You Are Who You Eat With

Academic Peer Effects from School Lunch Lines

Jonathan L. Presler *

08/08/2022

Abstract

Using daily lunch transaction data from NYC public schools, I determine which students frequently stand next to one another in the lunch line. I use this 'revealed' friendship network to estimate academic peer effects in elementary school classrooms, improving on previous work by defining not only where social connections exist, but the relative strength of these connections. Equally weighting all peers in a reference group assumes that all peers are equally important and may bias estimates by underweighting important peers and overweighting unimportant peers. 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 my claim that these are friendship networks. I also compare the influence of friends from different periods in the school year and find that connections occurring around standardized testing dates are most influential on test scores.

Keywords: Peer effect, Network, Education.
JEL Codes: I21, C31

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, the APPAM conference, and two anonymous reviewers for useful comments. The opinions expressed are those of the author and do not represent the views of the NYC Department of Education or NIH.

*Sinquefield Center for Applied Economic Research at Saint Louis University, Saint Louis, MO 63103

1 Introduction

Schools are cliquish environments where students sort into social groups and develop friends, best friends, and acquaintances. The 2016 Gallup Student Poll finds that 84% of fifth graders have a best friend at school. It follows that within a classroom, students are not equally affected by each of their classmates – indeed, this would be a bold assumption. Yet little work examining peer effects attempts to measure variation in the strength of social ties. By misrepresenting the peer group in this way, we may bias estimates towards zero by overweighting unimportant peers and underweighting influential ones.

A growing body of research exploring peer effects point to the existence of peer effects in education. One important step towards harnessing the potential of peer effects is understanding their size. If we know peer effects exist but are negligible, it may not be worth attempting to harness their effects. But if they are large, it becomes deeply important to understand their impact on both treatments that explicitly rely on peer effects as a key mechanism (such as tracking, school choice, and school integration) and on treatments that do not (due to the multiplicative nature of spillovers). The ability to harness peer effects across all types of interventions could be quite powerful.

In this paper, I use daily lunch transaction data from NYC public schools to determine which students frequently stand next to one another in the lunch line as a 'revealed' friendship network. I use this friendship network to estimate academic peer effects in elementary school classrooms. School lunch is an unusual part of the day because student actions are not directly supervised and do not have direct academic consequences. It is a relatively unstructured and social time, during which students' primary concern is with whom they are able to socialize. This makes it an ideal context in which to observe social connections between students and the resulting peer effects.

I find significant peer effects for both math and reading test scores, with higher spillover effects in math than in reading. The network I construct is based on daily interactions, allowing me to explore the evolution of friendships over different periods in the school year. My results show that friends at the time of the test have the largest impact on test scores. This suggests that friendships change over the course of the school year, and the influence of those who are no longer friends may be short-lived relative to current friends.

The contributions of this paper are as follows. First, I measure opportunities for interaction between students in the lunch line and demonstrate a novel approach for revealing a friendship network. Data in which social connections are observed are rare and typically not administrative.[1] Second, my measure of social interaction allows me to both refine the class-

[1] Friendship surveys may be unreliable or include significant noise. Landini et al. (2016) finds discrepancies between parent and student friendship surveys, and highlights the lack of reciprocity in both surveys.

room environment according to a revealed friendship network and properly weight classmates in order to overcome the binary nature of connections (students are either friends or they are not). I weight friendships on a continuous scale of importance, meaning that my results are not predicated on overweighting unimportant students and underweighting important peers as group averages necessarily do. Third, to my 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 me to decompose the peer effect into its social and contextual components,[2] a distinction which is important for accurate estimation of spillover effects.

This paper continues as follows. In Section 2, I motivate this work in the context of related research. I then describe the data I use in this paper and how I construct the sample in Section 3. Section 4 discusses how I measure connection between students. Once we understand the network structure, I discuss identification issues, how I overcome them, and the model I estimate in Section 5. Section 6 report my baseline results and shows that a student's friends have a significant and sizable influence on their academic performance. I conduct a number of robustness checks in Section 7. This includes providing evidence that the network is indeed a friendship network as well as exploring the effects of friends at different points in the school year. Section 8 concludes.

2 Background

2.1 Friendship Peer Effect Mechanisms

Researchers have long believed that peers play an influential role in education production (Coleman et al., 1966). Classmates undoubtedly share social experiences – these could include particularly disruptive classmates or a set of norms governing socially acceptable behavior. The costs and benefits resulting from exposure to classmates may be independent of friendship – particularly if teachers actively break up friendship groups during group work. Below, I discuss several mechanisms through which peer effects may operate, which are more salient within friendship groups.

It is helpful to think about the incentives students face in school when thinking about the influence of friends and classmates. While schools may be maximizing student academic

Higher reciprocity leads them to treat the student survey as correct, but the large number of unreciprocated connections may be concerning - both in the survey used in this paper and in work using Add Health or other friendship surveys.

[2] These effects are also called the endogenous and exogenous effect, as in Manski (1993). More discussion of the identification of these effects can be found in section 5.1.

performance given a set of inputs,[3] students are unlikely to be maximizing academic performance (Lavecchia et al., 2016). Academic performance is linked to a future benefit in the labor market, but students are myopic and social concerns are present today.[4] Akerlof and Kranton (2002) argue that a student is motivated by social identity.

Glaeser and Scheinkman (1999) discuss several mechanisms through which social interactions may create spillovers. They define "learning interactions" as social interactions through which knowledge is transferred. In the classroom context, this may be group work in which students help one another understand course material. Learning interactions also include students learning about new trends, or what is (or is not) socially acceptable. Having classmates who learn new skills may make it easier to learn those skills – through group work or hearing a student's perspective on the material. Having friends who learn new skills may make it socially acceptable to learn those skills as well.

In addition, a student's peers may shape their beliefs about the returns to effort. Having motivated friends with positive views of school or teachers may lead students to view them positively as well. Consider a model in which there are two potential states: education is worthwhile and education is not worthwhile. Which of these states is true remains unknown to students during their education, but students update their beliefs as they learn the beliefs of their friends (Degroot, 1974, Patacchini et al., 2017).

"Stigma interactions" (Glaeser and Scheinkman, 1999) may be most relevant within friend groups. When behavior is stigmatized, students who deviate from accepted norms face punishment from the group. As more students deviate, the costs of deviation reduce. We may think about stigma in a signaling model, where a student wants to be accepted by the group and alters their behavior to gain that acceptance. In fact, the effect of peers and friends may be to add costs to students for their academic achievement. (Austen-Smith and Fryer, 2005, Richmond et al., 2012, Bursztyn et al., 2019).[5]

This paper remains agnostic to the specific mechanisms at play. The preceding discussion focuses on mechanisms that work through friendship groups, but there are undoubtedly other social effects in the classroom. For example, environmental factors such as disruptive peers

[3]This is not quite true, as pointed out in Hanushek (1986), but at minimum this can be a useful simplifying assumption to fix concepts.

[4]Lavecchia et al. (2016) illustrate this by writing "A 6-year-old does not go to school because she wants a better life. She must be persuaded that school is fun now, or given no better option [...] the 6-year-old is simply not conditioned to think about long-run consequences from immediate actions." They later note that similar myopic behavior continues at least into adolescence – well beyond the age of students in this paper.

[5]For example, under the "acting white" hypothesis, minority students incur a social cost for academic achievement (Austen-Smith and Fryer, 2005). In Richmond et al. (2012), students face competing pressures from "deviant" and "nondeviant" peers. Bursztyn et al. (2019) models two potential peer cultures. In one, students stigmatize effort. In the other, students value innate ability. The effect of both cultures is that students are disincentivized from publicly studying.

(Carrell et al., 2018) likely play a role but would not be captured in a friendship network. To get a full picture of social effects in a classroom, multiple model types need to be explored. In this paper, I innovate on our ability to determine who is a friend by using not only a revealed friendship network, but also the strength of connection with each friend.

2.2 Which Peers Matter?

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. A researcher decides which peers are relevant for each student and the structure of that network. This is known as the student's reference group. 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 focus of the education peer effects literature is the effect of peers on test scores. It is important to note that estimates vary widely[6] due to difference in context, methodology, and reference group definition. The discussion that follows focuses on reference group definitions used in the literature, but Vigdor and Ludwig (2010), Epple and Romano (2011), Sacerdote (2011), and Paloyo (2020) give excellent reviews of the academic peer effects literature broadly.

We see notable variation in who researchers include in the reference group. Researchers frequently use cohort (school-grade level) as a reference group in order to avoid concerns around within-school sorting into classrooms (Hoxby, 2000, Vigdor and Nechyba, 2007, and Carrell and Hoekstra, 2010). Burke and Sass (2013) find cohort effects near zero using a student fixed effect model incorporating average peer performance. However, they find that classroom peers produce meaningful peer effects. Their results suggest 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 students in other classrooms generally have fewer opportunities to interact and influence behavior than classmates do. Incorporating unimportant peers who contribute little to the outcome should bias estimates towards zero. Thus for questions of whether peer effects exist, detecting peer effects at the cohort level may be sufficient. But it becomes important to model the network as precisely as possible in order to measure the size of peer effects.

[6] In Sacerdote (2011), the results in Table 4.2 illustrate the wide range of estimates we see in the literature: 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. The papers I highlight here show the range of estimates, but most in the literature fall near the middle range.

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. Examples include assigned seat neighbors (Hong and Lee, 2017) and college roommates (Sacerdote, 2001, Zimmerman, 2003, and Stinebrickner and Stinebrickner, 2006). Notice that even as these reference group definitions may better capture members of the peer group, connections are still binary.[7] There is a tension here, which is that as the reference group narrows and the target group is more likely to be peers, more actual peers are almost certainly missed as well. Both of these issues are mediated under a scenario in which peers are revealed by student behavior and when the measure of connection is continuous.

Other work also explores the effect of homophily on peer effects by breaking the environment into groups based on shared characteristics. Arcidiacono and Nicholson (2005) study peer effects in medical school cohorts. They break up the cohort and investigate heterogeneous effects based on shared race and shared gender, finding effects along gender lines but not along racial ones. Lavy et al. (2012) finds girls are affected by high performing peers, while boys are not. Similarly, Fletcher et al. (2020) explores within-gender friendships and finds girls but not boys perform better when their friend have college educated mothers.

The work discussed so far focuses on how researchers have defined the peer environment (ex: classroom or cohort), and how that environment can be sliced to focus on relevant subgroups. Another way to define the reference group is to look beyond the school, cohort, or classroom entirely. This is possible when information on the network structure itself is observed. 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.[8] The friendship survey allows researchers to define a set of relevant peers from the school (networks are within school, rather than 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).[9]

[7]Sacerdote (2001) notes that roommates are only part of a student's social network, and his estimates using randomly assigned roommates mark a "lower bound" on total academic peer effects. However, there is also reason not to believe that roommates are likely to be meaningful members of a student's peer group. Stinebrickner and Stinebrickner (2006) has a nice discussion of these and related potential concerns.

[8]A litany of papers have been written using Add Health, including Lin (2010), Bifulco et al. (2011), Lin and Weinberg (2014), Hsieh and Lin (2017), 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.

