IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17358

Estimating Peer Effects among College Students: Evidence from a Field Experiment of One-to-One Pairings in STEM

Robert Fairlie
Daniel Oliver
Glenn Millhauser
Randa Roland

OCTOBER 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17358

Estimating Peer Effects among College Students: Evidence from a Field Experiment of One-to-One Pairings in STEM

Robert Fairlie
UCLA, NBER and IZA

Daniel Oliver
Washington State Student Achievement Council

Glenn Millhauser
University of California, Santa Cruz

Randa Roland
University of California, Santa Cruz

OCTOBER 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Estimating Peer Effects among College Students: Evidence from a Field Experiment of One-to-One Pairings in STEM*

An extensive literature in the social sciences analyzes peer effects among students, but estimation is complicated by several major problems some of which cannot be solved even with random assignment. We design a field experiment and propose a new estimation technique to address these estimation problems including the mechanical problems associated with repeated observations within peer groups noted by Angrist (2014). The field experiment randomly assigns students to one-to-one partnerships in an important gateway STEM course at a large public university. We find no evidence of peer effects from estimates of exogenous peer effect models. We push further and estimate outcome-on-outcome models which sometimes reveal peer effects when exogenous models do not provide good proxies for ability. We find some limited evidence of small, positive outcome-on-outcome peer effects (which would have been missed without our new estimation technique). Standard estimation methods fail to detect peer effects and even return negative estimates in our Monte Carlo simulations because of the downward bias due to mechanical problems. Simulations reveal additional advantages of our technique especially when peer group sizes are fixed. Estimates of non-linear effects, heterogeneous effects, and different measures of peer ability and outcomes reveal mostly null effects but we find some evidence that low-ability peers negatively affect low-ability and medium-ability students. The findings in this setting of long-term, intensive interactions with classroom random assignment and "throwing everything at it" provide evidence of, at most, small positive peer effects contrasting with the common finding of large peer effects in previous studies in education.

JEL Classification: I21, I23

Keywords: peer effects, higher education, STEM, field experiment

Corresponding author:
Robert W. Fairlie
University of California, Los Angeles
Los Angeles, California 90095
USA

E-mail: rfairlie@ucla.edu

* We thank Lisa Morgan and Travis Wiggans for assistance in providing data on students, and to Lori Etow and Steven Luis for assistance in conducting the experiment in the labs. We thank Samantha Sweeney and Justyn Vanderplas for assistance with general questions about the undergraduate program. We thank Moumita Das and Tamar Sasson for research assistance. We thank Pia Basurto Preciado, Christina Felfe, Stefanie Fischer, Carlos Flores, Nicole Fortin, Laura Guiliano, Prashant Loyalka, Lester Lusher, Kyle Neering, Joseph Puglisi, Rachael Robnett, Maike Schlosser, Anastasia Wilson, and participants at Stanford, Cal Poly, the Society of Labor Economists Conference, APPAM Meetings, AEA Meetings, and the Student Success Equity Research Center Workshop for insightful comments on the project.

I. Introduction

Peer effects are one of the most studied questions in the social sciences. They are hypothesized to affect an extremely wide range of outcomes including, but not limited to, education, work, health, crime, risky behaviors, program participation and consumption (e.g., Sacerdote 2014, Angrist 2014). Understanding peer effects is especially important in education because of the growing availability of student performance data, choice of schools, and adaptive and personalized learning environments. This contemporary landscape in education heightens the potential for concentrations of students by ability and resulting peer effects. An extensive literature examines educational peer effects with most studies recovering large positive estimates although there is a wide range of estimates including examples of negative effects of high ability peers on low ability peers (Sacerdote 2014; Angrist 2014; Carrell et al. 2013; Feld & Zolitz 2017).

Peer effects are notoriously difficult to estimate and interpret, and we know relatively little about how they work. Perhaps the two most difficult problems to overcome, despite being well understood, are the "reflection problem" and the self-selection of peer groups (Manski 1993, 1995). The reflection problem is commonly addressed by simply including measures of student ability that pre-date the academic outcome of interest (which also addresses the potential problem of common shocks to the academic outcome). But, there is concern that predetermined measures of ability do not always capture the peer's ability that is specific or particular to the academic outcome. Thus, an argument can be made for also estimating outcome-on-outcome models which might reveal peer effects when predetermined proxies for ability in the subject of interest do not. On the second problem of self-selection, several recent studies rely on random assignment of peer groups (most commonly outside of the classroom): for example, dormitory roommates (Sacerdote 2001; Garlick 2018; Zimmerman 2003; Han & Li 2009), U.S. Air Force Academy students in social-network groups (Carrell et al. 2009; Carrell et al. 2013), professional golfers (Guryan et al. 2009), Dutch business school students (Booij et al. 2017; Feld and Zolitz 2017; Golsteyn et al. 2021), student gym use (Babcock et al. 2020), boarding school roommates in Peru (Zárate 2023), and student peer assessment groups (Kamei & Ashworth 2023).

