarily learn math in school. We see stronger in-school math effects than reading effects in Nye et al. (2004) which shows that teachers have a greater impact on math scores than reading scores.

Notice that the controls are performing as expected. Own student lag scores are highly significant and important. Male students perform slightly better in math but worse in reading than their female peers. English language learners and poor students do worse than native speakers and students who are not poor. The comparison group for ethnicity is Hispanic students, because these are the modal student in NYC public schools, and whites and Asians do better than them, while blacks do worse. Most of the exogenous effects are not statistically important, with the exception of friends in the Asian/other group, which has a large positive impact. Taking the math estimate, this means having all friends in the Asian/other group improves own math performance by 0.12 standard deviations as opposed to having Hispanic friends (the baseline reference group), all else equal. Notice that the exogenous effect is not a multiplier effect, but simply shows the effect

15. In math, the average top performer across all classrooms scores 1.52 standard deviations better than the mean. For reading the average top performer scores 1.55 standard deviations above the mean.

16. The average bottom performer across all classrooms scores 1.39 standard deviations worse than the mean in math, and 1.51 standard deviations worse in reading

of having friends from this group type. Surprisingly, the previous performance of friends does not appear to matter in math, but it is quite important in reading.

5.2 Evolution of the friendship network over time

The friendship network is constructed over repeated observations of the lunch line. As a result, we can explore different portions of the school year to see whether the strength of the friendship network varies over the year. It is ambiguous whether connections should be more important at the start or at the end of the year. Patacchini et al. (2017) provides evidence that students who are friends for longer periods of time are more important than short term friends, so we might expect that connections at the start of the year are more influential. On the other hand, students are still getting to know one another at the start of the year, so we may observe more noise as students sort into friendships. Additionally, testing occurs towards the end of the school year, so we might expect connections closer to the test date are most important.

Table 10 divides the year into halves, thirds, and quarters to compare friendship importance over time. When dividing the year into halves, we construct two proximity matrices according to equation 3.1. The first proximity matrix uses the set of days from the first half of the year, and the second matrix uses the days from the second half of the year. When dividing the year into three and four components, each matrix is constructed from the corresponding set of lunch line observations (days).

The first portion of the year seems to be most important, suggesting that friendships formed early and in place the longest are most important - even within the shorter time horizon of a single school year. This could be the result of carryover from who students know the previous year (so that they really are long-term friends), or they could be the result of newly formed friendships. Students do not have agency over their classroom assignment, and we show in Appendix C that class assignment is consistent with randomness along observable characteristics. While some connections will carry over from the previous year, there is no reason to think the strongest connections carry over, outside those that randomly get placed in the same classroom. This is also consistent with a story in which students either continue or begin long-term relationships with a small group of students and then try to branch out over the course of the year.[17]

5.3 Discussion

To our knowledge, Lin (2010) is the closest study to our own in terms of methodology, so we compare estimates. Lin uses the Add Health data to estimate peer effects on GPA using a similar spatial autoregressive model with maximum likelihood estimation. It is important to note that the sample in Lin (2010) is older

17. Future versions of this paper will further explore selection.

than our own sample, because the youngest students in the Add Health survey are in the seventh grade. Students nominate up to five other students of each gender from their school as friends. Thus the network structure involves connections between each student and up to ten or twenty students in the school depending upon whether connections need to be reciprocated. Each of these students is equally weighted, because the researcher does not observe the relative importance of these friends. Lin (2010) estimates the magnitude of peer effects such that they improve GPA by 7.85% of the mean GPA. This is comparable with our own results, in which math is above and reading below this estimate.

Our approach is to appropriately weight classmates based on the revealed friendship network. This should reduce the importance of students who are not friends, and increase the importance of students who are, thereby better targeting important peers. If friends are more important than a random peer, we should expect our results to be larger than those who give all friends equal weight. There are three potential reasons we do not see this stronger effect, which we will address one by one.

First, as touched upon previously, the papers look at different contexts. Social spillovers may be smaller for younger students than for older ones. Table 5 shows the results for our model when we separately run the model for each grade. The estimates for social spillovers in reading are statistically indistinguishable from one another when we compare fourth graders and fifth graders. However, the math spillover effect is much stronger in the fifth grade than in the fourth grade, which would be consistent with the conjecture that social spillovers increase with age. Thus differences in age may explain why we do not find stronger effects than an unweighted social network as in Lin (2010).

Second, while the models are similar, the data are different and dictate different models. Our data allow us to look within the classroom, and so include classroom (teacher) fixed effects. Add Health only allows network (school) fixed effects to be used. If friends share classes, we might expect them to do better or worse together due to teacher effects which cannot be controlled for in the Add Health data. A difficulty with Add Health is that we do not know if students are friends because they take classes together, or choose to take classes together because they are friends. Additionally, students are unrestricted in their choice of friends in Add Health. Thus students can nominate students outside their classes who are important and the choice set for important peers is less restrictive than nominations within a classroom. However, this last confounder is unlikely to be a large source of bias, as work such as Burke and Sass (2013) show that classmates are much more important than those outside the classroom.

Third, we could be incorrect in our supposition that friends have differing levels of importance. It may be that what really matters is the extensive margin - whether students are friends, not the intensive margin of how close the friends are. Using the Add Health data, Patacchini et al. (2017) find that students who are friends in both waves have stronger effects on one another than those who are friends in only one wave (although this paper does not look at academic outcomes). This suggests that there is indeed a hierarchy in friendships, even in the connections laid out in the Add Health network. In our data, we cannot distinguish between who is and who is not a friend outside our weighted framework. We can test whether our model performs better than randomly selected peers, and we discuss this in greater detail in section 6.1. We find no evidence of a peer effect when we randomly assign line orders, suggesting that this third option is not correct. We conclude that the reason our estimates are not larger than Lin (2010) must be due to a combination of context and model specification.

Our results provide evidence of strong spillover effects using a revealed friendship network. These results are on par with estimates found using a similar method. Our estimates are smaller than some reduced form estimates which do not attempt to disentangle endogenous and exogenous effects, such as those in Hoxby (2000). We are consistent with Hoxby (2000) in that we find stronger effects in math than in reading. Our results are larger than those Burke and Sass (2013) find for elementary school students using student and teacher fixed effects in combination with mean peer achievement. We can attribute this difference to a combination of methodology and definition of the reference group.

6 Robustness Checks

6.1 Random Lunch Lines

By construction, students in each of the networks we construct share a classroom, so we might expect that they are socially important to one another regardless of proximity in the lunch line. To test whether the spillover we estimate is simply a result of the students sharing a classroom, we randomly shuffle the lunch line for each day of the year and re-estimate the model. Results are found in Table 6. The placebo estimate for math is a statistically insignificant 0.004 and for reading it is also insignificant at 0.016. While the endogenous effect (as well as all the components of the exogenous effects) are insignificant, the own effects perform similarly to the baseline model. We conclude that students in close proximity to one another in the lunch line are socially more important to one another than a randomly selected classmate.

6.2 Alternate Distances in the Lunch Line

When constructing our baseline network, we chose to connect students who are next to one another in line. If friendship groups are larger than pairs, a larger distance may be appropriate. For example, if we observe friends A, B, and C in line together, we miss the connection between A and C if we restrict our analysis to students who are next to one another. We can increase our distance from one to test whether this is the appropriate group. Table 8 reports estimates when including multiple networks for each additional distance between two and six. While the inclusion of additional networks slightly dampens the effect from a distance of one, the results remain robust to the inclusion of these networks. Additionally, no higher distance is statistically significant. This indicates that students being next to one another in line is the strongest signal of connection we can measure, and while students who are further apart in line may also be friends, this signal is too noisy to be meaningful. Given the large number of observations throughout the year, we are able to observe a spectrum of connection strengths. For example, we may observe the same group of friends C, A, and B the next day. With only these two observations, we measure a stronger connection between A and B than either has with C. This is because we always observe both A and B next to one another.

It is possible to estimate the model with a single network, using higher distances (rather than putting each distance into its own separate network). Table 9 shows the results of this network specification for the same set of distances. The effects appear to be stronger than our baseline model. Notice that the point estimates in math increase and then decline after a distance of four. For reading the point estimates continue to rise in each specification. Standard errors are increasing for both outcomes, indicating additional noise from the inclusion of students who do not matter. We expect that the cause is scenarios like in the toy example with students A, B, and C discussed above. Students who are friends are likely near one another in line, even if they are not next to one another. What these models pick up are larger friend groups. While Table 8 shows that no other distance is important on its own, these estimates pick up the extent to which larger friendship groups (or possibly friends of friends) are important. We provide this as suggestive evidence that a broader set of students matter for performance in reading than in math.

6.3 Removing potentially ordered classrooms

When we measure variation in the line order, not all classrooms appear to give students agency over their location in line. We introduce a measure M of within-classroom noise in section 3.1, which is discussed in greater detail in appendix B.

Table 7 shows the results of removing classrooms which exhibit small levels of variation in the observed line order. We remove classrooms with M less than 0.05, 0.1, and 0.15. These values signify low levels of variation in lunch line order throughout the year, which may be attributable to students having no agency over their position in line. Appendix B further discusses properties and behaviors of the measure M. In each case, the peer effect increases in M.

We provide this as evidence that the mechanism is friendship rather than the effect of time spent in the line. In classrooms with very stable line orders, students spend time in the lunch line with their neighbors more consistently than in classrooms with more variation. If the spillover mechanism was due to this time, we would expect a decrease in estimates when these classrooms are removed. We see the opposite, suggesting that connections we observe in classrooms with more autonomy are more meaningful. In addition to suggesting that friendship is the mechanism through which these spillovers are working, this suggests that peer effect estimates in my baseline model may be biased towards zero because of the additional noise added to the model by these ordered classrooms.

6.4 Alternate Proximity Matrix Specifications

For our main network specification, we used the order in which we observe students go through the lunch line. Another option is to use the actual timing of transactions as a measure of distance, where students further apart socially will go through the lunch line further apart from one another. In order to estimate the model, we need to construct a proximity matrix similar to the ordinal version discussed previously. To do this, we average the time between each student in the classroom on each day both students participate in the lunch line. This creates a distance matrix, which we convert into a proximity matrix by taking the reciprocal of each entry. As before, we row-normalize this proximity matrix.