[9]There are 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. Hsieh and Lin (2017) explores connection variation based on gender and racial homophily. Lin and Weinberg (2014) explores differences in effect based on whether the friendship was reciprocated or not. In each of these, connections are categorized differently (ex: a separate network is used for long and short term peers in Patacchini et al. (2017)), but remain binary.

Most work focuses on the scope of the reference group - defining (or assuming) who is and who is not a relevant peer. Little work measures connection strength - how influential each relevant peer is. Lin (2010) points out that "the ideal model should contain the weighted average of the peer variables, with weight determined by the importance of a friend, as opposed to a mean peer variable" before noting that this sort of data is not common.

In this paper, I define the scope of the reference group as the classroom. I then measure the intensity of connection between classmates and use this network structure to determine social spillovers on math and reading test scores. In order to measure connection intensity, I use administrative point of service data from the New York City Department of Education to observe daily lunch transactions. I 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.[10] This means that friendships are revealed rather than stated, and their revelation means that friendships need to be reciprocated.[11] Additionally, friendships change during the school year, and we observe the result of daily decisions students make rather than a single survey snapshot. Second, the 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,[12] and I 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 perfect collinearity between group mean characteristics and mean expected group outcome (the reflection problem).[13] This enables us to disentangle the social effect from the contextual effect. This is a policy relevant distinction, as the social effect is a multiplier and the contextual effect is not.[14]

[10] There is some concern that friendship surveys may not be accurate. They could be aspirational (which could explain unreciprocated friendships in Lin and Weinberg, 2014). Landini et al. (2016) provides evidence of unreliability in friendship surveys by comparing a student friendship survey to a survey of who their parents think are their friends. Both display a significant number of unreciprocated friendships (as in Add Health), although the parent survey displays more.

[11] Lin and Weinberg (2014) use Add Health to show that reciprocated friendships are stronger than unreciprocated friendships on a variety of outcomes, including academics.

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

[13] As described in Manski (1993), and in section 5.1.

[14] 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

3 Data and Sample Construction

3.1 Data

This paper uses student-level administrative data on student academic performance, school and classroom codes, and socio-demographic characteristics from the 2018 academic year, with lag outcomes coming from the 2017 academic year. I link this to student lunch transactions at the point of service (POS) for 2018.

Student-level demographic data comes from the New York City Department of Education (NYCDOE). These data include socio-demographic characteristics such as gender, race, age, grade, residential zip code, an indicator of eligibility for free or reduced-price lunch, and whether a student is an English language learner. Data also include class assignment as well as both current year and lagged mathematics and reading test scores.

By academic year 2018 the POS system was in 93.4% of elementary schools, serving 95.5% of grade 1-5 students. POS data indicate the exact timing of student lunch transactions (precise to the second) for each day in the school year. I 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 proximity to one another in the lunch line. 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. I provide additional background on the POS system in Appendix C.

3.2 Sample

The sample is taken from the universe of students in the NYC public schools for academic year 2018. I examine elementary school students for two main reasons. First, elementary school students typically remain with the same set of classmates for the school day, as opposed to middle and high school students who tend to switch classes. Second, elementary school students are more likely to participate in the school lunch program than middle or high school students. This is likely because 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 my sample is 65.8%. I limit analysis to students in general education classrooms. Additional sample selection details can be found in Appendix D.

girls is important (the peer effect), or if it is related to, for example, behavioral differences where young boys demand more teacher time (a contextual 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.

Table 1 reports summary statistics for this sample. Test scores are normalized z-scores across grade level in the school district, so the sample is slightly higher performing than average. Average class size in my 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.

4 Network Construction

This paper uses a novel approach to measure contact between students and reveal the classroom friendship network. I observe the timing of every lunch transaction in the POS system each day during the school year, precise to the second. I transform this daily timing data into orderings of students in the lunch line. This allows us to think about distance as the physical proximity of students to one another. Because lunch is a relatively unstructured and social time, a primary concern for students is who they can socialize with in the line and then during lunch. I transform the observed order into a social distance by looking at whether any two students i and j are within some threshold distance (number of students) of one another. My baseline model uses a threshold distance of one - whether two students are next to one another in line. For robustness, I also look at larger threshold distances in Section 7.2.

The next step is to zoom out to the full year, such that students frequently observed near one another on individual days are considered friends. Some students do not participate in school lunch every day (or are absent from school), complicating this process. 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).

I start constructing the network by averaging daily observations together, akin to what De Giorgi et al. (2010) do with classes.[15] This proximity matrix represents the percent of days each student is near each other student. However, students with low participation will appear to have artificially low connections using this measure. To address this, I row-normalize the average proximity matrix such that each row sums to one. [16] Row-normalization changes the interpretation of the network. The denominator changes from the number of school days to

[15]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. We can think of the networks in this paper as a large number of partially overlapping networks (for each day of the school year). To my knowledge, this is the largest set of partially overlapping networks used, and the result is a network of connections that are essentially continuous.

[16]Row normalization is common in the literature because it improves interpretability of results by weighting the reference group into percentage terms. Each row i indicates influence due to student i's relevant peers.

the number of days student i is present. 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, I have addressed the issue of not observing who a student would choose to be near if they did attend. However, it is not clear that I 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, I construct a proximity matrix for the percent of times we observe students near one another when both are present:

$$p_{ij} = \frac{\sum_{d=1}^{D} S_d(i,j)}{\sum_{d=1}^{D} \delta_d(i,j)} \text{ for } i \neq j \text{ ; and } p_{ij} = 0 \text{ for } i = j \tag{4.1}$$

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 above, for estimation I create the weighting matrix W by row-normalizing proximity matrix P such that each row sums to one. [17] In Appendix F, I provide a simple illustration showing how I convert daily observations into a weighting matrix.

Admittedly there are other ways I could construct the proximity matrix, and there are potential concerns with this method. Perhaps most concerning is that low-participation students could appear overly important for those they stand near when they do participate. I address this using an alternate proximity matrix in Appendix G.2.

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.

The line-up process itself remains unobserved, but I discuss how students may choose

[17] While averaging the daily observations as in De Giorgi et al. (2010) and the method described in equation 4.1 lead to different proximity matrices, row normalization makes the resulting weighting matrices identical.

their location in line in Appendix E. Additionally, I provide evidence that students have at least some agency over this decision by considering whether students are in roughly the same order each day. To do this, I 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 1 shows the distribution of M and that the bulk of classrooms (average measure value is 0.203) are closer to a uniformly random distribution (value of 0.25) than fully ordered (value of zero), but that there is more order than there would be under complete randomness.[18] 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.[19] In Section 7.3 I remove these classrooms as a robustness check and test other thresholds.

5 Methods

5.1 Identification

Identification of peer effects is notoriously difficult due to the reflection problem and selection into peer groups. In this section, I discuss these threats and how I address them in this paper.

In his seminal paper, Manski (1993) discusses three different effects which may be captured in a naive model of peer effects and relates these effects to the reflection problem. The first is the social effect,[20] which is often the effect of interest to researchers and policymakers. The social effect results when one student's performance affects the performance of another. For example, we observe a social 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 social effect is a multiplier, capturing spillovers due to interaction. The second effect is the contextual effect, which controls for student characteristics in the reference group. For example, we might expect that wealthier students

[18] Appendix B outlines measure behavior under changes in class size, participation rate, and randomness.

[19] 134 of 4,077 classrooms are below the 0.1 threshold.

[20] Manski, 1993 calls this the endogenous effect. The endogenous effect is so named because it directly places the outcome vector 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 5.2 and about the estimation procedure in Appendix A. This paper uses the term social effect but these terms are equivalent.

perform better on tests, all else equal. As a result classroom performance may increase with wealth, but this is due to student characteristics rather than student interactions. The final effect Manski (1993) 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 these factors are not related to the students or any social interactions between them.

In many reduced form models of peer effects, we cannot distinguish between contextual 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 social and contextual effects (Bramoullé et al., 2009). This is because the collinearity issue arises when individuals share a reference group. In this paper, 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. Using a network composed of individual-level reference groups in a SAR model allows separate identification of social and contextual effects.[21]

Correlated effects are commonly addressed using fixed effects (Bifulco et al., 2011, Ajilore et al., 2014, Lin, 2015, Horrace et al., 2016), and I 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. Additionally, there are a number of other potential correlated effects. These can arise if certain types of students are differentially treated, either directly or through shared experience. For example, gender or racial groups may experience discrimination or differing levels of expectation and support. I include fixed effects for several groups, including racial, gender, free or reduced price qualification, and English language learner status.

The selection problem occurs when students select into peer groups with similar characteristics, such that outcomes may be correlated. This issue can also be considered an omitted variable problem. In this paper, there are two levels of selection to consider: assignment into the classroom[22] and into the friendship network. Selection into the classroom is the result of

[21] 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 my 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.

[22] This issue of how students are sorted into the classroom has been explored in other literatures as well,

student residential location[23] and administrator decision. I test whether student characteristics explain classroom assignment and show that class assignment based on observed student characteristics is consistent with randomness in Appendix H.[24]

Within the classroom, we could observe correlations between friends' test scores if students select friends based on academic performance, falsely identifying these correlations as peer effects. I do two things to address this. First, I use group fixed effects, which is a common method used to mitigate the problem of endogenous group formation.[25] The idea is to incorporate group effects for shared characteristics along which students may form friendships (homophily). In addition to classroom fixed effects, I include gender and race fixed effects as well.[26] However, neither of these are directly linked to academic performance, and it remains possible that a student's academic performance affects their friendship choices. For example, high ability students may choose other high ability students as their friends, resulting in an upwardly biased estimates if I do not control for peer ability. To address this, I control for both a student's own lagged performance and the lagged performance of their friends. This reduces the scope of the selection concern. Rather than concern that high ability students associate with one another, we are now concerned that students with high performance gains associate with one another, while controlling for a classroom effect.[27] In Section 7, I conduct

such as the teacher quality literature. Here, controlling for lagged student performance appears to mitigate nearly all selection bias due to student assignment to classrooms (using an extensive set of controls and quasi random assignment Chetty et al., 2014, with random assignment Kane et al., 2013). The models used in this paper include both a classroom fixed effect and lagged student performance, making them comparable to value-added models of teacher effects.

In the teacher quality work, one is concerned that students are non-randomly assigned into classrooms. If high achieving students are consistently assigned to a teacher, that teacher may appear a better teacher than they are. In peer effects work, we are again concerned that students are non-randomly assigned into classrooms. If high performing students are assigned to the same classroom, the correlated performance of these students may erroneously be attributed to peer effects. By controlling for own and peer past performance, as well as the classroom correlated effect (which we can think of as a teacher effect), we mitigate this sorting issue.

[23]Exercising school choice is not uncommon in NYC public schools, but choice of school is based at least in part on residential location. Mader et al. (2018) describes the state of school choice in New York City. They find that the decision to opt into a choice school is heavily based on student residential location and the quality of their zoned school. I include residential zip code in an attempt to capture the similar choices faced by neighboring students. Explicitly modeling school choice is beyond the scope of this paper.