But, random assignment does not solve several additional problems in the estimation of peer effects. First, a problem that has been under-appreciated, but emphasized by Angrist (2014) and Guryan et al (2009), arises when peers groups are formed or assigned from a larger group "without replacement." The common practice of random assignment within this "urn" results in a mechanical negative correlation. This induces biases in balance checks and peer effect estimates derived from outcome-on-outcome models. Second, there is the often-overlooked problem of not knowing who interacts with whom in the classroom and those interactions are likely to be self-selected even with random assignment to classrooms. In large lecture halls and even relatively small classrooms, students can generally choose with whom they sit near and interact in the classroom, and that information is almost never observed. Third, variation in the measure of peer ability, which is typically represented by average classmate ability, decreases rapidly as class size increases. Random assignment to classrooms clearly cannot fix this problem.[1] Fourth, it is not straightforward to test different functional forms for estimating peer effects in classrooms because of the problem of identifying which classmates are influential and which classmates are not.[2] For example, is student performance most affected by one or two disruptive students or is the ability of the top students in the classroom the most important?

We design a field experiment that randomly assigns classroom peers and propose a novel estimation technique to address these problems. The large-scale experiment is designed to estimate peer effects in an important, real-world setting in higher education: interaction effects between classmates in introductory Chemistry laboratory courses. We chose these courses because students work closely together in pairs for the entire term and because these courses are important – they are an essential, gateway course for the Sciences, pre-Med majors, and many other STEM majors.[3] As part of the experiment, students in every lab section associated with the introductory sequence in Chemistry at a large public research university were randomly assigned a partner. The study involves roughly 5,000 students, 7,000 course observations, and 3,500 peer groups over five academic years, 2014-2019. Students are graded independently on tests and

---

[1] Previous studies with larger groups experimentally manipulate the composition of these peer groups to increase the range of support or to increase interactions between low and high ability students (Booij, Leuven and Oosterbeek 2017; Carrell, Sacerdote and West 2013).
[2] For example, the underlying assumption for using average classroom peer ability is that every student in a classroom of 30 students has an equal, 1/29 effect on other students. Another problem, for example, is that increasing the number of high-achieving students in a group may induce low-achieving students to form subgroups among each other, which might not have occurred if there were fewer high-achieving students (Carrell, Sacerdote and West, 2013).
[3] The Introduction to Chemistry sequence which includes the labs is the gateway requirement to STEM majors, including Chemistry, Biology, Bioengineering, Environmental Studies, Environmental Science, Earth Sciences, Ecology and Neuroscience. It is also commonly taken by students in many other STEM majors (e.g. Physics, Computer Science, and Cognitive Science).

assignments, and grades in the labs are based on an absolute scale. Detailed administrative data from the university allow for estimating peer effects on a wide range of course outcomes, including numeric scores, drops, grades on different types of assignments over the term, and downstream outcomes such as continuation to more advanced chemistry classes or ultimate selection of a Chemistry or other STEM major.

Partly with the goal of taking a comprehensive approach that also allows for estimation of outcome-on-outcome models, we design and implement a novel, but straightforward, estimation technique that includes one-observation-per-peer group (OOPG) to address mechanical biases that are not addressed by random assignment. The OOPG technique involves randomly including one observation from each peer group so that peer observations are not repeated. This satisfies Angrist's (2014) intuitive suggestion of making a "clear distinction between the subjects and their peers" when estimating peer effect models and is more in line with an RCT. We show in Monte Carlo simulations that the technique recovers peer effects when they exist and null estimates when they do not exist. We also show how the technique works for each of the estimation applications in peer group settings (i.e. exogenous effects, outcome-on-outcome effects, randomization checks, and fixed urn sizes).