Table 13 shows the results of using time as a measure of distance. The biggest contrast from the baseline model is that spillovers in math are statistically insignificant and smaller than those for reading (which are statistically significant). The point estimate in reading is smaller than my baseline estimate, but it is not statistically different. For both models, other controls behave similarly for the most part. The biggest exception is the exogenous effect of the lag test score in reading, which is negative and significant at the 5% level, whereas it is positive in the baseline reading model model. Additionally, racial exogenous effects are less significant.

Precisely what is being measured by a temporal distance is more difficult to pin down relative to the ordinal system discussed previously. As a result, this is not our preferred method. There are a number of ways this measure introduces unnecessary noise and uncertainty into our measure. For example, it is unclear what meaning to ascribe to a 10 second pause between adjacent transactions relative to a minute pause between them. Does the longer time signify a social distance - where students who are closer socially try to stick together temporally as well, perhaps waiting for one another and going through the POS system in quick succession? Or is it likely that the person operating the POS system had technical difficulties or was distracted by something occurring elsewhere in the lunch line or kitchen? Similarly, if there is a set of three transactions in relatively quick succession, does this indicate a friend group, or does this indicate students (or a cashier) who is relatively more adept at navigating the lunch line? We could imagine that some students are slower than others, and consistently have larger gaps in time before or after them. Does this slowness somehow make that student less likable? The temporal distance between students could just as easily be due to indecision in the lunch line, ease of distraction, chattiness with cafeteria workers, and the like. Additionally, does it make sense to ascribe the same weight to all connections with a 10 second gap between them, even if there is another transaction between them? These are some of the questions raised by using a temporal measure of social distance, and sufficient reason for us to prefer an ordinal or spatial measure of social distance.

There are other ways we could define proximity, each with positives and negatives. In this section, we examine an alternative method for defining proximity which addresses the concern that low attendance students may have outsized effects on those near them. Section 3.2 describes our method for constructing the proximity matrices. Recall that the two main difficulties caused by student non-participation are a missed signal of who they choose to be near and removal of the absent student from the choice set of other students. The definition of proximity proposed here builds upon our baseline measure. Our previous definition of proximity measures the percent of days two students participate in lunch during which they choose to be near one another. To address the problem posed by low-participation students, we can require students to participate a certain number of days together before their connection can be evaluated. The proximity matrix in equation (3.1) is altered such that:

$$p_{ij} = \begin{cases} \dfrac{\sum_{d=1}^{D} S_d(i,j)}{\sum_{d=1}^{D} \delta_d(i,j)}, & \text{if } i \neq j \text{ and } \displaystyle\sum_{d=1}^{D} \delta_d(i,j) \geq \eta \\ 0, & \text{otherwise} \end{cases} \tag{6.1}$$

where η is the threshold number of days students must both participate in before we count their connection. Connections between students when one or both of them are low participation students are reduced to zero unless we see enough participation from both students on the same days. The potential drawback of this method is that we lose all or most connections with low-participation classmates and may be throwing away useful signals. The benefit is that these signals we lose may be noisy.

Tables 11 and 12, we explore thresholds of between two and ten days that students must participate on the same day before we evaluate their connection. Our results remain relatively unchanged in each specification. We conclude that our baseline model is robust to the concern that low participation students may appear overly important.

7 Conclusion

This paper measures contact between students in the school lunch line as an indicator of connection strength between students. This allows us to define not only the scope of the reference group, but measure the intensity of the relationships within that environment. We use the revealed friendship network to separately identify endogenous and exogenous effects using a linear in means model.

Our results indicate significant social spillovers in both math and reading. These effects are stronger in math than in reading. In math, a one standard deviation improvement in peer performance results in an increase in own performance between 7.5% and 11.1% of a standard deviation. This is a significant effect, on par with our estimates of the performance gap between black and white students. Social spillovers have a multiplier effect, magnifying other inputs in the education production function. This suggests an alternative interpretation, which is that for a given intervention that improves math scores, between 7.5% and 11.1% of improvements occur through the peer effects mechanism. Spillovers are lower in reading, where an increase in peer performance of one standard deviation improves own performance by between 4.1% and 6.3% of a standard deviation.

Our measure of connection is constructed using daily observations of student contact in the lunch line. The daily nature of this data allows us to look at the evolution of these connections over time. We find evidence that connections formed at the beginning of the year are most important, which is consistent with the findings in Patacchini et al. (2017) that long-term friends are more influential than short-term friends.

There are situations in which we want to measure average peer effects within a context, such as when measuring the average effect of exposure to a specific type of student. For example, it is useful to know the spillover effects for disruptive peers, as in Carrell et al. (2018). However, when it is important to understand the strength of the overall peer effect within the classroom or to understand mechanisms through which policies may be working, it is important to take into account the connection complexities found in the school social network.

Peer effects are believed to be important for a number of important policy discussions such as tracking, school choice, and school integration. In order to effectively weight the costs and benefits of these policies, accurate estimates of the relevant social spillovers are imperative. Estimates that rely on the average peer effect within a reference group may understate the overall effect by not appropriately weighting the most relevant peers for each student. In this paper, we measure the strength of connection between elementary school students sharing a classroom. Understanding the relative importance of peers within the network accurately weights which peers are important for each individual and provides stronger estimates of the peer effect.

References

Ajilore, Olugbenga, Aliaksandr Amialchuk, Wei Xiong, and Xinyue Ye. 2014. "Uncovering peer effects mechanisms with weight outcomes using spatial econometrics." *The Social Science Journal* 51 (4): 645–651. ISSN: 0362-3319. https://doi.org/https://doi.org/10.1016/j.soscij.2014.07.008. http://www.sciencedirect.com/science/article/pii/S0362331914000834.

Bifulco, Robert, Jason M. Fletcher, and Stephen L. Ross. 2011. "The Effect of Classmate Characteristics on Post-secondary Outcomes: Evidence from the Add Health." *American Economic Journal: Economic Policy* 3, no. 1 (February): 25–53. https://doi.org/10.1257/pol.3.1.25. http://www.aeaweb.org/articles?id=10.1257/pol.3.1.25.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of peer effects through social networks." *Journal of Econometrics* 150 (1): 41–55. ISSN: 0304-4076. https://doi.org/https://doi.org/10.1016/j.jeconom.2008.12.021. http://www.sciencedirect.com/science/article/pii/S0304407609000335.

Burke, Mary A., and Tim R. Sass. 2013. "Classroom Peer Effects and Student Achievement." *Journal of Labor Economics* 31 (1): 51–82. https://doi.org/10.1086/666653. eprint: https://doi.org/10.1086/666653. https://doi.org/10.1086/666653.

Carrell, Scott E., Mark Hoekstra, and Elira Kuka. 2018. "The Long-Run Effects of Disruptive Peers." *American Economic Review* 108, no. 11 (November): 3377–3415. https://doi.org/10.1257/aer.20160763. http://www.aeaweb.org/articles?id=10.1257/aer.20160763.

Coleman, J.S., E.Q. Campbell, C. Hobson, J. McPartland, A. Mood, F.D. Weinfield, and R. York. 1966. *Equality of Educational Opportunity.* Technical report OE-38001. U.S. Government Printing Office, Washington, DC: National Center for Educational Statistics.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli. 2010. "Identification of Social Interactions through Partially Overlapping Peer Groups." *American Economic Journal: Applied Economics* 2, no. 2 (April): 241–75. https://doi.org/10.1257/app.2.2.241. http://www.aeaweb.org/articles?id=10.1257/app.2.2.241.

Elliott, Marc N., Peter A. Morrison, Allen Fremont, Daniel F. McCaffrey, Philip Pantoja, and Nicole Lurie. 2009. "Using the Census Bureau's surname list to improve estimates of race/ethnicity and associated disparities." *Health Services and Outcomes Research Methodology* 9, no. 2 (April): 69. ISSN: 1572-9400. https://doi.org/10.1007/s10742-009-0047-1. https://doi.org/10.1007/s10742-009-0047-1.

Horrace, William, Hyunseok Jung, Jonathan Presler, and Amy Ellen Schwartz. 2019. "What makes a classmate a peer? Examining which peers matter in NYC."

Horrace, William C., Xiaodong Liu, and Eleonora Patacchini. 2016. "Endogenous network production functions with selectivity." Endogeneity Problems in Econometrics, *Journal of Econometrics* 190 (2): 222–232. ISSN: 0304-4076. https://doi.org/https://doi.org/10.1016/j.jeconom.2015.06.005. http://www.sciencedirect.com/science/article/pii/S0304407615001748.

Hoxby, Caroline. 2000. *Peer Effects in the Classroom: Learning from Gender and Race Variation.* Working Paper, Working Paper Series 7867. National Bureau of Economic Research, August. https://doi.org/10.3386/w7867. http://www.nber.org/papers/w7867.

Laschever, Ron A. 2013. *The Doughboys Network: Social Interactions and the Employment of World War I Veterans.* Technical report. June. https://ssrn.com/abstract=1205543.

Lee, Lung-fei, and Jihai Yu. 2010. "Estimation of spatial autoregressive panel data models with fixed effects." *Journal of Econometrics* 154 (2): 165–185. ISSN: 0304-4076. https://doi.org/https://doi.org/10.1016/j.jeconom.2009.08.001. http://www.sciencedirect.com/science/article/pii/S030440760900178X.

Lin, Xu. 2010. "Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables." *Journal of Labor Economics* 28 (4): 825–860. https://doi.org/10.1086/653506. eprint: https://doi.org/10.1086/653506. https://doi.org/10.1086/653506.

———. 2015. "Utilizing spatial autoregressive models to identify peer effects among adolescents." *Empirical Economics* 49, no. 3 (November): 929–960. ISSN: 1435-8921. https://doi.org/10.1007/s00181-014-0897-4. https://doi.org/10.1007/s00181-014-0897-4.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60, no. 3 (July): 531–542. ISSN: 0034-6527. https://doi.org/10.2307/2298123. eprint: http://oup.prod.sis.lan/restud/article-pdf/60/3/531/4468725/60-3-531.pdf. https://doi.org/10.2307/2298123.