[24]Of the tested variables, only English Language Learner (ELL) status appears not random. This is likely because public schools in New York state offer programs for ELL students, including transitional bilingual and dual language programs. It is also important to note that even when students are systematically placed into classrooms (or tracked), this assignment process is not determined by the student. Students select friends based on the choice set they are given – in this case, that is their classmates.

[25]This is particularly common in work using Add Health or other explicit networks, including Bramoullé et al. (2009), Lin (2010), Bifulco et al. (2011), Lin and Weinberg (2014), Hsieh and Lin (2017), and Patacchini et al. (2017).

[26]Horrace et al. (2021) explores several factors of homophily. Gender and race are two of the strongest factors along which NYC elementary students sort.

[27]It is plausible that students know (and sort based on) their classmate's performance levels, but less plausible that they know their classmate's performance gains. For example, students may compare quiz

two robustness checks related to these selection issues. One is an alternative specification, under which the results remain largely unchanged. In the other, I consider the evolution of friendships during the school year. I find little evidence that students change their friendship networks based on academic performance.

5.2 Baseline Model

This paper uses a revealed friendship network to measure academic spillovers in the classroom. For the baseline model, I construct a within-classroom network according to equation 4.1. In the linear in means model I use, this network is multiplied by both the outcome Y and student characteristics X so that we can separately identify endogenous and contextual effects. Below is the basic format of the linear in means model I estimate:

$$Y = \alpha + \lambda WY + \theta WX + X\beta + \phi + U \tag{5.1}$$

where Y is the outcome of interest, α is a constant, W is the weighting matrix as defined in Section 4, ϕ 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 contextual 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 contextual effect because it captures the spillover effects resulting from social interaction, whereas the contextual component controls for student characteristics. References to estimates of the peer or social effect refer to this endogenous effect.[28]

I estimate this model using Maximum Likelihood Estimation and follow Horrace et al. (2021) and Lee and Yu (2010). Details of the estimation procedure are found in Appendix A.

results, but may be unlikely to compare score improvements over the previous quiz.

[28]It is important to note that estimates of the social effect from model 5.1 are multiplier effects. This means that interpretation of the estimated structural parameter $\hat{\lambda}$ is done by converting the result as below:

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

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.

6 Results

6.1 Baseline Results

Table 2 reports 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. The 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. I also include these zip codes in the contextual effect. The first line of Table 2 shows a math peer effect for students in the lunch line together of 0.078, which is statistically significant. As discussed in Section 5.2 this is a multiplier effect, and so we interpret this as a multiplier of 1.085 or an increase of 0.085 units. The peer effect for reading is also significant, but smaller at 0.043, or a multiplier of 1.045. The fact that both estimates of the peer 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 8.5% of a standard deviation. This is approximately 90% of 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 35% 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.[29] This is mirrored in low performers.[30] I 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 I row normalize, 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,

[29] 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.

[30] 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

rather than an average student, is 0.026 in math and 0.014 in reading. This is equivalent to nearly half the effect of poverty and over one quarter 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 quarter the effect of poverty and 11% of the black-white test score gap (the gap 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 contextual 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 0.18 standard deviations as opposed to having all Hispanic friends (the baseline reference group), all else equal. Notice that the contextual effect is not a multiplier effect, but simply shows the effect 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, where gender of friends is also important.

6.2 Evolution of the friendship network over time

I construct the friendship network using repeated observations of the lunch line throughout the school year. This allows me to explore different periods within the school by defining separate networks for each period in the school year. I explore how the influence of the friendship network on test scores varies as the year progresses. 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 may still be getting to know one another at the start of the year, so we may observe more noise as students sort into friendships.[31] Additionally, testing occurs towards the end of the school year, so we might

[31] We may expect more of this kind of noise in larger schools, where students are less likely to have already shared a class, or for students who are new to the school.

expect connections closer to the test date are most important.

Table 3 divides the year into between two and six periods in order to compare friendship importance over time. When dividing the year into halves, as in columns (1) and (6), I construct two proximity matrices according to equation 4.1. The first matrix uses the set of days from the first half of the year, and the second uses the days from the second half of the year. When dividing the year into additional periods, each matrix is constructed from the corresponding set of lunch line observations (days). This means that as we move between columns, the length of time represented by Period 1 is decreasing. In model (1), Period 1 represents about 90 school days, and in model (5) it represents about 30 school days.

We now discuss which set of friends is most influential on standardized test scores. Table 3 marks in bold the period in which the standardized test occurred. In models dividing the year into four or fewer periods (models 1-3 for math and 6-8 for reading) the first period appears important. This effect becomes insignificant when I include additional periods, but the test period remains significant throughout. Periods after the test never influence test scores. This supports my claim that these are peer effects, as we would not expect student behavior after a test to influence the test scores unless we were also picking up another signal. These results suggest that while there are important long-term effects of peers (Carrell et al., 2018), there are also important short term effects as well. That is, to the extent that friendships evolve, past friends likely have less relevance than current friends.[32] Several of the mechanisms discussed in section 4 likely have short term salience. For example, students may share beliefs about the importance of the test, and update these beliefs based on friends up to the testing date (Patacchini et al., 2017). Similarly, social norms around the acceptability of effort may not develop until closer to the test date Bursztyn et al. (2019). Future work will further explore changes in the network over time.

7 Robustness Checks

Below, I discuss several robustness checks. I conduct a placebo test using random lunch line pairings, test additional distance measures and definitions, and remove classrooms in which students may not have agency to order themselves. I conduct additional robustness checks in Appendix G. Specifically, I relax some sample restrictions, limit to schools with only one POS terminal, explore alternate proximity matrix definitions, look at selection on ability, examine whether there are differences in effect when more students shared a classroom the previous year, and address a concern that friendships may endogenously evolve based on performance.

[32]Friends who remain in both periods may be most influential (Patacchini et al., 2017), but my measure of social connection should appropriately rank long-term friends with other friends in the current period.

7.1 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 I estimate is simply a result of the students sharing a classroom, I randomly shuffle the lunch line for each day of the year and re-estimate the model. Results are found in Table 4. The placebo estimate for math is a statistically insignificant 0.004 and for reading it is also insignificant at 0.016. While the social effect (as well as all the components of the contextual effects) are insignificant, the own effects perform similarly to the baseline model. I 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.

7.2 Alternate Distances in the Lunch Line

When constructing the baseline network, I modeled connections between students who are directly 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. I explore distances greater than one to test whether this is the appropriate group. Table 5 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. While students who are further apart in line may also be friends, this signal is too noisy to be meaningful.

It is possible to model higher distance thresholds within a single network (rather than putting each distance into its own separate network). Table 6 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 point estimates for both math and reading increase and then decline. Standard errors are increasing for both outcomes, indicating additional noise from the inclusion of students who do not matter. These models are able to pick up larger friend groups, but they have higher potential to include non-friends as well. While Table 5 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.

7.3 Removing potentially ordered classrooms

Some teachers may not give students agency over their position in line, assigning a specific order to their students. I introduced a measure M of within-classroom noise in Section 4, which I discuss in greater detail in appendix B. M ranges between 0 under perfectly consistent ordering and 0.25 under uniform randomness. Figure 1 shows the distribution of M for all classrooms. The mean classroom has a noise measure of 0.203. Classrooms with low noise measure M indicate few changes in the ordering of students in the lunch line throughout the year. This could indicate external interference with a student's ability to choose their location in the lunch line. Appendix B further discusses properties and behaviors of the measure M.

Table 7 shows the results of removing classrooms which exhibit small levels of variation in the observed line order. I 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. In each case, the peer effect increases in M, indicating that when students have the freedom to choose who they stand next to in the lunch line, the students they choose are more influential.

I 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 time spent together in the lunch line, 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.

7.4 Discussion

This section compares my results with other findings in the literature. It is important to note that estimates vary due to differences in context, methods, and definition of the reference group. The focus of this paper is the definition of the reference group, but to my knowledge there is no other work that explores an explicit friendship network for elementary school students.[33] This makes comparison challenging, as there will be differences in at least one of context or methods, as well as how the reference group is defined.

Lin (2010) uses the Add Health survey to estimate peer effects on GPA. She finds that an

[33] Landini et al. (2016) is the closest I know, as they survey students in Italian primary schools to construct a friendship network. However they do not use this network to estimate peer effects, but rather explore differences in parent and student perceptions of school social networks.

exogenous increase in peer performance by one standard deviation increases own performance by of 0.22 standard deviations. The methodology Lin uses is similar to that in this paper, but the context is quite different. Students in Add Health are middle and high school students (grades 7-12), so they are older than the elementary age students in this paper, and the scope of the network is school-level rather than within-classroom.[34]

Burke and Sass (2013) explore classroom effects for fourth and fifth graders in Florida public schools using a student fixed effects model. Similar to my results, they find larger effects in math (0.0292) than reading (0.0271), although the difference is less pronounced than in this paper. These effects are smaller than those presented here, but this could be due to the underweighting influential peer and overweighting unimportant peers that necessarily occurs when constructing reference groups at the classroom level.

It is possible that by narrowing in on the relevant peers exclusively addresses the concern of overweighting unimportant peers. Hong and Lee (2017) use assigned seats for Korean college students to show a hierarchy of which students are relevant. The pairs assigned to sit next to one another are most relevant, with influence decreasing for students sitting further away. They estimate peer effects between 0.02 and 0.03.[35] It is important to note the similarity in these estimates with my own discussion around the influence of a best friend. The key estimates in Hong and Lee (2017) involve only one other student in the classroom.

Lin and Weinberg (2014) also explores within-network heterogeneity in friendship effects, again using the Add Health data. They estimate effects for separate networks of friends in which connections are reciprocated, are nominated but not reciprocated, and are not chosen at all. As expected, the strongest effect is for students who reciprocate connections (0.17) followed by nominated friends (0.12), and finally peers that are not chosen (0.07). Notice that even the peers who are not nominated are found to be important for peer effects, although the effect is smaller.[36] This highlights the importance of not only exploring highly influential peers, but the network as a whole.

A single revealed friendship network that appropriately weights all classmates provides a view of friendship within the classroom. It takes into account the number of strong and weak connections for each student, such that we get a clear picture of the average friendship effect in the classroom. There is benefit to separating networks to understand influencers,[37] as well

[34] Students in Add Health may select into similar classes, but this is not observed. Grades may be correlated with teachers (ex: a teacher may be more or less generous in their grading), and this is also not observed. These factors may lead to selection bias in these data that increases the correlation in academic outcomes between students. This issue may be less of a concern for behavior-based outcomes (such as smoking).

[35] Specifically, they model the effect of their neighbor's midterm score on their final score in the class.

[36] Their models include network fixed effects, which alleviates concern that we may simply be seeing within-school correlations (ex: students in wealthier schools doing better). However, these estimates do not include a contextual effect, although they note that their results are robust to its inclusion.