We chose Chemistry labs also because the one-to-one pairings in the labs ensure that the randomly assigned peers interact with each other. The random assignment of lab partners in our experiment addresses the problem of not knowing who interacts with whom or endogenous sorting within a randomly assigned classroom or group. The one-to-one matching of peers also maximizes potential variation in peer abilities (i.e. variation in classmate averages increases as class sizes become smaller). Furthermore, one-to-one matching makes it easier to test different functional forms for estimating peer effects because it removes the potentially conflating problems of identifying which classmates are influential and which classmates are not.

The experiment solves additional problems in estimating peer effects. First, changes in instructional practices or competition in grading can make it difficult to isolate actual student-to-student peer effects. Chemistry labs in our experiment are standardized in content and grading practices, and we include section fixed effects (i.e. specific to course-term-section), thus isolating student-to-student peer effects by removing instructor effects and the potential adjustments of instructors to classroom compositions. Second, grading is not based on a curve which removes concerns regarding college students competing for relative grades potentially undermining beneficial peer interactions. Third, linear-in-means models can result in attenuation and other biases from measurement error of actual peer groups or functional form misspecification. In estimating exogenous peer effects there is still potential for attenuation bias due to classical measurement error in the proxy for peer ability. We address problems associated with measurement error noted in Angrist (2014), and measurement problems more generally (e.g. sample attrition and item non-response), by using the same measures for the student and the classmate average (i.e. partner) and administrative data for all measures.

Estimates from standard linear-in-means models from our experiment indicate null effects on course scores, grades, dropout rates and pass rates. The findings are robust to several different measures of peer ability and additional student outcomes, and the estimates are precise enough that we can rule out large positive effects. We push further and estimate outcome-on-outcome models that can sometimes identify peer effects even when exogenous peer effect models do not. Naive estimates from outcome-on-outcome models indicate null estimates for course scores and grades, but negative estimates on dropout rates and pass rates. We demonstrate that these estimates are driven by mechanical biases. Using Monte Carlo estimates we can show that even in the presence of zero peer effects the outcome-on-outcome model will provide negative estimates. We explore three techniques to fix the problem, but only our novel one-observation-per-peer-group (OOPG) technique works in all applications (and importantly for the outcome-on-outcome model). We also demonstrate how the technique works for balance checks even in the case of having no variation in urn sizes. After correcting for biases, we find a small positive peer effect estimate for the outcome-on-outcome model for course scores and grades. Thus, the mechanical bias creates a downward bias on the outcome-on-outcome estimates which could potentially conceal a positive peer effect without using the OOPG technique. The OOPG technique also detected an outcome-on-outcome peer effect when the exogenous peer effect model did not.

We then dive much deeper into the main exogenous peer effect estimates and further explore the reflection problem. Focusing on specific assignments in the labs or the timing of grading over the term we also do not find evidence of peer effects. Focusing on longer-term outcomes and related outcomes we find no evidence of peer effects. We estimate models allowing for non-linear effects, heterogeneous effects and different measures of outcomes and ability, and find predominately evidence of null effects. The only exception is that we find some evidence that low-ability peers negatively affect low-ability and medium-ability (but not high-ability) students. In this setting of highly interactive, long-term interactions in a gateway STEM course in which peers are randomly assigned and throwing everything at it including estimating outcome-on-outcome models, we find evidence of, at most, small positive peer effects.

Our study contributes to two important literatures. First, it contributes to the relatively small literature on peer effects in higher education and even smaller literature using random assignment of peer groups in college classrooms.[4] Feld and Zolitz (2017) and Golsteyn, Non and Zolitz (2021) study random assignment of students to sections of 10-15 students in problem-based learning courses at a Dutch business school. They find that while middle- and high-ability students benefit from higher-ability peers, low-ability students do not, and that students perform better with persistent peers. Although not focused on the classroom but in more general social-network peer groups, Carrell et al. (2009) study the effects of peer groups among randomly-assigned groups of 30 students in the U.S. Air Force Academy and find large peer effects and non-linearities in the magnitude of peer effects by student ability.[5] In a follow-up study, Carrell et al. (2013) assign half to peer groups designed to maximize the academic performance of the lowest ability students. They find that lower-achieving U.S. Air Force Academy students do worse when assigned to squads with more high-achieving peers. Subsequent studies also focus on altering the composition of peer groups. Kamei & Ashworth (2023) randomly assign partners in an introductory economics course in the U.K. to provide peer feedback and discuss an assignment for the course. They find that individual performance on the assignment and other areas of the course are enhanced in pairs with a greater spread in ability. Booji et al. (2017) randomly assign the composition of roughly 50 different tutorial groups with 40 students each at a Dutch business and economics program to achieve a wide range of support. They find that when switching from ability mixing to three-way tracking substantially reduces dropout rates and increases academic performance for middle and low-ability students.