Nye, Barbara, Spyros Konstantopoulos, and Larry V. Hedges. 2004. "How Large Are Teacher Effects?" *Educational Evaluation and Policy Analysis* 26 (3): 237–257. https://doi.org/10.3102/01623737026003237. eprint: https://doi.org/10.3102/01623737026003237. https://doi.org/10.3102/01623737026003237.

Patacchini, Eleonora, Edoardo Rainone, and Yves Zenou. 2017. “Heterogeneous peer effects in education.” *Journal of Economic Behavior and Organization* 134:190–227. ISSN: 0167-2681. https://doi.org/https://doi.org/10.1016/j.jebo.2016.10.020. http://www.sciencedirect.com/science/article/pii/S0167268116302499.

Ryan, Ronan, Sally Vernon, Gill Lawrence, and Sue Wilson. 2012. “Use of name recognition software, census data and multiple imputation to predict missing data on ethnicity: application to cancer registry records.” *BMC medical informatics and decision making* 12 (January): 3. ISSN: 1472-6947. https://doi.org/10.1186/1472-6947-12-3. http://europepmc.org/articles/PMC3353229.

Sacerdote, Bruce. 2001. “Peer Effects with Random Assignment: Results for Dartmouth Roommates*.” *The Quarterly Journal of Economics* 116, no. 2 (May): 681–704. ISSN: 0033-5533. https://doi.org/10.1162/00335530151144131. eprint: http://oup.prod.sis.lan/qje/article-pdf/116/2/681/5375285/116-2-681.pdf. https://doi.org/10.1162/00335530151144131.

———. 2011. “Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?,” edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 3:249–277. Handbook of the Economics of Education. Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-444-53429-3.00004-1. http://www.sciencedirect.com/science/article/pii/B9780444534293000041.

Vigdor, Jacob, and Thomas Nechyba. 2007. “Peer effects in North Carolina public schools.” In *In Schools and the equal opportunities problem, ed. Ludger Woessmann and Paul.* MIT Press.

8 Tables

Table 1: Comparing students in schools with a POS system to the full sample

	NYC Student Population		Has POS System		
	Freq.	Percent	Freq.	Percent	difference:
Borough					
Manhattan	182,794	15.64%	164,258	15.59%	-0.06%
Bronx	246,967	21.14%	218,101	20.70%	-0.44%
Brooklyn	350,124	29.96%	312,236	29.63%	-0.33%
Queens	321,262	27.49%	296,443	28.13%	0.64%
Staten Island	67,307	5.76%	62,721	5.95%	0.19%
Total:	1,168,454	100%	1,053,759	100%	
Ethnicity					
hispanic	472,229	49.76%	429,074	50.41%	0.66%
black	302,744	31.90%	265,098	31.15%	-0.75%
white	174,105	18.34%	156,966	18.44%	0.10%
asian other	214,698	22.62%	198,241	23.29%	0.67%
Total:	949,078	100%	851,138	100%	

Data are from the New York City Department of Education (NYCDOE). Table depict differences between all schools and those with a point of sale (POS) system for all students (over all grades). Ethnicity information is not known for all students.

Table 2: Sample Selection Process

Students	Number Drop	Percent Drop	Percent Remaining	Transactions	Number Drop	Percent Drop	Percent Remaining
All 4th and 5th graders at schools using POS systems for the full year							
145,495			100.00%	16,179,728			100.00%
Participate in school lunch							
140,090	5,405	3.71%	96.29%	16,174,323	5,405	0.03%	99.97%
Have a test lag for either math or reading							
129,234	10,856	7.46%	88.82%	15,118,176	1,056,147	6.53%	93.44%
Have a test score for either math or reading							
125,890	3,344	2.30%	86.53%	14,861,855	256,321	1.58%	91.85%
Transaction time is between 10:00am and 2:00pm							
125,771	119	0.08%	86.44%	14,705,898	155,957	0.96%	90.89%
Removing transactions not occuring with student's class							
125,700	71	0.05%	86.39%	14,582,405	123,493	0.76%	90.13%
Removing transactions which are simultaneous for the entire class							
125,559	141	0.10%	86.30%	14,568,447	13,958	0.09%	90.04%
Class size is at least 20 students							
102,606	22,953	15.78%	70.52%	12,010,146	2,558,301	15.81%	74.23%

The table depicts how many students (and corresponding transactions) are lost at each point of the sample selection process.

Table 3: Summary Stats

variable	mean	sd	N
lunch_part_rate	0.658	0.275	102,606
lunch length	15.69	24.31	12,010,146
lunch time	12.07	0.79	12,010,146
class_size	25.57	3.21	102,606
female	0.506	0.500	102,606
grade4	0.489	0.500	102,606
grade5	0.511	0.500	102,606
ever poor	0.845	0.362	102,606
ell	0.125	0.331	102,606
ethnicity:			
hispanic	0.408	0.492	102,606
black	0.190	0.392	102,606
white	0.165	0.371	102,606
asian_other	0.237	0.425	102,606
Borough:			
manhattan	0.099	0.299	102,606
bronx	0.213	0.409	102,606
brooklyn	0.282	0.450	102,606
queens	0.335	0.472	102,606
staten_island	0.070	0.256	102,606
zmath	0.082	0.950	101,948
zread	0.088	0.953	102,244

Summary statistics for our selected sample. Lunch length calculated in minutes. Lunch time is in hours, so the mean lunch time is equivalent to 12:04.

Table 4: Baseline Model

	Math		Reading	
Endogenous Effect	0.089**	(0.009)	0.053**	(0.009)
Own Effect:				
lag test score	0.730**	(0.002)	0.660**	(0.003)
female	-0.011**	(0.004)	0.055**	(0.004)
ELL	-0.094**	(0.006)	-0.193**	(0.007)
Asian/other	0.177**	(0.005)	0.157**	(0.006)
black	-0.019**	(0.005)	-0.043**	(0.006)
white	0.076**	(0.006)	0.088**	(0.007)
ever poor	-0.059**	(0.005)	-0.056**	(0.006)
Exogenous Effect:				
lag test score	-0.007	(0.010)	0.049**	(0.011)
female	0.005	(0.009)	-0.031**	(0.011)
ELL	0.020	(0.021)	0.048	(0.027)
Asian/other	0.121**	(0.019)	0.085**	(0.022)
black	-0.011	(0.020)	-0.037	(0.023)
white	0.045*	(0.021)	0.006	(0.025)
ever poor	-0.007	(0.019)	0.001	(0.023)
observations:	100,156		94,838	

Models include classroom fixed effects, own zip code fixed effects, and zip code exogenous fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

Table 5: Baseline model by grade

	Fourth Grade		Fifth Grade	
	Math	Reading	Math	Reading
Endogenous Effect	0.070**	0.054**	0.105**	0.047**
	(0.013)	(0.013)	(0.013)	(0.013)
Own Effect:				
lag test score	0.701**	0.677**	0.761**	0.644**
	(0.003)	(0.004)	(0.003)	(0.004)
female	-0.029**	0.043**	0.007	0.066**
	(0.005)	(0.007)	(0.005)	(0.006)
ELL	-0.116**	-0.137**	-0.069**	-0.254**
	(0.008)	(0.010)	(0.008)	(0.011)
Asian/other	0.172**	0.172**	0.182**	0.144**
	(0.007)	(0.009)	(0.007)	(0.008)
black	-0.026**	-0.040**	-0.013	-0.047**
	(0.008)	(0.009)	(0.007)	(0.009)
white	0.081**	0.093**	0.071**	0.083**
	(0.008)	(0.010)	(0.008)	(0.009)
ever poor	-0.078**	-0.053**	-0.040**	-0.060**
	(0.007)	(0.009)	(0.007)	(0.008)
Exogenous Effect:				
lag test score	0.010	0.031	-0.025	0.067**
	(0.015)	(0.017)	(0.015)	(0.016)
female	0.001	-0.042**	0.007	-0.021
	(0.013)	(0.016)	(0.013)	(0.015)
ELL	0.028	0.047	0.011	0.046
	(0.029)	(0.038)	(0.029)	(0.039)
Asian/other	0.111**	0.115**	0.129**	0.052
	(0.027)	(0.032)	(0.026)	(0.029)
black	0.000	-0.029	-0.016	-0.052
	(0.029)	(0.034)	(0.027)	(0.032)
white	0.050	0.030	0.038	-0.026
	(0.031)	(0.036)	(0.029)	(0.034)
ever poor	-0.024	0.024	0.006	-0.029
	(0.027)	(0.032)	(0.027)	(0.032)
observations:	48,862	46,015	50,956	48,576

Models include classroom fixed effects, own zip code fixed effects, and zip code exogenous fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

Table 6: Placebo

	Math		Reading	
	parameter	s.e.	parameter	s.e.
Endogenous Effect	0.004	0.032	0.016	0.032
Own Effect:				
lag test score	0.734**	0.003	0.661**	0.003
male	-0.012**	0.003	0.053**	0.004
ELL	-0.094**	0.007	-0.193**	0.009
Asian/other	0.183**	0.006	0.158**	0.007
black	-0.018**	0.006	-0.048**	0.007
white	0.074**	0.006	0.087**	0.008
ever poor	-0.056**	0.006	-0.056**	0.007
Exogenous Effect:				
lag test score	0.075	0.038	0.029	0.041
male	-0.020	0.042	0.008	0.050
ELL	-0.011	0.083	0.045	0.105
Asian/other	0.031	0.071	-0.055	0.082
black	0.010	0.076	-0.132	0.087
white	-0.068	0.076	-0.034	0.089
ever poor	0.061	0.068	0.026	0.079
observations:	100,156		94,838	

Models include classroom fixed effects, own zip code fixed effects, and zip code exogenous fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