[37] Some mechanisms of interest involve exposure to student types, such as bad apple or shining light models.

as to modeling the network as a whole.[38] To my knowledge, this is the first paper to measure connection strength between students and estimate the peer effect on academic outcomes.[39]

8 Conclusion

Models of peer effects and social networks typically define reference groups using group participation (ex: cohort or classroom) or a friendship network (ex: Add Health). These define the scope of the reference group, but without incorporating connection strength they necessarily underweight important peers and overweight important peers. In this paper, I construct a revealed friendship network using repeated observations of which students choose to stand near one another in the lunch line.

I use this revealed friendship network to separately identify social and contextual effects and find evidence of significant social effects on both math and reading test scores. These effects are stronger in math than in reading. In math, a one standard deviation improvement in peer performance increases own performance between 7.5% and 11.1% of a standard deviation. This is a large effect, on par with my 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. In addition, I exploit the daily nature of this data to look at the evolution of these connections over time and find evidence that social connections which occur during the test period are the most influential connections on test scores.

Many researchers and policymakers believe peer effects are consequential for a number of important policy discussions such as tracking, school choice, and school integration. In order to effectively weigh the costs and benefits of these policies, accurate estimates of 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, understating the influence of peers. Methods that measure the strength of social connections are critical for understanding how social networks operate.

[38] An interesting avenue for future work combines these different mechanisms, such as friendship and disruption, in the same model. Taken together, these different models provide a fuller picture of the complexities involved in modeling social spaces, and allows for more complex questions as well. For example, what is the effect of being friends with, as opposed to simply being classmates with, a disruptive peer?

[39] De Giorgi et al. (2010) develops variation in connection strength using the number of shared classrooms college students attend. They estimate the peer effect on major choice.

References

Ajilore, Olugbenga, Aliaksandr Amialchuk, Wei Xiong, and Xinyue Ye (2014). "Uncovering peer effects mechanisms with weight outcomes using spatial econometrics". In: *The Social Science Journal* 51.4, pp. 645–651.

Akerlof, George A. and Rachel E. Kranton (Dec. 2002). "Identity and Schooling: Some Lessons for the Economics of Education". In: *Journal of Economic Literature* 40.4, pp. 1167–1201.

Arcidiacono, Peter and Sean Nicholson (Feb. 2005). "Peer effects in medical school". In: *Journal of Public Economics* 89.2-3, pp. 327–350.

Austen-Smith, David and Jr. Fryer Roland G. (May 2005). "An Economic Analysis of "Acting White"*". In: *The Quarterly Journal of Economics* 120.2, pp. 551–583.

Bifulco, Robert, Jason M. Fletcher, and Stephen L. Ross (Feb. 2011). "The Effect of Classmate Characteristics on Post-secondary Outcomes: Evidence from the Add Health". In: *American Economic Journal: Economic Policy* 3.1, pp. 25–53.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009). "Identification of peer effects through social networks". In: *Journal of Econometrics* 150.1, pp. 41–55.

Brand, David (2019). *Why do Some NYC School Kids Still Eat Lunch Before Some of Us Have had Breakfast?* City Limits.

Burke, Mary A. and Tim R. Sass (2013). "Classroom Peer Effects and Student Achievement". In: *Journal of Labor Economics* 31.1, pp. 51–82.

Bursztyn, Leonardo, Georgy Egorov, and Robert Jensen (May 2019). "Cool to be Smart or Smart to be Cool? Understanding Peer Pressure in Education". In: *The Review of Economic Studies* 86.4, pp. 1487–1526.

Calderon, Valerie J. (2017). *How to Keep Kids Excited About School*. Gallup.

Carlson, Kurt A. and Jacqueline M. Conard (2011). "The Last Name Effect: How Last Name Influences Acquisition Timing". In: *Journal of Consumer Research* 38.2, pp. 300–307.

Carrell, Scott E., Mark Hoekstra, and Elira Kuka (Nov. 2018). "The Long-Run Effects of Disruptive Peers". In: *American Economic Review* 108.11, pp. 3377–3415.

Carrell, Scott E. and Mark L. Hoekstra (Jan. 2010). "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids". In: *American Economic Journal: Applied Economics* 2.1, pp. 211–28.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (Sept. 2014). "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates". In: *American Economic Review* 104.9, pp. 2593–2632.

Cobb-Clark, Deborah A. and Stefanie Schurer (2012). "The stability of big-five personality traits". In: *Economics Letters* 115.1, pp. 11–15.

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

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli (Apr. 2010). "Identification of Social Interactions through Partially Overlapping Peer Groups". In: *American Economic Journal: Applied Economics* 2.2, pp. 241–75.

Degroot, Morris H. (1974). "Reaching a Consensus". In: *Journal of the American Statistical Association* 69.345, pp. 118–121.

Elliott, Marc N., Peter A. Morrison, Allen Fremont, Daniel F. McCaffrey, Philip Pantoja, and Nicole Lurie (Apr. 2009). "Using the Census Bureau's surname list to improve estimates of race/ethnicity and associated disparities". In: *Health Services and Outcomes Research Methodology* 9.2, p. 69.

Epple, Dennis and Richard E. Romano (2011). "Chapter 20 - Peer Effects in Education: A Survey of the Theory and Evidence". In: ed. by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson. Vol. 1. Handbook of Social Economics. North-Holland, pp. 1053–1163.

Fletcher, Jason M., Stephen L. Ross, and Yuxiu Zhang (2020). "The consequences of friendships: Evidence on the effect of social relationships in school on academic achievement". In: *Journal of Urban Economics* 116, p. 103241.

Glaeser, Edward and Jose Scheinkman (1999). *Measuring Social Interactions.* Harvard Institute of Economic Research Working Papers. Harvard - Institute of Economic Research.

Hanushek, Eric A. (1986). "The Economics of Schooling: Production and Efficiency in Public Schools". In: *Journal of Economic Literature* 24.3, pp. 1141–1177.

Hong, Sok Chul and Jungmin Lee (2017). "Who is sitting next to you? Peer effects inside the classroom". In: *Quantitative Economics* 8.1, pp. 239–275.

Horrace, William C., Hyunseok Jung, Jonathan Presler, and Amy Ellen Schwartz (2021). "What makes a classmate a peer? Examining which peers matter in NYC". Working Paper.

Horrace, William C., Xiaodong Liu, and Eleonora Patacchini (2016). "Endogenous network production functions with selectivity". In: *Journal of Econometrics* 190.2. Endogeneity Problems in Econometrics, pp. 222–232.

Hoxby, Caroline (Aug. 2000). *Peer Effects in the Classroom: Learning from Gender and Race Variation.* Working Paper 7867. National Bureau of Economic Research.

Hsieh, Chih-Sheng and Xu Lin (2017). "Gender and racial peer effects with endogenous network formation". In: *Regional Science and Urban Economics* 67, pp. 135–147.

Kane, Thomas J., Daniel F. Mccaffrey, Trey Miller, and Douglas O. Staiger (2013). "Have we identified effective teachers? Validating Measures of Effective Teaching using Random Assignment". In: *Research Paper. MET Project. Bill & Melinda Gates Foundation.*

Landini, Fabio, Natalia Montinari, Paolo Pin, and Marco Piovesan (2016). "Friendship network in the classroom: Parents bias on peer effects". In: *Journal of Economic Behavior & Organization* 129.C, pp. 56–73.

Laschever, Ron A. (June 2013). *The Doughboys Network: Social Interactions and the Employment of World War I Veterans.* Tech. rep.

Lavecchia, A.M., H. Liu, and P. Oreopoulos (2016). "Chapter 1 - Behavioral Economics of Education: Progress and Possibilities". In: ed. by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Vol. 5. Handbook of the Economics of Education. Elsevier, pp. 1–74.

Lavy, Victor, Olmo Silva, and Felix Weinhardt (2012). "The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools". In: *Journal of Labor Economics* 30.2, pp. 367–414.

Lee, Lung-Fei and Jihai Yu (2010). "Estimation of spatial autoregressive panel data models with fixed effects". In: *Journal of Econometrics* 154.2, pp. 165–185.

Lin, Xu (2010). "Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables". In: *Journal of Labor Economics* 28.4, pp. 825–860.

— (Nov. 2015). "Utilizing spatial autoregressive models to identify peer effects among adolescents". In: *Empirical Economics* 49.3, pp. 929–960.

Lin, Xu and Bruce A. Weinberg (2014). “Unrequited friendship? How reciprocity mediates adolescent peer effects”. In: *Regional Science and Urban Economics* 48, pp. 144–153.

Mader, Nicole, Clara Hemphill, and Qasim Abbas (May 2018). *The Paradox of Choice: How School Choice Divides New York City Elementary Schools*. Report. The Center for New York City Affairs.

Mammadov, Sakhavat (2022). “Big Five personality traits and academic performance: A meta-analysis”. In: *Journal of Personality* 90.2, pp. 222–255.

Manski, Charles F. (July 1993). “Identification of Endogenous Social Effects: The Reflection Problem”. In: *The Review of Economic Studies* 60.3, pp. 531–542.

Nye, Barbara, Spyros Konstantopoulos, and Larry V. Hedges (2004). “How Large Are Teacher Effects?” In: *Educational Evaluation and Policy Analysis* 26.3, pp. 237–257.

Paloyo, Alfredo R. (2020). “Chapter 21 - Peer effects in education: recent empirical evidence”. In: *The Economics of Education (Second Edition)*. Ed. by Steve Bradley and Colin Green. Second Edition. Academic Press, pp. 291–305.

Patacchini, Eleonora, Edoardo Rainone, and Yves Zenou (2017). “Heterogeneous peer effects in education”. In: *Journal of Economic Behavior and Organization* 134, pp. 190–227.

Richmond, Melanie J, Robin J Mermelstein, and Aaron Metzger (2012). “Heterogeneous friendship affiliation, problem behaviors, and emotional outcomes among high-risk adolescents”. In: *Prevention Science* 13.3, pp. 267–277.

Roberts, Brent and Wendy DelVecchio (Feb. 2000). “The Rank-Order Consistency of Personality Traits from Childhood to Old Age: A Quantitative Review of Longitudinal Studies”. In: *Psychological bulletin* 126, pp. 3–25.

Ryan, Ronan, Sally Vernon, Gill Lawrence, and Sue Wilson (Jan. 2012). “Use of name recognition software, census data and multiple imputation to predict missing data on ethnicity: application to cancer registry records”. In: *BMC medical informatics and decision making* 12, p. 3.

Sacerdote, Bruce (May 2001). “Peer Effects with Random Assignment: Results for Dartmouth Roommates*”. In: *The Quarterly Journal of Economics* 116.2, pp. 681–704.

— (2011). “Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?” In: ed. by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Vol. 3. Handbook of the Economics of Education. Elsevier, pp. 249–277.

Stinebrickner, Ralph and Todd R. Stinebrickner (2006). “What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds”. In: *Journal of Public Economics* 90.8, pp. 1435–1454.

Vigdor, J. and J. Ludwig (2010). “Neighborhoods and Peers in the Production of Schooling”. In: *Economics of Education*. Ed. by Dominic J. Brewer and Patrick J. McEwan. Elsevier Science, pp. 163–169.

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.