[4] A larger and more established literature focuses on peer effects in primary and secondary school. The limited focus on higher education is likely due complications associated with college students having more choice over schools, subject matter, courses and sections than younger students.

[5] As noted in Carrell et al. (2009) "Conditional on a few demographic characteristics, the students in our study are randomly assigned to a peer group in which they live in adjacent dorm rooms, dine together, compete in intramural sports together, and study together. They have limited ability to interact with other students outside of their assigned peer group during their freshman year of study."

Our study is the first to use random assignment of one-to-one matched student groups that interact with each other in the classroom throughout the entire term for roughly 3,500 peer groups over a five-year span and two different courses. Our findings differ from these in that we find no evidence of peer effects from exogenous peer effect models but find some evidence from outcome-on-outcome models that would not have been detected without our new estimation technique. We also find no evidence of negative effects from high-ability peers unto students of any ability level. Peer effects in an intensive, continual interaction but separately graded course appear to be minimal at most, contrasting with the findings from many previous studies.

We also provide an analysis and discussion of the experimental design that may be useful for future research in the large and rapidly growing literature on estimation of peer effects more broadly. We designed the experiment to address the major methodological concerns raised in the peer effects literature, and provide a useful template for examining all of these issues. We also propose and validate a new estimation technique that addresses mechanical problems recently noted by Angrist (2014).

The remainder of the paper is organized as follows. In Section II we describe the experimental setting and data. Section III provides a detailed discussion of methodological challenges in estimating peer effects and presents randomization checks. Section IV presents the main peer effects estimates. Section V presents outcome-on-outcome peer effect estimates and explores alternative estimation strategies including our novel procedure. Section VI provides additional results on the timing and type of course assignments, and longer-term outcomes. Section VII explores estimates of non-linear-in-means peer effects, and Section VIII concludes.

II. Experiment and Data

Experimental Setting

For the experiment, we randomized all student pairings in introductory Chemistry labs from Winter Quarter 2015 to Spring Quarter 2019 at a large, public university. The university has a total enrollment of roughly 20,000 students. Total enrollment in all labs observed for our study is 5,537 (3,902 students). Enrollment in the 330 unique Chemistry labs is capped at 18 (mean=16.8). Average enrollment in the large-lecture introductory Chemistry courses is 348.

We chose Chemistry labs to estimate peer effects in higher education because students work closely together in pairs for the entire term and because these courses are important – they are an essential, gateway course for the Sciences, pre-Med majors, and many other STEM majors. One-to-one student interaction in lab partnerships especially facilitate an analysis of peer effects. In a classroom of 300 students, or even 30 students, it is very difficult to identify which classmates have the most influence on a particular student and students can generally choose which other students they want to sit near and interact with in the class potentially avoiding or reducing gender bias and discrimination from peers. In Chemistry labs, student pairs work very closely together the entire term, but take individual assessments and are graded on their own knowledge of the subject material. The one-to-one matching in Chemistry labs removes this peer measurement problem and provides an intensive interaction between students. Additionally, we avoid the concern that random assignment creates little, or essentially no, variation in average ability levels in classrooms when there are large classes. The assigned lab partner has one ability level removing concerns over using mean, median, low or high values of ability levels across students in classrooms.

The Introduction to Chemistry sequence at the university covers a standard set of topics, similar to other large research universities. The laboratory classes associated with this sequence are also standard. The sequence requires a minimum of pre-calculus before enrolling, but most students have already taken calculus. The sequence involves extensive use of math throughout the coursework. Students generally take Chem 1A, 1B and 1C in consecutive quarters. The two labs (Chem 1M and 1N) are associated (but not required) with the second and third quarter courses in the sequence, respectively.

The Introduction to Chemistry sequence which includes the labs is the gateway requirement to a diverse set of STEM majors, including Chemistry, Biology, Biochemistry, Molecular Biology, Bioengineering, Environmental Studies, Environmental Science, Earth Sciences, Ecology and Neuroscience. It is also commonly taken by students in many other STEM majors (e.g. Physics, Computer Science, and Cognitive Science). Chemistry labs develop a broad skillset including a strong mathematical component (e.g. statistics, linear regression, physical processes, experimental measurement, and instrumentation).