Table 7: Removing classrooms with little variance in observed line order

	Cutoff is 0.05		Cutoff is 0.1		Cutoff is 0.15	
	Math	Reading	Math	Reading	Math	Reading
Endogenous Effect	0.090**	0.053**	0.096**	0.056**	0.100**	0.061**
	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)
Own Effect:						
lag test score	0.731**	0.660**	0.730**	0.660**	0.730**	0.661**
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
female	-0.012**	0.055**	-0.011**	0.053**	-0.012**	0.052**
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)
ELL	-0.093**	-0.193**	-0.093**	-0.192**	-0.093**	-0.190**
	(0.006)	(0.007)	(0.006)	(0.008)	(0.006)	(0.008)
Asian/other	0.176**	0.157**	0.176**	0.157**	0.176**	0.156**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
black	-0.018**	-0.043**	-0.018**	-0.042**	-0.017**	-0.041**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)
white	0.075**	0.087**	0.075**	0.087**	0.075**	0.087**
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
ever poor	-0.058**	-0.056**	-0.057**	-0.054**	-0.057**	-0.053**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
Exogenous Effect:						
lag test score	-0.008	0.048**	-0.009	0.058**	-0.005	0.068**
	(0.011)	(0.012)	(0.011)	(0.012)	(0.012)	(0.013)
female	0.009	-0.030**	0.009	-0.027*	0.009	-0.030*
	(0.009)	(0.011)	(0.010)	(0.012)	(0.010)	(0.012)
ELL	0.021	0.045	0.025	0.042	0.023	0.049
	(0.021)	(0.028)	(0.022)	(0.029)	(0.024)	(0.031)
Asian/other	0.129**	0.090**	0.131**	0.099**	0.140**	0.102**
	(0.019)	(0.022)	(0.020)	(0.023)	(0.021)	(0.024)
black	-0.009	-0.047	0.009	-0.053*	0.002	-0.056*
	(0.020)	(0.024)	(0.021)	(0.025)	(0.023)	(0.026)
white	0.047*	0.006	0.053*	0.012	0.061**	0.012
	(0.021)	(0.025)	(0.022)	(0.026)	(0.023)	(0.027)
ever poor	-0.007	-0.001	-0.008	0.003	-0.016	0.007
	(0.019)	(0.023)	(0.020)	(0.023)	(0.021)	(0.024)
observations:	99,275	94,022	97,029	91,920	92,962	88,121

Models include classroom fixed effects, own zip code fixed effects, and zip code exogenous fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

Table 8: Multiple Distance levels

	Distance=2		Distance=3		Distance=4		Distance=5		Distance=6	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading	Math	Reading
Endogenous Effect:										
D=1	0.077**	0.038**	0.076**	0.039**	0.076**	0.038**	0.076**	0.038**	0.075**	0.039**
	(0.008)	(0.009)	(0.008)	(0.009)	(0.008)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)
D=2	0.013	0.019	0.012	0.019	0.013	0.018	0.013	0.019	0.014	0.019
	(0.010)	(0.011)	(0.010)	(0.011)	(0.010)	(0.011)	(0.010)	(0.011)	(0.010)	(0.011)
D=3			0.012	0.000	0.015	0.001	0.015	0.000	0.014	-0.001
			(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
D=4					-0.017	-0.008	-0.016	-0.01	-0.015	-0.01
					(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
D=5							-0.002	0.006	-0.002	0.005
							(0.011)	(0.011)	(0.011)	(0.011)
D=6									-0.004	0.01
									(0.011)	(0.011)
Own Effect:										
lag test score	0.731**	0.661**	0.731**	0.660**	0.731**	0.661**	0.731**	0.661**	0.731**	0.660**
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
female	-0.009*	0.056**	-0.009*	0.055**	-0.009*	0.054**	-0.008*	0.055**	-0.008*	0.055**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)
ELL	-0.093**	-0.194**	-0.092**	-0.197**	-0.090**	-0.200**	-0.091**	-0.201**	-0.091**	-0.203**
	(0.006)	(0.007)	(0.006)	(0.008)	(0.006)	(0.008)	(0.006)	(0.008)	(0.006)	(0.008)
Asian/other	0.177**	0.158**	0.176**	0.156**	0.175**	0.156**	0.174**	0.157**	0.173**	0.158**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
black	-0.020**	-0.044**	-0.018**	-0.043**	-0.017**	-0.040**	-0.015*	-0.040**	-0.014*	-0.040**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)
white	0.076**	0.089**	0.076**	0.088**	0.075**	0.088**	0.075**	0.088**	0.075**	0.088**
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
ever poor	-0.057**	-0.055**	-0.057**	-0.057**	-0.056**	-0.058**	-0.056**	-0.059**	-0.057**	-0.058**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
observations	100,156	94,838	100,156	94,838	100,156	94,838	100,156	94,838	100,156	94,838

Models include classroom fixed effects, own zip code fixed effects, and zip code exogenous fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

Table 9: Distances greater than one

	Distance=2		Distance=3		Distance=4		Distance=5		Distance=6	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading	Math	Reading
Endogenous Effect	0.092**	0.060**	0.102**	0.065**	0.101**	0.064**	0.094**	0.067**	0.091**	0.073**
	(0.011)	(0.011)	(0.014)	(0.014)	(0.016)	(0.016)	(0.018)	(0.018)	(0.019)	(0.020)
Own Effect:										
lag test score	0.732**	0.661**	0.732**	0.662**	0.733**	0.662**	0.733**	0.662**	0.733**	0.662**
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
male	-0.010*	0.055**	-0.010**	0.056**	-0.010**	0.056**	-0.011**	0.054**	-0.011**	0.054**
	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
ELL	-0.093**	-0.193**	-0.093**	-0.195**	-0.092**	-0.197**	-0.092**	-0.198**	-0.092**	-0.200**
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.008)	(0.006)	(0.008)	(0.006)	(0.008)
Asian/other	0.180**	0.159**	0.182**	0.160**	0.182**	0.161**	0.183**	0.161**	0.183**	0.162**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
black	-0.019**	-0.044**	-0.019**	-0.044**	-0.018**	-0.043**	-0.017**	-0.042**	-0.017**	-0.042**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.007)
white	0.077**	0.088**	0.078**	0.089**	0.078**	0.089**	0.079**	0.089**	0.078**	0.090**
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
ever poor	-0.058**	-0.056**	-0.057**	-0.057**	-0.057**	-0.058**	-0.057**	-0.058**	-0.057**	-0.058**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
Exogenous Effect:										
lag test score	-0.001	0.050**	-0.004	0.038*	0.006	0.040*	0.013	0.041	0.009	0.031
	(0.013)	(0.014)	(0.016)	(0.018)	(0.018)	(0.020)	(0.021)	(0.022)	(0.023)	(0.024)
male	0.000	-0.038**	0.003	-0.047**	0.003	-0.056**	0.006	-0.056**	0.005	-0.054**
	(0.011)	(0.013)	(0.012)	(0.015)	(0.014)	(0.016)	(0.015)	(0.018)	(0.017)	(0.020)
ELL	0.029	0.021	0.039	-0.026	0.071	-0.072	0.064	-0.093	0.055	-0.123
	(0.027)	(0.036)	(0.034)	(0.044)	(0.039)	(0.051)	(0.045)	(0.058)	(0.050)	(0.065)
Asian/other	0.102**	0.064*	0.101**	0.059	0.087**	0.058	0.074*	0.061	0.062	0.077
	(0.023)	(0.026)	(0.027)	(0.032)	(0.031)	(0.036)	(0.035)	(0.040)	(0.038)	(0.044)
black	-0.021	-0.062*	-0.007	-0.059	0.028	-0.024	0.054	-0.005	0.056	0.013
	(0.026)	(0.030)	(0.032)	(0.037)	(0.037)	(0.043)	(0.041)	(0.048)	(0.046)	(0.053)
white	0.041	0.019	0.059*	0.015	0.037	0.016	0.044	0.032	0.034	0.053
	(0.025)	(0.029)	(0.030)	(0.035)	(0.034)	(0.039)	(0.037)	(0.043)	(0.040)	(0.047)
ever poor	0.014	-0.001	0.034	-0.026	0.040	-0.033	0.045	-0.041	0.046	-0.042
	(0.022)	(0.026)	(0.026)	(0.031)	(0.030)	(0.035)	(0.033)	(0.038)	(0.035)	(0.041)
observations:	100,156	94,838	100,156	94,838	100,156	94,838	100,156	94,838	100,156	94,838

Models include classroom fixed effects, own zip code fixed effects, and zip code exogenous fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

Table 10: Evolution of friendship importance over the school year

	Periods=2		Periods=3		Periods=4	
	Math	Reading	Math	Reading	Math	Reading
Endogenous Effect:						
Period 1	0.045**	0.030**	0.034**	0.032**	0.026**	0.023**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
Period 2	0.038**	0.021*	0.028**	0.001	0.029**	0.006
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Period 3			0.027**	0.025**	0.024**	0.019*
			(0.009)	(0.009)	(0.009)	(0.009)
Period 4					0.01	0.008
					(0.008)	(0.008)
Own Effect:						
lag test score	0.730**	0.660**	0.730**	0.659**	0.730**	0.659**
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
female	-0.010*	0.055**	-0.010*	0.056**	-0.010*	0.056**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
ELL	-0.093**	-0.193**	-0.093**	-0.193**	-0.093**	-0.192**
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
asian/other	0.177**	0.157**	0.177**	0.156**	0.176**	0.156**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
black	-0.019**	-0.043**	-0.019**	-0.044**	-0.019**	-0.043**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
white	0.075**	0.087**	0.075**	0.087**	0.074**	0.086**
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
ever poor	-0.058**	-0.055**	-0.057**	-0.056**	-0.056**	-0.056**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
observations	100,156	94,838	100,156	94,838	100,156	94,838

Models include classroom fixed effects, own zip code fixed effects, and zip code exogenous fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