Zimmerman, David J. (Feb. 2003). “Peer Effects in Academic Outcomes: Evidence from a Natural Experiment”. In: *The Review of Economics and Statistics* 85.1, pp. 9–23.

9 Tables

Table 1: Summary Statistics

variable	mean	sd	N
Lunch participation 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
Grade 4	0.489	0.500	102,606
Grade 5	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
z-math	0.082	0.950	101,948
z-read	0.088	0.953	102,244

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

Table 2: Baseline Model

	Math		Reading	
Social Effect:	0.078**	(0.009)	0.043**	(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	0.058**	(0.005)	0.056**	(0.006)
Contextual Effect:				
lag test score	0.001	(0.010)	0.056**	(0.011)
female	-0.006	(0.009)	0.030**	(0.011)
ELL	0.018	(0.021)	0.046	(0.027)
Asian/other	0.123**	(0.019)	0.086**	(0.022)
black	-0.011	(0.020)	-0.038	(0.023)
white	0.046*	(0.021)	0.007	(0.025)
ever poor	0.003	(0.019)	-0.002	(0.023)
N	100,156		94,838	

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

Table 3: Evolution of friendship importance over the school year

	Math				
Social Effect:	(1)	(2)	(3)	(4)	(5)
Period 1	0.040**	0.027**	0.020*	0.014	0.007
	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
Period 2	**0.040****	0.032**	0.022*	0.007	0.016
	(0.010)	(0.011)	(0.010)	(0.010)	(0.009)
Period 3		**0.021***	**0.028****	0.032**	0.017
		(0.010)	**(0.010)**	(0.010)	(0.010)
Period 4			0.006	**0.030****	0.011
			(0.010)	**(0.010)**	(0.010)
Period 5				-0.009	**0.030****
				(0.009)	**(0.010)**
Period 6					-0.009
					(0.009)
N	100,156	100,156	100,156	100,156	100,156

	Reading				
Social Effect:	(6)	(7)	(8)	(9)	(10)
Period 1	0.030**	0.029**	0.021*	0.014	0.010
	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)
Period 2	**0.019**	0.002	-0.001	0.012	0.015
	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)
Period 3		**0.020***	**0.024***	-0.001	-0.008
		(0.010)	**(0.011)**	(0.010)	(0.010)
Period 4			0.002	**0.027****	0.007
			(0.010)	**(0.010)**	(0.010)
Period 5				-0.003	**0.021***
				(0.009)	**(0.010)**
Period 6					0.001
					(0.009)
N	94,838	94,838	94,838	94,838	94,838

Each column presents a model in which I divide the school year into additional periods, such that friends from each period are considered separate networks. Periods in bold denote the period in which the standardized test occured. Models include classroom fixed effects, own zip code fixed effects, and zip code contextual fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

Table 4: Placebo Tests

	Math		Reading	
Social Effect:	-0.016	(0.032)	-0.007	(0.033)
Own Effect:				
lag test score	0.733**	(0.003)	0.661**	(0.003)
female	-0.009**	(0.003)	0.053**	(0.004)
ELL	-0.095**	(0.007)	-0.194**	(0.009)
Asian/other	0.180**	(0.006)	0.159**	(0.007)
black	-0.021**	(0.006)	-0.048**	(0.007)
white	0.077**	(0.007)	0.088**	(0.008)
ever poor	-0.060**	(0.006)	-0.056**	(0.007)
Contextual Effect:				
lag test score	0.062	(0.039)	0.038	(0.042)
female	0.039	(0.043)	0.010	(0.051)
ELL	-0.044	(0.083)	0.025	(0.105)
Asian/other	-0.047	(0.071)	-0.039	(0.082)
black	-0.061	(0.076)	-0.130	(0.088)
white	0.012	(0.077)	-0.018	(0.090)
ever poor	-0.028	(0.068)	0.024	(0.080)
N	100,156		94,838	

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

Table 5: Multiple Distance levels

	Math			
Social Effect:	(1)	(2)	(3)	(4)
D=1	0.078**	0.082**	0.082**	0.081**
	(0.009)	(0.009)	(0.009)	(0.009)
D=2		0.005	0.004	0.004
		(0.012)	(0.012)	(0.012)
D=3			0.007	0.011
			(0.013)	(0.013)
D=4				-0.024
				(0.013)
N	100,156	100,156	100,156	100,156

	Reading			
Social Effect:	(5)	(6)	(7)	(8)
D=1	0.043**	0.045**	0.044**	0.043**
	(0.009)	(0.010)	(0.010)	(0.010)
D=2		0.011	0.015	0.014
		(0.012)	(0.013)	(0.013)
D=3			-0.018	-0.018
			(0.013)	(0.013)
D=4				-0.015
				(0.014)
N	94,838	94,838	94,838	94,838

Models include classroom fixed effects, own zip code fixed effects, and zip code contextual fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level. Columns (1) and (5) are reproduced from Table 2, which defines proximity between students as students standing directly next to one another in the lunch line on a given day. Other columns present results from a similar model that differs only with the addition of networks for distances greater than one. For example, columns (2) and (6) adds a network for students standing in the lunch line with a single student between them. The inclusion of these additional networks coincides with the inclusion of contextual effects for each network.

Table 6: Distances greater than one

	Math			
	D = 1	D $\leq$ 2	D $\leq$ 3	D $\leq$ 4
	(1)	(2)	(3)	(4)
Social Effect	0.078**	0.099**	0.114**	0.112**
	(0.009)	(0.013)	(0.016)	(0.019)
N	100,156	100,156	100,156	100,156

	Reading			
	D = 1	D $\leq$ 2	D $\leq$ 3	D $\leq$ 4
	(5)	(6)	(7)	(8)
Social Effect	0.043**	0.064**	0.062**	0.054**
	(0.009)	(0.013)	(0.017)	(0.020)
N	94,838	94,838	94,838	94,838

Models include classroom fixed effects, own zip code fixed effects, and zip code contextual fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level. Columns (1) and (5) are reproduced from Table 2, which defines proximity between students as students standing directly next to one another in the lunch line on a given day. Other columns present results from a similar model that differs only with the increase of the bandwidth to distances greater than one. For example, columns (2) and (6) considers students friends if they are directly next to one another in the line, or have one student between them.

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

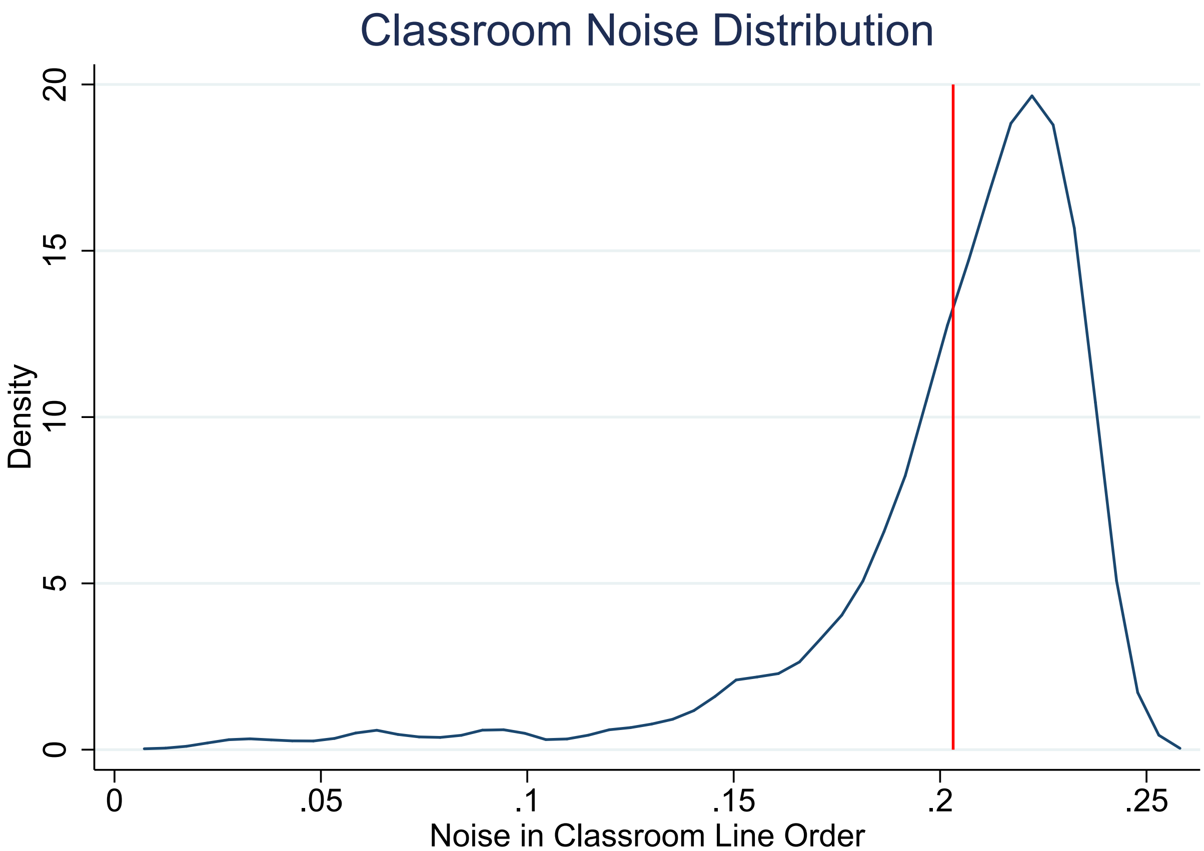
	Math			
	Baseline	M $\geq$ 0.05	M $\geq$ 0.1	M $\geq$ 0.15
	(1)	(2)	(3)	(4)
Social Effect	0.078**	0.083**	0.090**	0.093**
	(0.009)	(0.009)	(0.010)	(0.010)
N	100,156	99,275	97,029	92,962

	Reading			
	Baseline	M $\geq$ 0.05	M $\geq$ 0.1	M $\geq$ 0.15
	(5)	(6)	(7)	(8)
Social Effect	0.043**	0.048**	0.052**	0.056**
	(0.009)	(0.010)	(0.010)	(0.011)
N	94,838	94,022	91,920	88,121

Models include classroom fixed effects, own zip code fixed effects, and zip code contextual fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level. Columns (1) and (5) are reproduced from Table 2. Other columns present results from the same model run on subsamples excluding classrooms with a measure of within-classroom noise (M) less than the specified threshold.

10 Figures

Figure 1: Classroom Noise Distribution



Vertical line indicates mean, which is equal to 0.203. Includes one entry for each of 4,077 classrooms. Density constructed using an Epanechnikov kernel with bandwidth = 0.004.

A Estimation procedures

I 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. I 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, I follow the transformation approach discussed in Lee and Yu (2010) and used in Horrace et al. (2021). This method involves the deviation from the classroom mean operator $Q = \iota\iota'/n$ an $n \times n$ matrix where n is classroom size. I define the orthonormal within transformation matrix Q as $[P, \iota_n/\sqrt{n}]$. Following Lee and Yu (2010), I 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.