The laboratory curriculum, physical equipment and space are standardized across sections. The laboratory equipment and materials are uniformly disbursed from a central laboratory manager. Lab sections are held in eight different laboratory classrooms along one hallway in the Chemistry instruction building. The standardization across lab sections provides one of the most controlled environments for studying social interactions between students possible on a college campus.

Data and Measures

We obtained administrative data from the university for students enrolled in introductory Chemistry labs. We have detailed background baseline data measured prior to enrollment in the Chemistry labs, and post-enrollment data on several academic outcomes. The baseline data include measures of baseline ability and background demographic and academic characteristics.

We measure baseline ability in two main ways. We distinguish between low and high ability based on a student's performance in all previous courses (i.e. prior GPA). Another baseline measure of ability is the student's grade in Chemistry 1A, which is the first lecture course taken in the introductory sequence. Chemistry 1A is taken in a prior term to enrollment in the labs. As expected, grades in Chemistry 1A are a very strong predictor of performance in the lab. We also use grades in this course to define low and high ability students and lab partners.

Administrative data also provide a rich set of additional controls to include in the regressions and use for balance and sorting tests. These include baseline lab section fixed effects, a detailed set of race/ethnicity indicators, Educational Opportunity Programs (EOP) status[6], year in college, major interest, and declaration of major. It is important to note that the inclusion of lab section fixed effects in Equation (1) controls for the variation in performance due to different instructors, teaching assistants, rooms, lab courses (i.e. Chemistry 1M and 1N), academic years/terms, section times, and days of the week. Importantly, these fixed effects also control for the ability and demographic distribution of all students in the lab.

We focus on three sets of academic outcomes: i) lab course performance and outcomes, ii) same-term other course outcomes, and iii) future outcomes related to Chemistry and STEM. Our primary measure of overall course performance is the underlying numeric continuous total score in the class (i.e. scale of 0-100). We rescale this score by demeaning and dividing by the standard deviation. We also measure performance using the letter grade in the course converted to a 5-point scale (i.e. scaled similarly as a GPA measure, F=0.0 to A=4.0).[7] We also estimate

[6] EOP students are in-state residents who are first generation college going or have low expected family contribution towards financial support.
[7] Less than 1 percent of students take the course pass/no pass instead of taking it for a letter grade.

peer effects on the behavioral outcome of dropping the lab course that term. Peer effects might be stronger on this behavioral outcome than performance outcomes. A goal for many students to move on in STEM is to pass and complete the lab course. We also create a summary measure that captures whether the student passed and completed the course (i.e. did not fail or drop).

We also examine separate components of grades in the lab courses. We have scores on different assignments and exams. We also have scores at different periods within a term (e.g. first few weeks vs. last few weeks).

The second set of academic outcomes focuses on other courses taken by the student in the same term as the Chemistry lab. Peer effects might transfer over to other courses that term because, for example, a high-ability partner might motivate or help the student with homework or studying for other classes. This effect could be the strongest in concurrent Chemistry classes. We explore impacts on grades in the large lecture Chemistry courses taken concurrently with the lab.

The third set of academic outcomes focuses on whether lab partners affect student interest and continuation in Chemistry. We estimate whether lab partners influence subsequent course taking in Chemistry which is captured by future enrollment in the secondary sequence in organic chemistry for all students including those not required to take the sequence. We also estimate effects on declaring majors in Chemistry, and declaring majors in any STEM field.

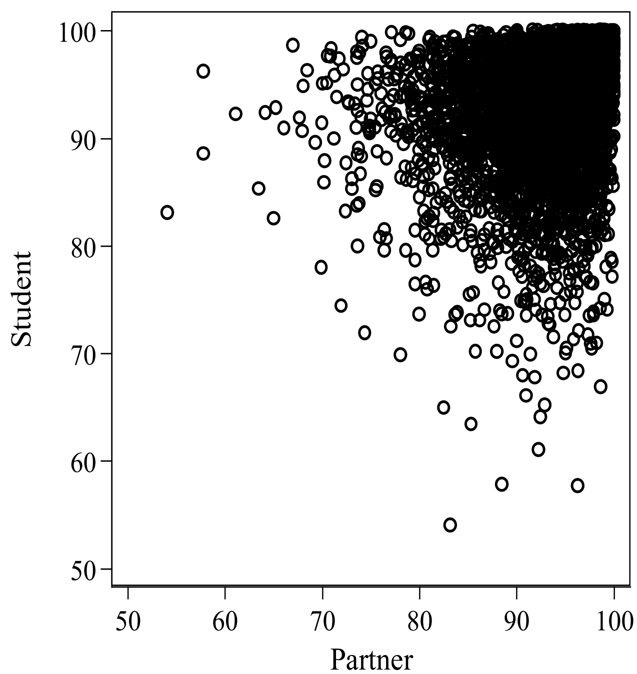
The combination of these administrative data sources provides a comprehensive view of the possible peer effects resulting from Chemistry labs. We can examine short run outcomes within the chemistry lab, dynamics of grades within the course, effects that spill over to the associated large-lecture Chemistry course, and longer-term outcomes such as enrolling in a subsequent Chemistry course and majoring in Chemistry or STEM fields.