Table 11: Students must participate together more than one day

	At least 2 days		At least 3 days		At least 4 days	
	Math	Reading	Math	Reading	Math	Reading
Endogenous Effect	0.079**	0.048**	0.081**	0.049**	0.081**	0.049**
Own Effect:	(0.008)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)
lag test score	0.730**	0.660	0.730	0.660	0.730	0.660
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
female	-0.010*	0.055**	-0.010*	0.055**	-0.010*	0.055**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
ELL	-0.093**	-0.193**	-0.093**	-0.193**	-0.093**	-0.193**
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
black	-0.018**	-0.043**	-0.019**	-0.043**	-0.019**	-0.043**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
white	0.076**	0.087**	0.076**	0.087**	0.076**	0.087**
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
Asian/other	0.177**	0.157**	0.177**	0.157**	0.177**	0.157**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
ever poor	-0.058**	-0.056**	-0.058**	-0.057**	-0.058**	-0.057**
Exogenous Effect:	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
lag test score	0.000	0.041	-0.001	0.042	0.000	0.042
	(0.010)	(0.011)	(0.010)	(0.011)	(0.010)	(0.011)
female	0.000	-0.027*	0.003	-0.027*	0.003	-0.029**
	(0.009)	(0.011)	(0.009)	(0.011)	(0.009)	(0.011)
ELL	0.025	0.033	0.021	0.033	0.019	0.032
	(0.020)	(0.027)	(0.020)	(0.027)	(0.020)	(0.027)
black	-0.007	-0.039	-0.009	-0.049*	-0.010	-0.051*
	(0.019)	(0.022)	(0.019)	(0.022)	(0.019)	(0.022)
white	0.031	-0.007	0.033	-0.026	0.013	-0.014
	(0.019)	(0.023)	(0.020)	(0.023)	(0.019)	(0.023)
Asian/other	0.096**	0.070**	0.093**	0.058**	0.088**	0.060**
	(0.018)	(0.020)	(0.018)	(0.020)	(0.018)	(0.020)
ever poor	0.009	-0.008	0.016	-0.008	0.012	-0.005
	(0.017)	(0.020)	(0.017)	(0.020)	(0.017)	(0.020)
observations:	100,156	94,838	100,156	94,838	100,156	94,838

Models include classroom fixed effects, own zip code fixed effects, and zip code exogenous fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

Table 12: Students must participate together more than one day (continued)

	At least 5 days		At least 8 days		At least 10 days	
	Math	Reading	Math	Reading	Math	Reading
Endogenous Effect	0.080**	0.049**	0.083**	0.054**	0.079**	0.054**
Own Effect:	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
lag test score	0.730**	0.660	0.730	0.659	0.730	0.659
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
female	-0.011**	0.055**	-0.011**	0.056**	-0.011**	0.057**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
ELL	-0.093**	-0.193**	-0.093**	-0.193**	-0.093**	-0.194**
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
black	-0.019**	-0.043**	-0.019**	-0.043**	-0.019**	-0.043**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
white	0.076**	0.087**	0.076**	0.087**	0.076**	0.087**
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
Asian/other	0.178**	0.157**	0.177**	0.157**	0.178**	0.157**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
ever poor	-0.058**	-0.056**	-0.058**	-0.057**	-0.058**	-0.057**
Exogenous Effect:	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
lag test score	-0.004	0.041	-0.006	0.030	-0.006	0.028
	(0.010)	(0.011)	(0.010)	(0.011)	(0.010)	(0.011)
female	0.004	-0.029**	0.004	-0.034**	0.004	-0.035**
	(0.009)	(0.011)	(0.009)	(0.011)	(0.009)	(0.011)
ELL	0.017	0.033	0.017	0.021	0.016	0.022
	(0.020)	(0.026)	(0.020)	(0.026)	(0.020)	(0.026)
black	-0.010	-0.047*	-0.006	-0.038	-0.009	-0.038
	(0.019)	(0.022)	(0.019)	(0.022)	(0.018)	(0.022)
white	0.015	-0.012	0.018	-0.001	0.018	0.004
	(0.019)	(0.023)	(0.019)	(0.022)	(0.019)	(0.022)
Asian/other	0.090**	0.066**	0.095**	0.076**	0.097**	0.081**
	(0.017)	(0.020)	(0.017)	(0.020)	(0.017)	(0.020)
ever poor	0.024	-0.003	0.022	-0.021	0.021	-0.014
	(0.017)	(0.020)	(0.017)	(0.020)	(0.017)	(0.020)
observations:	100,156	94,838	100,156	94,838	100,156	94,838

Models include classroom fixed effects, own zip code fixed effects, and zip code exogenous fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

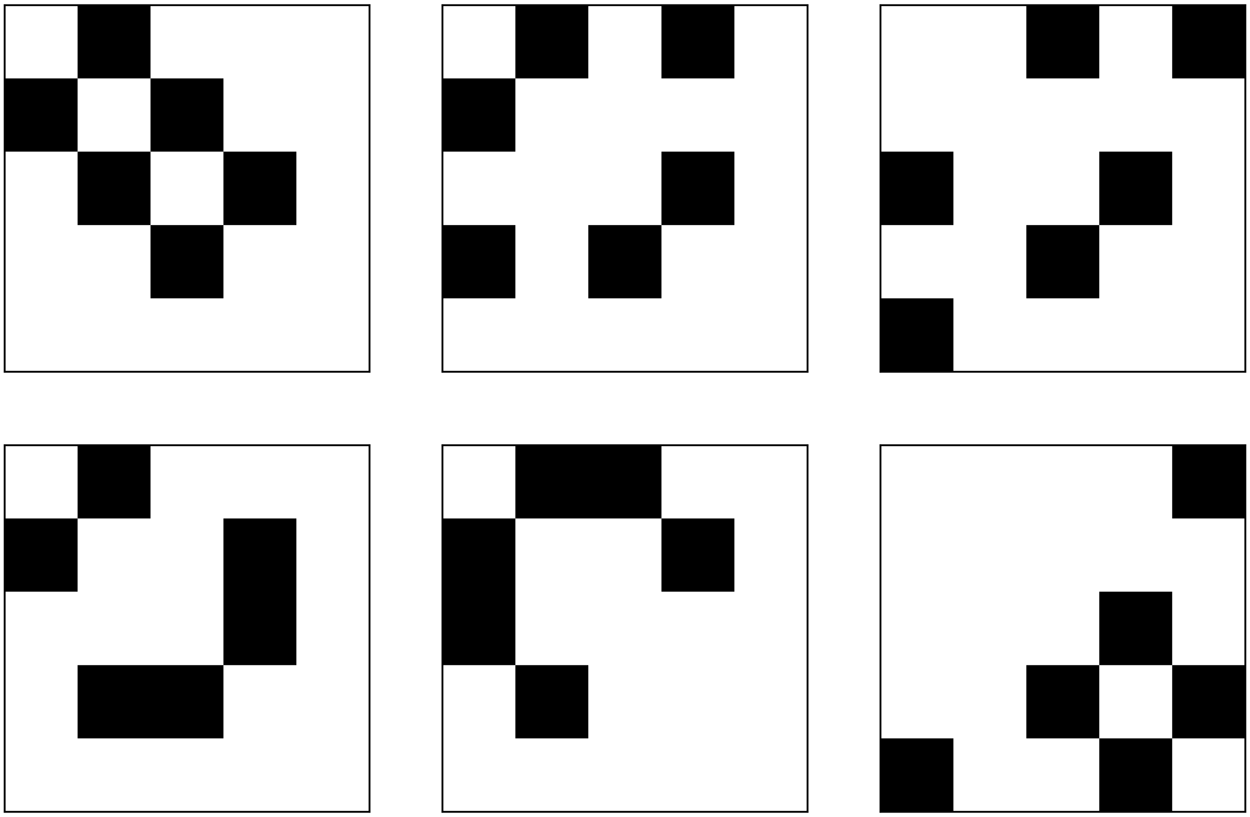
Table 13: Temporal proximity

	Math		Reading	
Endogenous Effect	0.025	(0.015)	0.046**	(0.015)
Own Effect:				
lag test score	0.731**	(0.002)	0.659**	(0.003)
female	-0.011**	(0.003)	0.051**	(0.004)
ELL	-0.094**	(0.006)	-0.196**	(0.007)
Asian/other	0.181**	(0.005)	0.161**	(0.006)
black	-0.019**	(0.005)	-0.042**	(0.006)
white	0.077**	(0.006)	0.088**	(0.007)
ever poor	-0.058**	(0.005)	-0.057**	(0.006)
Exogenous Effect:				
lag test score	-0.007	(0.013)	-0.033*	(0.013)
female	-0.011	(0.011)	-0.034**	(0.013)
ELL	-0.023	(0.018)	-0.014	(0.025)
Asian/other	-0.029	(0.017)	0.019	(0.020)
black	-0.010	(0.018)	0.002	(0.021)
white	-0.024	(0.019)	0.003	(0.023)
ever poor	0.008	(0.017)	0.013	(0.020)
observations:	100,156		94,838	

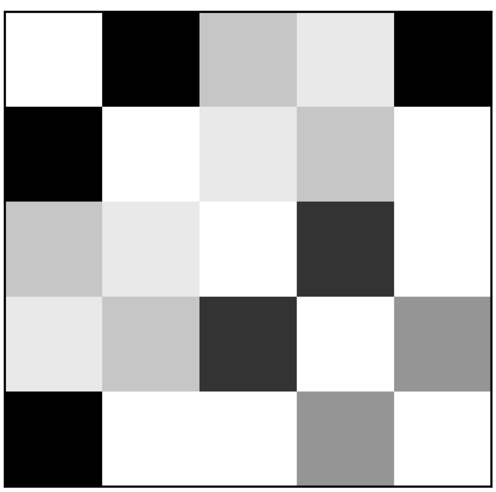
This table shows results using temporal proximity rather than physical (line order) proximity. Models include classroom fixed effects, own zip code fixed effects, and zip code exogenous fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

9 Figures

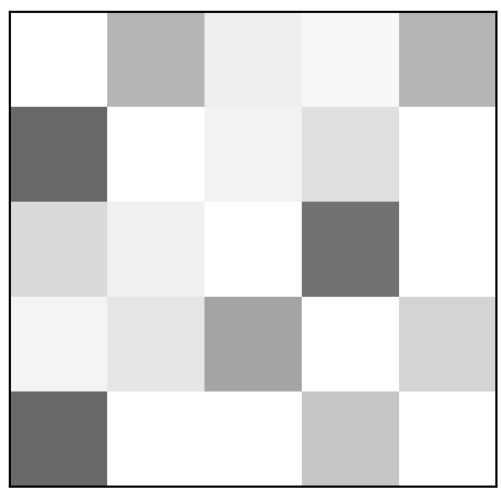
Figure 1: Constructing Proximity and Weighting Matrices