I 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:

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

I 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, I 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, I attempt to determine whether there is a large set of classrooms in which students are ordered. The measure I 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, I will detect this line order as having little noise.[40]

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, I 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, I 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 I 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, I 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 $q \cdot B_{ij}$. I 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. I 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, I replace $\min(k, B_{ij} - k)$ with

[40] External ordering according to last name is a common enough phenomenon that it has been deemed "The Last Name Effect" (Carlson and Conard, 2011).

$\binom{B_{ij}-1}{k}^{-1}\binom{B_{ij}-2}{k-1} = \frac{k(B_{ij}-k)}{B_{ij}-1} = \varphi$. 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 I 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 I will show when we finish constructing the measure. As before, I 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.[41] 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, I 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.

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.

[41] 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.

The measure is meant to detect noise, so I 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. I 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.

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.

It is important to note that this measure will be limited if the teacher attempts to be more equitable and alternates (for example) lining their students up alphabetically one day and reverse alphabetically another day, I 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, I 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 The point of sale system

NYCDOE began implementing a 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 C1 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.

Table C1: 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.

D Sample Selection

Table D1 illustrates the sample selection process. I begin with all general education fourth and fifth grade students in schools with a POS system in place for the entire academic year.[42] I cannot measure social connections to students who never participate in school lunch, and 3.7% of students fall into this category. I 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 I lose another 2.3% of students from test non-participation in both math and reading. I drop less than half a percent of students total due to the following three reasons. First, I 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.[43] Second, I 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 the estimates, so I remove these transactions. Third, I 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 I make is students in classrooms with less than 20 students, resulting in the largest loss of students from the sample (15.78%). I choose to look only at classrooms that are larger than twenty students because I am concerned classrooms that appear smaller may be integrated co-teaching (ICT) classrooms, and I do not want unobserved peers.[44]

[42] I restrict to fourth and fifth grade students because standardized tests begin in the third grade, and I include a lag test score in the model.

[43] Some lunch times are even more unreasonable than these bounds we place on lunch times, as in the article Brand (2019) "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.

[44] 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. NYC classrooms are limited to 32 students for elementary students, meaning that if 60% of an ICT classroom is general education, the class would be contain 19 general education students. A cutoff of 20 students allows me to avoid these classrooms, which are structurally different and may contain unobserved students with disabilities. However, as pointed out by a helpful reviewer, this may be too restrictive. In Appendix G.1 I test sensitivity to other cutoffs.

Table D1: 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.

E How do students order themselves?

It is worth discussing the implication of the observed lunch line order, as the ordering process is a black box, and the method of ordering likely varies by classroom. I 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. I then provide evidence that in the majority of classrooms, students have at least some agency over their position in line.

First, I 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, I 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.[45] 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. I do not observe names, and so I cannot test this hypothesis

[45] There is a potential threat to identification here. For example, unobserved personality traits (such as extroversion) may correlate with both academic performance and a preference for position in line (such as being first). In this case, we may see students queue according to personality type rather than friendship. Under this scenario, the network measure connects students with similar traits rather than social connection, potentially introducing bias into my estimates. However, the effect of most of the big five traits on academic performance appears to be small (Mammadov, 2022), suggesting that the associated bias would be small, if it exists. Additionally, it is unclear that personality would be associated with academic performance gains (rather than levels) - particularly if traits are relatively stable over time (Cobb-Clark and Schurer, 2012, Roberts and DelVecchio, 2000).

directly (although I do look at how much strict ordering exists in our sample). If students are ordered based on name, there is little reason to expect these students to be socially more important than other classmates.[46] 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 Horrace et al. (2021) 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 by classroom tables, so that students form a line within a subset of the classroom - they have autonomy within a constraint.[47] 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.

F Example Weighting Matrix Construction

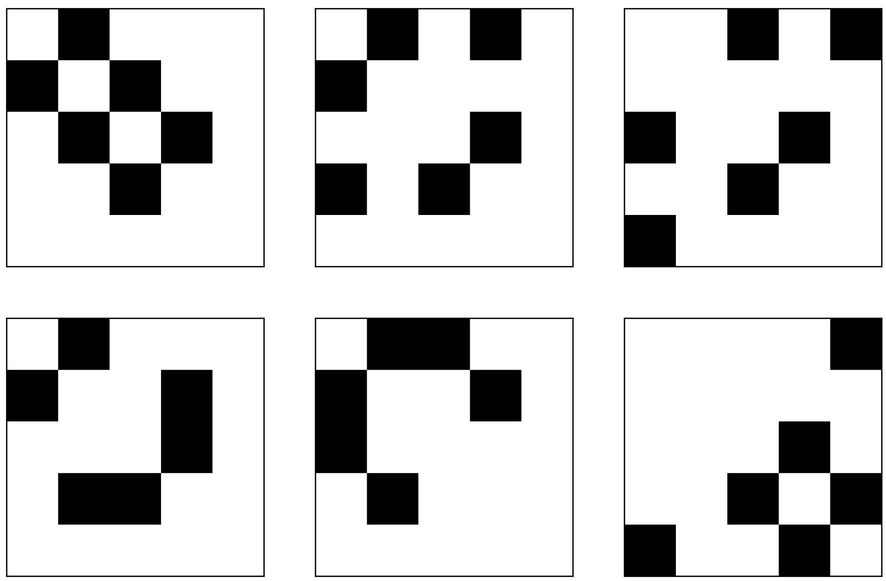
Figure F1 provides a simple example to illustrate how I convert daily observations into a proximity matrix (according to equation 4.1) and corresponding weighting matrix. In the example, we observe five students over six days, and one student is absent or not participating each day. Figure F1a shows 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 F1b applies equation 4.1 to the daily observations, calculating

[46] 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.

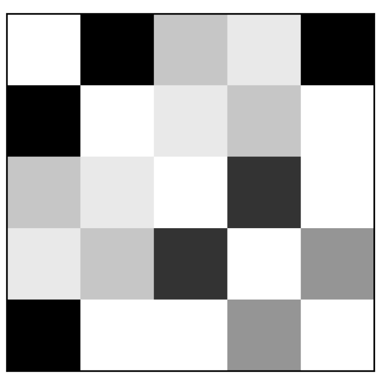
[47] We may think this might elicit a similar peer effect to that in Hong and Lee (2017), in which they measure the peer effect for Korean college students sitting next to each other in classrooms with assigned seats. However, I observe most students near more than just a handful of their classmates in the lunch line over the school year.

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 F1c 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.

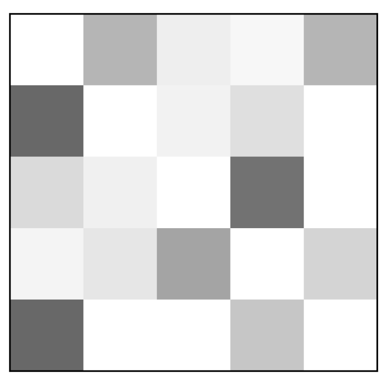
Figure F1: Constructing the Weighting Matrix



(a) Six example daily lunch line observations



(b) Weighting matrix



(c) Row Normalization

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]. (b) Shows the resulting weighting matrix, constructed using equation 4.1 and (c) shows the result of row-normalization.

G Additional Robustness Checks

G.1 Relaxed Sample Restrictions

In the main analysis, I restrict to classrooms that have at least 20 students. As discussed in Section 3, this is to avoid ICT classrooms. I relax this restriction here, and use alternative class size cutoffs to test the sensitivity of my results to this restriction. Table G1 reports results from classrooms with cutoffs of at least 15, 10, and 5 students. In the baseline model, re-reported here in columns (1) and (5), I restrict my sample to classrooms with at least 20 students. In columns (2) and (6), I require at least 15 students. The models in columns (3) and (7) require at least 10 students, and in columns (4) and (8) I require only 5 students in a classroom. Notice that the results are very similar at each cutoff. For Math scores, the social effect ranges between 0.078 and 0.079 and they are not statistically different from one another. For Reading the range is between 0.043 and 0.046. Again, these estimates do not statistically differ. I conclude that my results are not sensitive to the class size cutoff.

G.2 Alternate Proximity Matrix Specifications

There are other ways to define proximity in the lunch line. In this section I examine an alternative proximity definition to address a concern that low participation students may have outsized effects on those near them in the lunch line on the few days they participate. Section 4 describes my 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 on this baseline measure. My 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, I require students to participate a certain number of days together before their connection can be evaluated. The proximity matrix in equation (4.1) is altered such that:

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

where η is the threshold number of days students must both participate in before I 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. A potential drawback of this method is the loss of all or most connections with

Table G1: Sensitivity to Alternate Class Size Cutoffs

	Math			
	CS $\geq$ 20	CS $\geq$ 15	CS $\geq$ 10	CS $\geq$ 5
	(1)	(2)	(3)	(4)
Social Effect	0.078**	0.078**	0.079**	0.078**
	(0.009)	(0.008)	(0.008)	(0.008)
N	100,156	114,390	116,574	117,010

	Reading			
	CS $\geq$ 20	CS $\geq$ 15	CS $\geq$ 10	CS $\geq$ 5
	(5)	(6)	(7)	(8)
Social Effect	0.043**	0.043**	0.046**	0.044**
	(0.009)	(0.008)	(0.008)	(0.008)
N	94,838	112,147	114,257	114,649

I rerun my baseline specification, but test the sensitivity to alternate class size cutoffs. In the baseline model, re-reported here in columns (1) and (5), I restrict my sample to classrooms with at least 20 students. In columns (2) and (6), I require at least 15 students. The models in columns (3) and (7) require at least 10 students, and in columns (4) and (8) I require only 5 students in a classroom. Models include classroom fixed effects, own zip code fixed effects, and zip code contextual fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

low-participation classmates. However the signals we lose may be noisy.

Table G2 reports results for thresholds of between two and ten days that students must participate on the same day before we evaluate their connection. The results remain relatively unchanged in each specification.

Table G2: Students must participate together more than one day

	Math					
	Days ≥ 1	Days ≥ 2	Days ≥ 3	Days ≥ 4	Days ≥ 5	Days ≥ 6
	(1)	(2)	(3)	(4)	(5)	(6)
Social Effect	0.078**	0.079**	0.079**	0.076**	0.075**	0.072**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
N	100,156	100,156	100,156	100,156	100,156	100,156

	Reading					
	Days ≥ 1	Days ≥ 2	Days ≥ 3	Days ≥ 4	Days ≥ 5	Days ≥ 6
	(7)	(8)	(9)	(10)	(11)	(12)
Social Effect	0.043**	0.044**	0.044**	0.043**	0.042**	0.042**
	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
N	94,838	94,838	94,838	94,838	94,838	94,838

Columns (1) and (5) are reproduced from Table 2, which has no minimum requirement for number of days students must stand next to one another in order to consider a connection between them. Other columns present results from a similar model that differ only with the introduction of a minimum number of days. For example, the models represented in columns (2) and (6) eliminate a connection between two students unless they are next to one another in the lunch line for at least two days. This minimum number of days is increased to three in columns (3) and (7), and to four in (4) and (8). Models include classroom fixed effects, own zip code fixed effects, and zip code contextual fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

G.3 Selection on Ability

The models presented so far include group fixed effects for a number of dimensions, including race, gender, and zip code. These are used to mitigate the problem of endogenous group formation.[48] We may be concerned that students select their friends based on ability as well.

[48]See Bramoullé et al. (2009) for a discussion and application. Group fixed effects are common, including in Lin (2010), Bifulco et al. (2011), Lin and Weinberg (2014), Hsieh and Lin (2017), and Patacchini et al. (2017).

If this is the case, some correlation observed between the performance of friends would be due to the choice of students to be friends with classmates who perform similarly, rather than due to social spillovers. I replace the lag test score in my baseline model with a binned lag test score. We can think of lag test score as a measure of ability, and in this way I control for correlations between students with similar ability levels. I run multiple models in which I bin students using between two and ten quantiles, meaning these are groups of between 13 and 2.6 students for the average classroom.

Table G3 reports the results of this exercise. Estimates for the social effect remains large and significant. In math, the effect remains pretty similar, although it is dampened for the larger groups. For reading, the effect is consistently larger for all estimates. The pattern that emerges is that as group size decreases (more quantiles), the estimated social effect increases. This suggests that if fourth and fifth graders select their friends along ability lines, they do so at large levels of aggregation. That is, whether a classmate is in the same half or quartile of students may influence a student's selection of their friends, but students are less likely to care whether another student is at a specifically similar level.

G.4 Cafeteria Layout

The estimates presented in this paper rely on accurately determining the order of students in the lunch line. To construct this measure, I use the timestamp associated with each student's transaction. However, the layout of the cafeteria (which is unobserved) may complicate my measurement. For example, if the same lunch line feeds into multiple POS terminals, we might see a scenario in which two friends were in line together, but they separate to complete their transactions. This separation may allow another transaction to be recorded between the friends' transactions, introducing noise into my measure of social connection.

To partially address this issue, I remove schools from the sample that have more than one POS terminal. This is a smaller sample, containing 56-57% of the students in the baseline models. Table G4 reports these results. For this reduced sample, the social effect for reading is 0.081, which is larger point estimate than the estimate of 0.078 in the baseline model, but this difference is not statistically different. The reading point estimate is increased as well, being 0.056 compared to the 0.043 in the baseline model. This suggests that there may be noise caused by students shuffling between multiple POS terminals when those are present, and my baseline results may understate the true social effect due to these revealed friendship networks. However, these results are no longer representative of NYC as a whole, as schools with more than one POS terminal are, on average, larger than those with only one POS terminal. This increase in the social effect estimate is also consistent with stronger peer

Table G3: Binned Lag Test Scores

	Math							
	Baseline (1)	2 Qtiles (2)	3 Qtiles (3)	4 Qtiles (4)	5 Qtiles (5)	6 Qtiles (6)	8 Qtiles (7)	10 Qtiles (8)
Social Effect	0.078**	0.053**	0.053**	0.049**	0.057**	0.078**	0.082**	0.093**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
N	100,156	100,156	100,156	100,156	100,156	100,156	100,156	100,156

	Reading							
	Baseline (9)	2 Qtiles (10)	3 Qtiles (11)	4 Qtiles (12)	5 Qtiles (13)	6 Qtiles (14)	8 Qtiles (15)	10 Qtiles (16)
Social Effect	0.043**	0.075**	0.061**	0.075**	0.099**	0.087**	0.104	0.127**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.069)	(0.009)
N	94,838	94,838	94,838	94,838	94,838	94,838	94,838	94,838

Columns (1) and (9) are reproduced from Table 2, which uses a continuous lag test score. Other columns present results from similar models that differ by binning the lag test score into between 2 and 10 quantiles. For example, the models represented in columns (2) and (10) include an indicator if the student's lag performance is above the median in the student's classroom. The number of quantiles increases to three in columns (3) and (11), to four in columns (4) and (12), and so on. Models include classroom fixed effects, own zip code fixed effects, and zip code contextual fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

effects in smaller schools or smaller classrooms.[49]

Table G4: Schools with a Single POS Terminal

	Math		Reading	
Social Effect:	0.081**	(0.012)	0.056**	(0.012)
N:	56,576		53,846	

In these models, I restrict my sample to just the schools in which there is a single POS terminal. Models include classroom fixed effects, own zip code fixed effects, and zip code contextual fixed effects. Parameters with * are significant at the 5% level and ** at the 1% level.

G.5 Friends from Last Year

I use classroom assignment to break up existing peer groups from the previous year, and have shown that classroom assignment is consistent with randomness on observed characteristics (Appendix H). But it could be that the connections I observe are simply friendships that carry over from the previous year, as some students in a current classroom surely shared the previous year classroom. I test whether this is a concerning source of selection. To do this, I measure the percentage of students that shared a classroom the previous year. Precisely, the measure is:

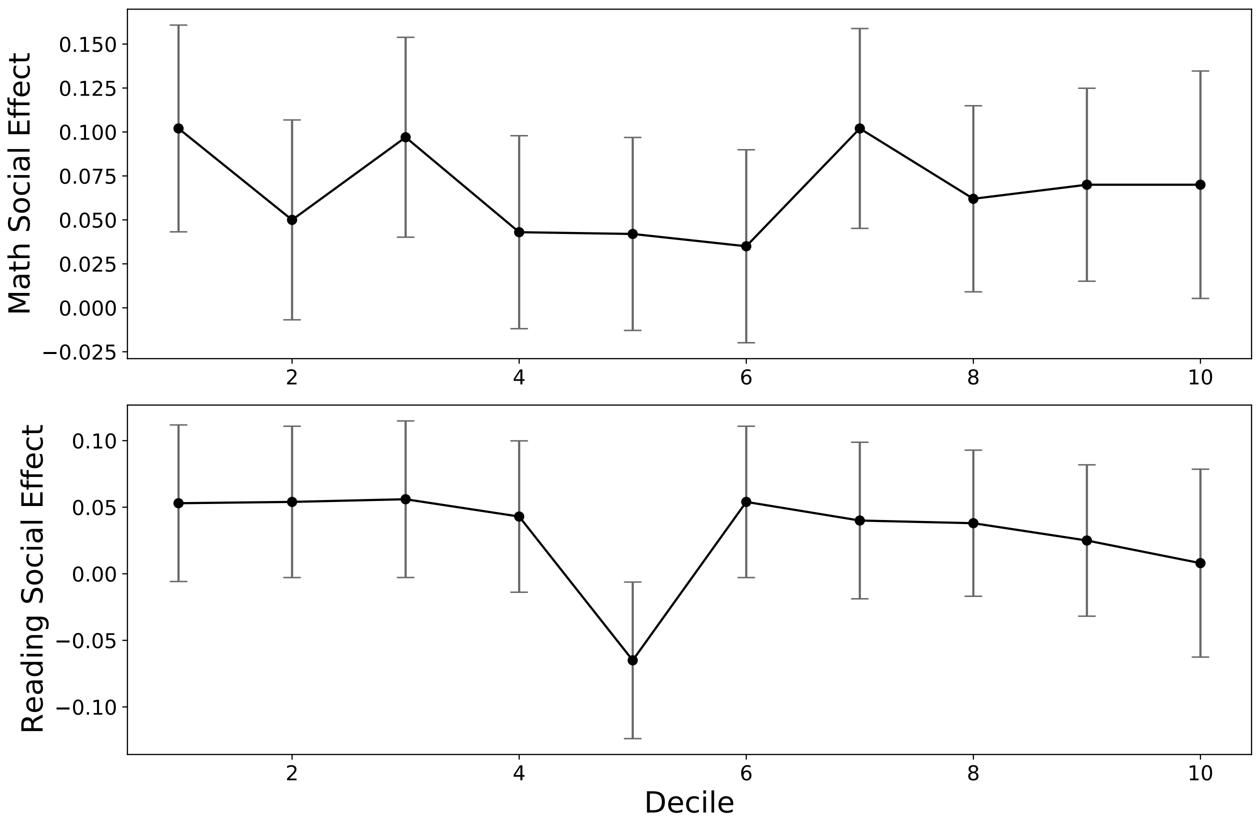
$$S_c = \frac{1}{n_c} \sum_i \frac{1}{n_c - 1} \sum_{i \neq j} \mathbb{1}(i, j) \tag{G.2}$$

Where $\mathbb{1}_d(i, j)$ is an indicator function that takes a value of one when i and j shared a classroom in the previous year, and zero otherwise. S_c is measured at the classroom (c) level, with each classroom having n_c students. This means that if S_c takes a value of 1, every student in the class shared a classroom the previous year, and a value of 0 means no students shared a classroom in the previous year.

I then split the sample by decile of S_c and estimate the social effect for math and reading for each subsample. If we see higher peer effect estimates for classrooms with higher values of S_c, this may indicate that friendships from the previous year carry over and are much more important than new current-year friendships. Figure G1 plots the results of these estimates for both math and reading test scores, with 95% confidence intervals. I find little to no pattern in these estimates, and the results remain relatively consistent across deciles.

[49] Average class size in schools that have only one POS terminal is 25.2 compared with 26.0 in schools with

Figure G1: Classroom Noise Distribution



Graph reports estimates from ten models for each outcome. Models are subsamples for each decile of measure S_c, a measure of how many students in the classroom were classmates the year prior. Bars indicate a 95% confidence interval.

G.6 Evolution of Friendships

Friendships change and evolve over time. If we observe students sorting based on academic performance as the year progresses, this would indicate sorting based on academic performance and introduce potential bias. Additionally, if friendships evolve in preparation for the test, this would explain the strong effect observed in the period before the test occurs.

To address this concern, I calculate how different each student's lag performance is from their friends' lag performance in each period of the current academic year. Specifically, I calculate the following difference D_{ti}:

$$D_{ti} = |Y_{i,y-1} - W_{ct} Y_{y-1}| \tag{G.3}$$

where t indicates the period within academic year y. $Y_{i,y-1}$ is student i's own lag performance. W_{ct} represents the lunch line network in classroom c, and the term $W_{ct}Y_{y-1}$ weights the lagged performance of student i's classmates. I then calculate the average difference in my sample. A low value of $\bar{D}_{ti}$ indicates students are not very different from their friendship groups - that students are not sorting along academic lines. A high value would indicate that students are sorting according to academic performance. If students are changing friends based on academic performance, comparing this difference across periods would detect changes in network composition. For example, if students have an affinity for similarly performing classmates (homophily), and friendship networks evolve towards this affinity, we would expect $\bar{D}_2$ to be smaller than $\bar{D}_1$.

Table G5 reports average differences $\bar{D}_t$ for my sample. As before, I divide the year into between two and six periods in order to compare over time.[50] Each column presents the sample average difference $\bar{D}_t$ for each period t, ranging from two periods in columns (1) and (6) to six periods in columns (5) and (10). Periods in bold denote the period in which the standardized test occurred. The D_t never statistically differ from one another, suggesting that friendships do not evolve based on academic performance.

more than one POS terminal.

[50] Recall from the discussion of Table 3 that when I divide the year into halves, as in columns (1) and (6), I construct two proximity matrices according to equation 4.1. The first proximity matrix refers to days from the first half of the year, and the second matrix refers to days in the second half. As additional periods are added, the length of time represented by Period 1 decreases. In column (1), Period 1 represents about 90 school days, and in column (5) it represents about 30 school days.