(a) Six example daily lunch line observations



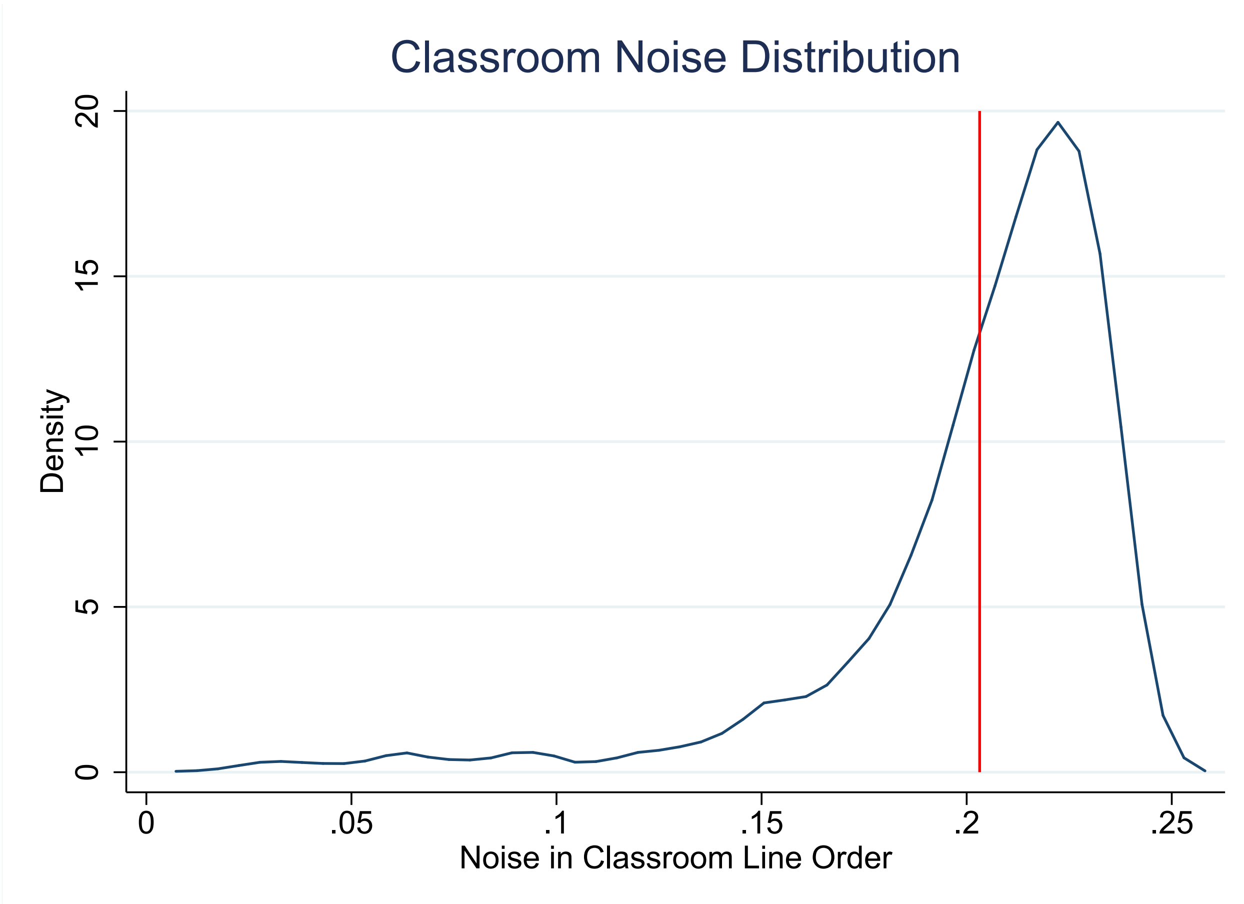
(b) Resulting proximity matrix



(c) Resulting weighting matrix

Example daily observations in (a) are constructed from example line orders [A,B,C,D]. [C,D,A,B], [D,C,A,E], [C,D,B,A], [D,B,A,C], and [A,E,D,C]. Notice that each day includes one absence, so that we can observe how this affects our measure of connection. (b) Shows the resulting proximity matrix, which is constructed according to equation 3.1 and (c) shows the result when we row-normalize this into a weighting matrix. Notice that students C and D are constructed to have a strong connection (near one another 5 of 6 days, and students A and B are constructed to have a weaker but still strong connection. Student E is next to A on the rare occasion E is present. We see these connections bear out in both the proximity matrix and in the final weighting matrix.

Figure 2: Classroom Noise Distribution



Vertical line indicates mean, which is equal to 0.2031616. Includes one entry for each of 4,077 classrooms. Density constructed using an Epanechnikov kernel with bandwidth = 0.004.

A Estimation procedures

We estimate a spatial autoregressive (SAR) model of the form:

$$Y = \lambda WY + \theta WX + X\beta + U \tag{A.1}$$

Where X contains a constant and the fixed effects for simplicity of notation. We estimate the model using maximum likelihood estimation (MLE), and so assume U is iid$(0, \sigma^2)$. Note that if we do not assume normality of the error term, this becomes quasi-maximum likelihood esetimation (QMLE).

In order to maintain the interdependencies of the error terms and incorporate classroom fixed effects, we follow the transformation approach discussed in Lee and Yu (2010) and used in W. Horrace et al. (2019). This method involves the deviation from the classroom mean operator $Q = \iota\iota'/n$ an $n \times n$ matrix where n is classroom size. We define the orthonormal within transformation matrix Q as $[P, \iota_n/\sqrt{n}]$. Following Lee and Yu (2010), we premultiply our model by P':

$$P'Y = \lambda P'WY + P'\theta WX + P'X\beta + P'U \tag{A.2}$$

This means our log likelihood function takes the form:

$$ln\mathcal{L}(\lambda, \beta, \sigma^2) = -\Big(\frac{n-1}{2}\Big)\Big[ln(2\pi) + ln(\sigma^2)\Big] + ln|I - \lambda P'WP| - \frac{\bar{e}'Q\bar{e}}{2\sigma^2} \tag{A.3}$$

Where $\bar{e} = P'Y - \lambda P'WQY - P'WQX\theta - P'X\beta$ After some algebra, we can rewrite this with only Q (and not P):

$$ln\mathcal{L}(\lambda, \beta, \sigma^2) = -\Big(\frac{n-1}{2}\Big)\Big[ln(2\pi) + ln(\sigma^2)\Big] - ln(1-\lambda) + ln|I - \lambda W| - \frac{e(\lambda,\xi)'Qe(\lambda,\xi)}{2\sigma^2} \tag{A.4}$$

where $\xi = (\theta, \beta)'$, $\mu = (WX, X)$ and $e(\xi) = Y - \lambda WY - WX\theta - X\beta = Y - \lambda WY - \mu\xi$. Notice that the parameter space for λ must be restricted such that its magnitude is less than one in order to guarantee that both $|I - \lambda W|$ will be strictly positive and $\ln(1-\lambda)$ is well defined.

We simplify estimation by concentrating out the ξ and σ^2 using first order conditions. Thus:

$$\xi^\star(\lambda) = (\mu'Q\mu)^{-1}\mu'Q(Y - \lambda WY) \tag{A.5}$$

and

$$\sigma^{2\star}(\lambda,\xi) = \frac{e'(\lambda,\xi)Qe(\lambda,\xi)}{n-1} \tag{A.6}$$

This simplifies estimation substantially, as we need only maximize in one dimension. We get some cancellation from the $\sigma^{2\star}$ and our likelihood function for an individual class becomes:

$$ln\mathcal{L}(\lambda) = -\Big(\frac{n-1}{2}\Big)\Big[ln(2\pi) + 1 + ln[\sigma^{2\star}(\lambda)]\Big] - ln(1-\lambda) + ln|I - \lambda W| \tag{A.7}$$

We sum the likelihoods over all classrooms to obtain the complete likelihood, analogous to the way Lee and Yu (2010) sum over the time periods.

In order to calculate the standard errors, we again follow Lee and Yu (2010) and estimate the asymptotic variance matrix V_{ML} as in the block matrix below:

$$V_{ML} = \begin{pmatrix} a & d & e \\ d & b & 0 \\ e & 0 & c \end{pmatrix}^{-1} \tag{A.8}$$

Such that:

$$\begin{aligned} a &= \frac{\partial \ln \mathcal{L}(\lambda,\xi,\sigma^2)}{\partial\lambda_k\partial\lambda_l} = (W_kG)'QW_lG/\sigma^2 + \mathrm{tr}[W_kGQW_lG] \\ a &= \frac{\partial \ln \mathcal{L}(\lambda,\xi,\sigma^2)}{\partial\xi^2} = \mu'Q\mu/\sigma^2 \\ c &= \frac{\partial \ln \mathcal{L}(\lambda,\xi,\sigma^2)}{\partial\sigma^4} = (n-1)/(2\sigma^4) \\ d &= \frac{\partial \ln \mathcal{L}(\lambda,\xi,\sigma^2)}{\partial\lambda_k\partial\xi} = (W_kG)'Q\mu\sigma^2 \\ e &= \frac{\partial \ln \mathcal{L}(\lambda,\xi,\sigma^2)}{\partial\lambda_k\partial\sigma^2} = \mathrm{tr}[QW_kG]/\sigma^2 \end{aligned} \tag{A.9}$$

where $G = (I - \sum_k \lambda_k W_k)^{-1}$. The standard errors are then the square roots of the diagonal of V_{ML}.

B Measuring Order Noise

When analyzing a social network based off the observed lunch line order, we may be concerned that students do not have agency over their place in line. If the students are ordered by some external factor, such as a teacher, then the interpretation of our results changes. As such, we attempt to determine whether there is a large set of classrooms in which students are ordered. The measure we intend to create will be able to determine whether a consistent order is used throughout the period of observation. If a teacher orders students alphabetically (for example) for lunch every day, we will detect this line order as having little noise.