Table G5: How different is student academic performance from their friends?

	Math				
	(1)	(2)	(3)	(4)	(5)
Period 1	0.604	0.606	0.608	0.609	0.610
	(0.478)	(0.480)	(0.481)	(0.482)	(0.483)
Period 2	**0.603**	0.606	0.608	0.609	0.610
	(0.478)	(0.480)	(0.482)	(0.483)	(0.484)
Period 3		**0.605**	**0.608**	0.609	0.611
		(0.479)	**(0.481)**	(0.483)	(0.484)
Period 4			0.606	**0.609**	0.610
			(0.480)	**(0.483)**	(0.483)
Period 5				0.608	**0.609**
				(0.481)	**(0.483)**
Period 6					0.609
					(0.482)

	Reading				
	(6)	(7)	(8)	(9)	(10)
Period 1	0.607	0.610	0.612	0.614	0.616
	(0.480)	(0.483)	(0.485)	(0.486)	(0.488)
Period 2	**0.607**	0.611	0.613	0.615	0.616
	(0.480)	(0.484)	(0.485)	(0.487)	(0.488)
Period 3		**0.610**	**0.613**	0.615	0.617
		(0.483)	**(0.486)**	(0.487)	(0.488)
Period 4			0.612	**0.615**	0.616
			(0.485)	**(0.487)**	(0.489)
Period 5				0.614	**0.616**
				(0.486)	**(0.488)**
Period 6					0.615
					(0.487)

I measure how different each student's lag performance is from their friends' lag performance, $D = |Y_{t-1} - W_{ct}Y_{t-1}|$. Each column presents the sample average difference $\bar{D}_t$ for each period t. Periods in bold denote the period in which the standardized test occured. The D_t never statistically differ from one another, suggesting that friendships do not evolve based on academic performance.

H 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.[51] 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. I have no insight into the assignment process, but I 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, I consider a series of multinomial logits as follows:

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

For each iteration of equation (H.1) I include a single grade g within a single school s, during a single year t. I 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.[52] X_i is a binary indicator variable for a characteristic of student i. Each iteration of equation (H.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 I collect the t-statistic for all school-grades. I then conduct my own random assignment of students to classrooms and run the same set of models, again collecting these t-statistics. I then compare the distributions of t-statistics from the observed and simulated models.

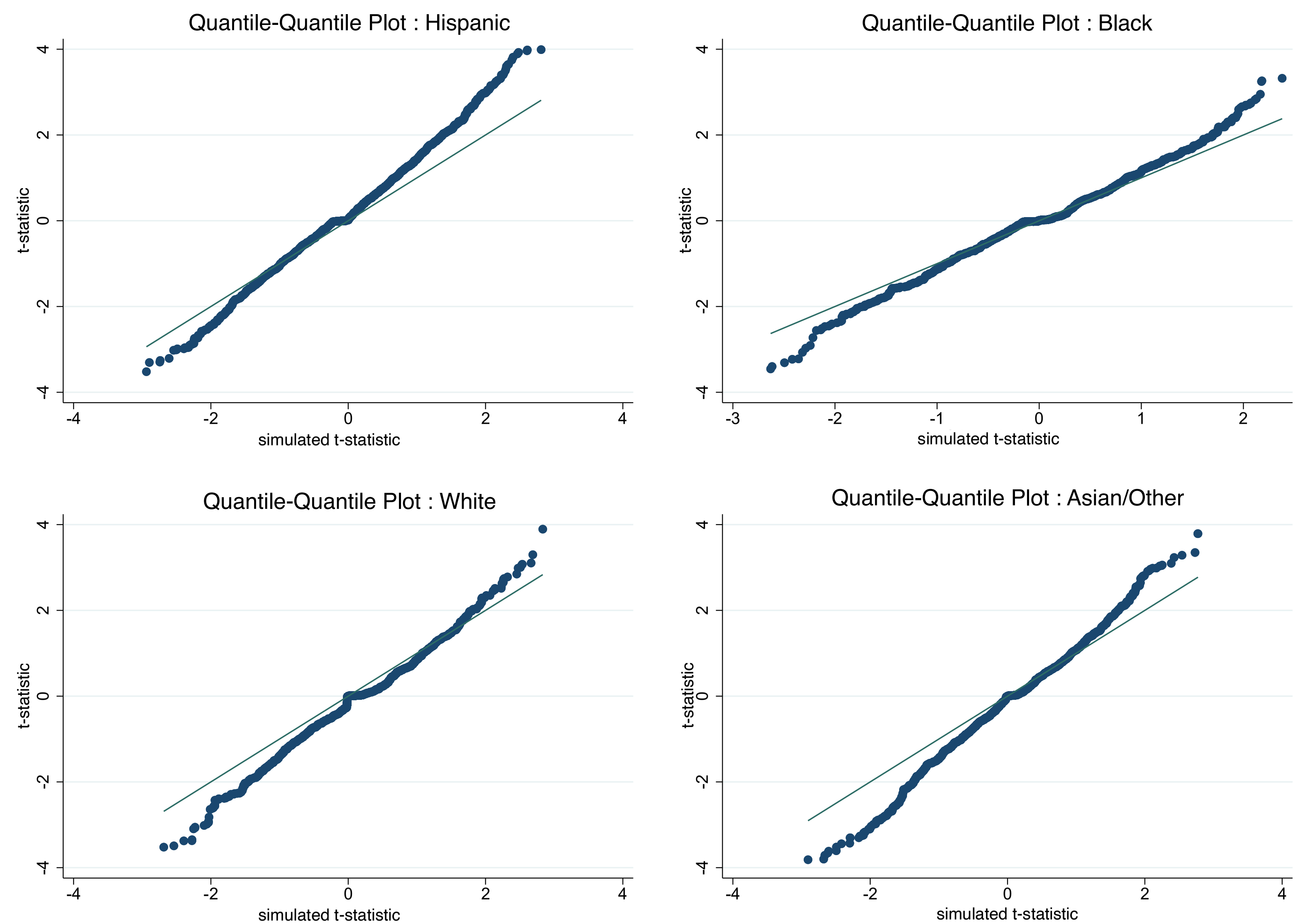
Figures G2 and G3 show the results of these tests. Each dot in these figures compares equal ranked t-statistics from the simulated and observed populations. If these distributions are the same, we should expect a 45 degree line. Most of the observed characteristics remain

[51]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.

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

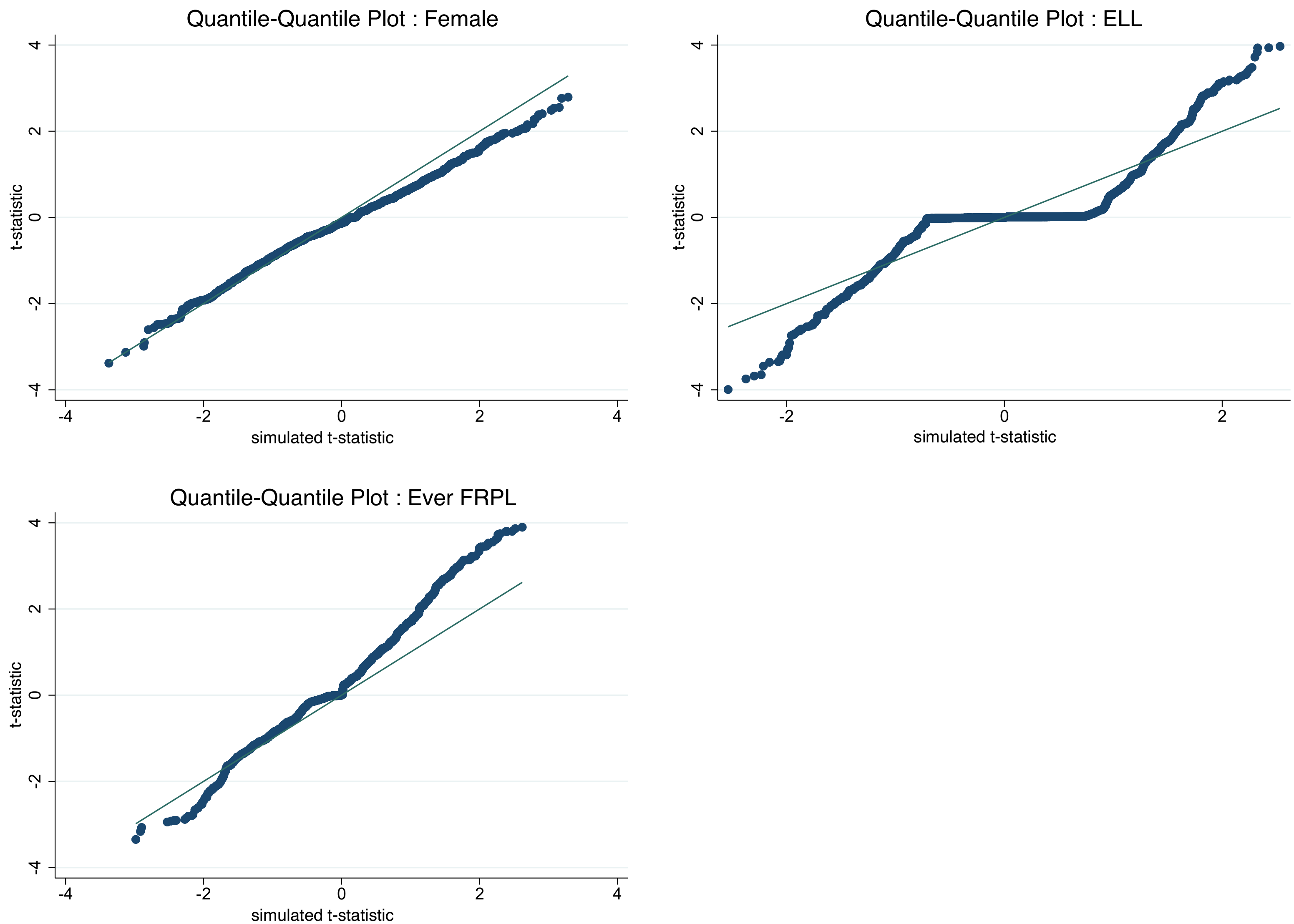
reasonably close to the 45 degree line, with the largest deviation at the tails of the distribution. A notable exception is the English language learner characteristic, which appears to deviate significantly from the 45 degree line. This indicates that classroom assignment may group English language 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, I conclude that we do not need to be concerned about selection into classrooms based on observed characteristics.

Figure G2: Quantile-Quantile Plots



I run a series of multinomial logits 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 I randomly assign students to classrooms. Thus I plot these two distributions against one another, and if the distributions are the same, we should expect a straight line of slope one. I argue that these provide evidence that class assignment is consistent with a random process.

Figure G3: Quantile-Quantile Plots



I run a series of multinomial logits 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 I randomly assign students to classrooms. Thus I plot these two distributions against one another, and if the distributions are the same, we should expect a straight line of slope one. I 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).