In determining a good measure of noise, the measure must have two specific characteristics. First, the measure needs to be invariant to classroom size so that we can compare noise across classrooms without concern that the driving factor is number of students. The second issue is that students do not participate every day, so the measure must be able to contend with varying student combinations and line sizes. Thus any bias in our measure cannot be a function of class size or lunch participation rate. For the purpose of developing intuition, we discuss first an intuitive measure that does not meet these criteria, and then its relationship to a measure that does.

Consider every pair of students (i, j) within a classroom. For these students, we define two quantities A_{ij} and B_{ij}. Let $r(i)$ be the rank of student i and $\mathbb{1}_d(i, j)$ be an indicator function for both students i and j being present at lunch on day d. Then $A_{ij} = \sum_D \mathbb{1}_d\big(r(i) < r(j)\big)$ and $B_{ij} = \mathbb{1}_d(i, j)$. These quantities allow us to determine the number of switches $C_{ij} = \min\big(A_{ij}, B_{ij} - A_{ij}\big)$ if we assume that the most common order is the "true" order of the students. The quantity $S = \sum_{i<j} C_{ij}$ gives the total number of inversions, and we normalize this by the number of observed pairs $B = \sum_{i<j} B_{ij}$, so that our noise measure is $M = S/B$. This has a nice interpretation as the chance that a given pair is swapped. However the measure does not quite have the properties we would like - the measure varies by size and participation rate.

To understand the issue, we look at the bias of our measure. For all pairs (i, j), there exists some probability q of swapped order, conditioning on the appearance of both students (i, j). This results in the expectation of the total number of times i and j swap order equal to $q \cdot B_{ij}$. We consider the expectation of our estimator: $\mathbb{E}[C_{ij}] = \sum_{k=0}^{B_{ij}} \binom{B_{ij}}{k} q^k (1-q)^{(B_{ij}-k)} \min(k, B_{ij} - k)$. This expectation varies with B_{ij}. When $B_{ij} \in \{0, 1\}$, $\mathbb{E}[C_{ij}] = 0$; when $B_{ij} \in \{2, 3\}$, $\mathbb{E}[C_{ij}] = B_{ij} q(1-q)$; and more complex objects as B_{ij} increases. We look for a $\tilde{C}_{ij}$ where $\mathbb{E}[\tilde{C}_{ij}] = B_{ij} q(1-q)$ regardless of B_{ij} (being invariant in B_{ij} should meet the requirements of invariance to class size and participation rate). To do this, we replace $\min(k, B_{ij} - k)$ with $\binom{B_{ij}-1}{k}^{-1} \binom{B_{ij}-2}{k-1} = \frac{k(B_{ij}-k)}{B_{ij}-1} = \varphi$. What is nice about φ is that it keeps much of the meaning of $\min(k, B_{ij} - k)$. Without loss of generality, we can say that $\min(k, B_{ij} - k) = k$. Notice that in both measures, $C_{ij=0}$ when $k = 0$, so we consider only $k > 0$. Then $1 \leq k \leq \frac{B_{ij}}{2}$. Thus $\frac{B_{ij}}{2} \leq B_{ij} - k \leq B_{ij} - 1$.

This implies $\frac{k}{2} < \frac{kB_{ij}}{2(B_{ij}-1)} \leq \frac{k(B_{ij}-k)}{B_{ij}-1} = \varphi \leq k$, and we see that φ is bounded by k and $\frac{k}{2}$, although it loses the nice interpretation of our estimator being the chance i and j are swapped. We do however gain invariance by size and absences, which we will show when we finish constructing the measure. As before, we normalize this by dividing by the number of observed pairs. The result is:

$$M = \frac{\sum_{i<j;B_{ij}>1} \tilde{C}_{ij}}{\sum_{i<j;B_{ij}>1} B_{ij}} \quad \text{where} \quad \tilde{C}_{ij} = \frac{A_{ij}(B_{ij} - A_{ij})}{B_{ij} - 1} \tag{B.1}$$

The omission we are left with is the case for $B_{ij} = 1$. Given probability p that a pair of students participate in lunch, the expectation that the students participate in lunch together only one time is $\mathbb{E}[B_{ij} = 1] = Dp(1-p)^{(D-1)}$. Average lunch participation is 65.8% over a school year of 180 days.[18] If two students participated 25% of the time (such that $p = .0625$), the chance of $B_{ij} = 1$ is less than e^{-9} if the participation of students i and j is independent. Given such a small chance, we ignore the scenario $B_{ij} = 1$ and forcibly remove such occurrences from the measure.

We can see that our measure is invariant to class size and participation rate in Table B1, which reports results of Monte-Carlo simulations under changes in class size and participation rate. Line orders are generated randomly for each of 180 days to simulate observation throughout the school year. Standard errors decrease in participation rate.

The measure is meant to detect noise, so we also simulate increases in randomness to show that the measure works as promised. Table B2 reports results of Monte-Carlo simulations on the measure of classroom noise in response to increasing levels of randomness. Line orders are generated randomly for X% of the 180 days, where X is in the percent random column. In all of these simulations, students have a 70% chance of participation in the line on any simulated day. The measure increases as randomness increases. We also show what may be apparent from the previous discussion, which is that when students are perfectly ordered in the classroom, the measure is zero. The average measure observed in the data is 0.203, which is consistent with between 50% and 60% randomness in the lunch line. This makes sense, as we expect variation in the line order, but as students reveal their preferences for line location and who to be in line with, we expect the sorting to be less than random. It is likely that classroom geography also plays a part in which groups of students are most likely in the front of the line on a consistent basis, further reducing the number of inversions detected by the measure. We expect that the lower level of randomness is not restricted to specific days of the year, as in the simulations, but rather each day has non-random variation.

It is important to note that this measure will be limited if the teacher attempts to be more equitable and

18. The school year is required to be at least 180 days, and is sometimes longer than this. In 2013, the school year was exactly 180 days.

alternates (for example) lining their students up alphabetically one day and reverse alphabetically another day, we would not detect this as an ordering (because there will be many line switches even without large changes in relative position). While limited in this way, we argue that the majority of orderings we might be concerned with (ex: alphabetical, height, location with the classroom, etc) will be detected by this measure.

C Classroom Assignment

There are two forms of selection that are important to address. The first is the assignment of students to their set of potential peers. This occurs through two channels: assignment of students to school and then to classrooms. The second is the selection of friends within the classroom.[19] In this section I provide evidence that, conditional on school attended, student assignment into classrooms is consistent with randomness over most observed characteristics.

A key assumption for our estimates to be causal is that that assignment of students to their choice of peers (classroom assignment) is random. We have no insight into the assignment process, but we do show that over most observed characteristics, assignment is consistent with randomness. The objective of this test is to show that classroom assignment is not a function of the observed characteristics along which sorting into friendships might occur. To do this, we consider a series of multinomial logits as follows:

$$Class_i = \alpha + X_i\beta_{gst} + \varepsilon_i \tag{C.1}$$

For each iteration of equation (C.1) we include a single grade g within a single school s, during a single year t. We exclude all school-grades for which there is only a single classroom, as these schools by definition assign their students to classrooms randomly (less than 5% of our sample are in cohorts with only one classroom). $Class_i$ indicates the classroom assignment for student i, and the number of options varies by school-grade.[20] X_i is a binary indicator variable for a characteristic of student i. Each iteration of equation (C.1) gives us an estimate β_{gst} and a t-statistic. The t-statistic tells us the significance of the characteristic for assignment at that school-grade, and we collect the t-statistic for all school-grades. We then conduct our own random assignment of students to classrooms and run the same set of models, again collecting these t-statistics. We then compare the distributions of t-statistics from the observed and simulated models.

19. In the current version, this within-classroom friendship selection is dealt with primarily through group fixed effects. Homophily plays an important role in who students select as their friends, and the models used in this paper include a large set of demographic characteristics to control for these sorting avenues - including gender, ethnicity, residential zip code, and others. This is in line with other literature which uses fixed effects for networks of importance to control for these sorting effects. However, I include additional information in my measure of connection strength. To the extent that this additional information is the result of sorting which is not controlled for by these avenues, further work needs to be done. Future versions of this paper will include a more thorough examination of this within-classroom sorting.

20. There are between 2 and 11 classrooms in a school-grade-year. The mean is 4.2 classrooms.

Figures A1 and A2 show the results of these tests. Each dot in these figures compares equal ranked t-statistics from the simulated and observed populations. If these distributions are the same, we should expect a 45 degree line. Most of the observed characteristics remain reasonably close to the 45 degree line, with the largest deviation at the tails of the distribution. A notable exception is the English language learner characteristic, which appears to deviate significantly from the 45 degree line. This indicates that classroom assignment may group English language learners into classrooms together. It is important to note that this test behaves best when the group sizes are similar in size, such as when we compare female and male students. English language learners make up only 12.5% of the population. That said, this is similar (slightly smaller than) the size of both white and Asian/other students in our sample, and both of these groups appear to behave better in this visual test. Thus, with the exception of English language learners, we conclude that we do not need to be concerned about selection into classrooms based on observed characteristics.

D Appendix Tables and Figures

Table B1: Monte Carlo simulations of classroom size and participation rates

Participation rate:	Class size 20		Class size 25		Class size 30	
25%	0.2501	(0.0027)	0.25	(0.0023)	0.2501	(0.0019)
50%	0.25	(0.0008)	0.25	(0.0007)	0.25	(0.0006)
75%	0.25	(0.0005)	0.25	(0.0004)	0.25	(0.0004)
100%	0.25	(0.0003)	0.25	(0.0003)	0.25	(0.0003)

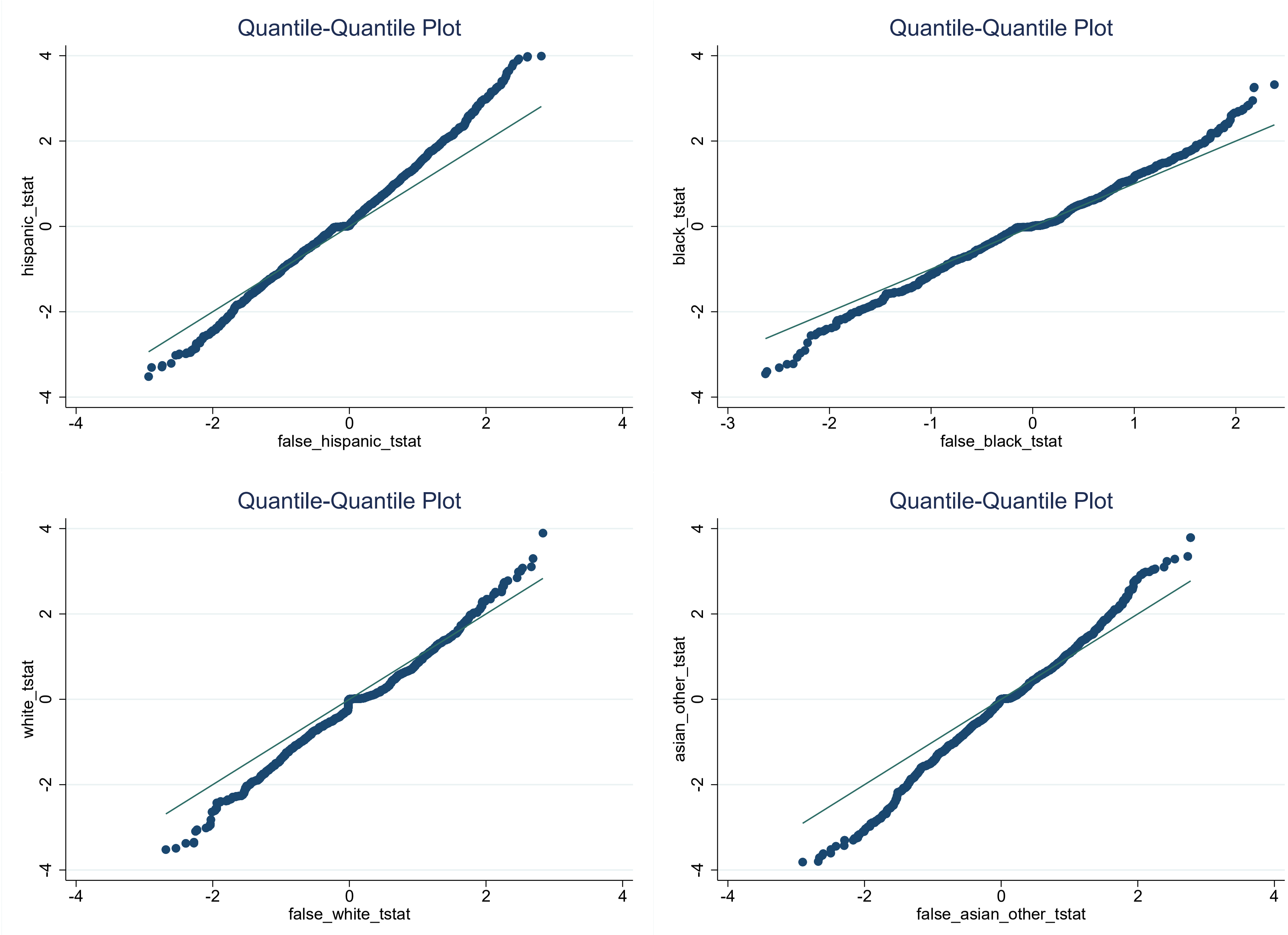
Table reports the results of Monte-Carlo simulations on the measure of classroom noise. Simulation is for 1,000 classrooms at each combination of class size and participation rate. Standard errors are in parenthesis. Line orders are generated randomly for each of 180 days to simulate observation throughout the school year. The measure is invariant to class size and participation rate, although standard errors increase as participation rate decreases.

Table B2: Monte Carlo simulations for different levels of randomness

Percent random:	mean	s.e.
0%	0.0000	(0.0000)
10%	0.0489	(0.0033)
20%	0.0902	(0.0040)
30%	0.1287	(0.0042)
40%	0.1603	(0.0039)
50%	0.1886	(0.0034)
60%	0.2104	(0.0029)
70%	0.2282	(0.0022)
80%	0.2402	(0.0015)
90%	0.2478	(0.0008)
100%	0.2500	(0.0004)
N	10,000	

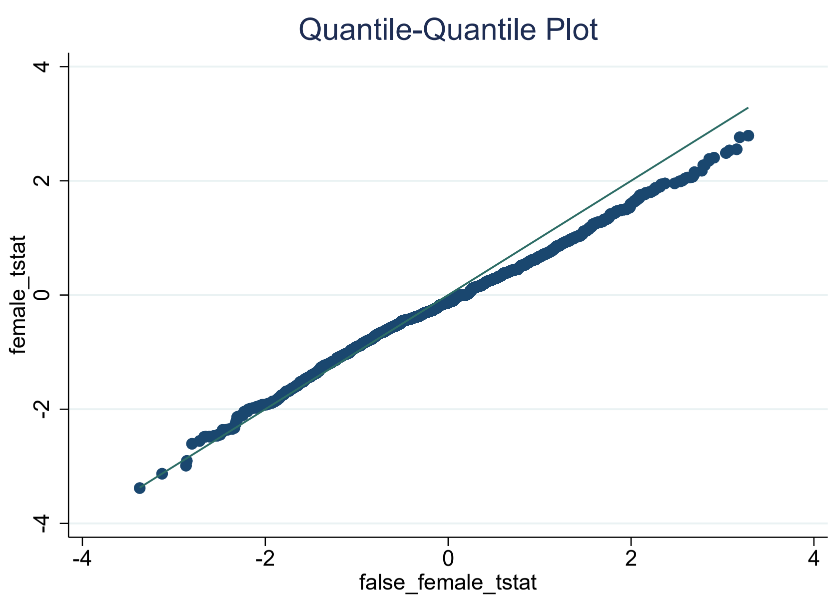
Table reports results of Monte-Carlo simulations on the measure of classroom noise. Simulation is for 10,000 classrooms at each combination of class size and participation rate. Line orders are generated randomly for X% of the 180 days, where X is in the percent random column. Students have a 70% chance of participation in the line on any simulated day. The measure increases as randomness increases.

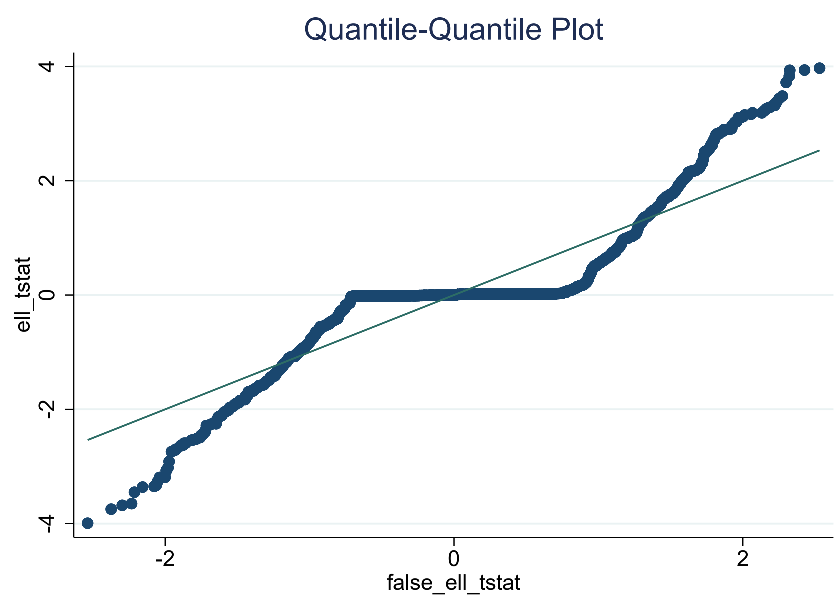
Figure A1: Quantile-Quantile Plots

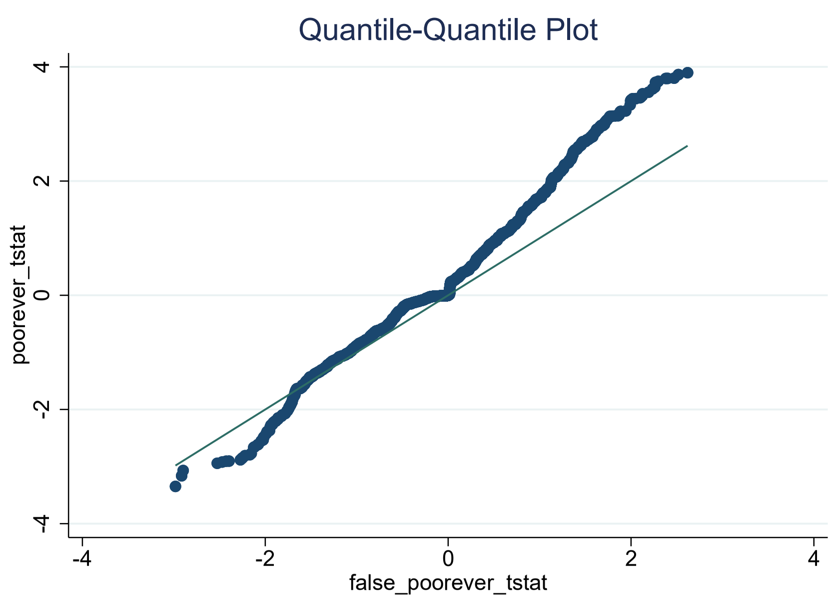


A series of multinomial logits are run to estimate the importance of each ethincity indicator in class assignment. In each pair of graphs, the left plots the t-statistics from these against the t-statisitics from a similar exercise in which we randomly assign students to classrooms. Thus we are plotting these two distributions against one another, and if the distributions are the same, we should expect a straight line of slope one. We argue that these provide evidence that class assignment is consistent with a random process.

Figure A2: Quantile-Quantile Plots







A series of multinomial logits are run to estimate the importance of gender, whether students are English language learners, or whether the student rides is poor. In each pair of graphs, the left plots the t-statistics from these against the t-statisitics from a similar exercise in which we randomly assign students to classrooms. Thus we are plotting these two distributions against one another, and if the distributions are the same, we should expect a straight line of slope one. We argue that these provide evidence that class assignment is consistent with a random process for most observed characteristics. This test performs best when groups are large (ex: females are about half the student population) and is more noisy when the network is small (ex: English language learners are only about 12.5% of the student population).