Grading

The continuous underlying score and corresponding letter grades are determined by several assignments and exams. Specifically, final scores in the class are based on the following assignments: Written procedure and data tables (7 assignments) 25% Pre-labs (7) 5%; In-labs (7) 35%; Reviews (7) 5%; Formal abstracts (2) 10%; Quizzes (7) 10%; Scholarship and week 1 worksheet 10%. These are standardized across all sections. In addition to the total (numeric) score in the lab course, we have the detailed gradebooks for each course. This includes four different types of assignment scores over the term: quizzes, pre-labs, in-labs, and reviews. For each week of the term, we have different score summaries allowing us to create first-half and second-half scores in the course and whether peer effects differ over the quarter.

Although students work together throughout the quarter in pairs these assignments and quizzes are done independently. Figure 1 displays a scatterplot of the student and their partner's total numeric score in the lab course. The results confirm that lab pairs do not receive the same scores in the course. The correlation coefficient on the two total scores is 0.19.

Figure 1: Relationship between Student and Partner Total Scores in the Lab Course



Focusing on specific assignments for the course (i.e. quizzes, pre-labs, in-labs, and reviews) we checked to determine if lab partners were graded jointly on these different types of assignments. We find similar results: students and their partners do not receive the same scores on any of these types of assignments (see Appendix Figure 1). The correlation coefficients are 0.07 for quizzes, 0.08 for pre-labs, 0.16 for in-labs, and 0.11 for reviews.

Randomization Process

We randomly assigned partners in all introductory Chemistry lab courses over the past five years. Students were assigned partners on the first day of sections, and these partnerships were maintained for the entire term. The process of randomization was deliberately transparent – students drew folded slips of paper with numbers between 1 and 9 from a large beaker. Students with matching numbers were paired. When only 16 students (or an even number of students below 18) were present for the draw, pairs of slips with the same number were either omitted from the beaker, or students with the lowest unmatched numbers were matched. When only 17 students (or an odd number of students) were present for the draw, the non-matching student was added to the lowest numbered pair. We drop these observations which represent only 2 percent of all partnerships.

III. Random Assignment and Peer Effect Estimation

We present our estimation strategy and discuss how our field experiment addresses the major problems that arise in the estimation of peer effects. The starting point for our econometric specification is to use the peer effect model with pre-determined peer measures. A simplified linear-in-means regression specification is:

(3.1) $\mathrm{Y}_{ig} = \beta_0 + \beta_1 A_{ig} + \beta_2\, \bar{A}_{-ig} + \varepsilon_{ig}$,

where Y_{ig} is the student's course outcome, A_{ig} is the student's ability level measured prior to the course (pre-course ability), $\bar{A}_{-ig}$ is mean group ability excluding student i, and ε_{ig} is the error term. In our case, the peer group includes only 2 people (the student and their lab partner) and *g*=2. Focusing on pairs, we estimate the model:

(3.2) $\mathrm{Y}_{ig} = \beta_0 + \beta_1 A_{ig} + \beta_2 A_{-ig}^{PT} + \theta' \mathrm{W}_{ig} + \lambda_s + \varepsilon_{ig}$,

where A_{-ig}^{PT} is the partner's academic ability, λ_s are section fixed effects, and W_{ig} is a vector of student characteristics (which include gender, race/ethnicity, EOP status, residential college, year in college, prior units accumulated, and major interest/declaration). Standard errors are clustered at the section level.

Recently, Angrist (2014) linked econometric models of peer effects to the behavior of IV estimators relative to OLS estimators. Therefore, the standard OLS estimate of peer effects can also be thought of as a function of the difference between the IV estimate of the student's course outcome on the student's pre-GPA and the OLS estimate of the same relationship. For the IV model group assignment dummies (i.e. partner fixed effects) are used as instrumental variables for the student's pre-course GPA. The use of group assignment dummies and no other controls in the first stage implies that the second-stage equation is the same as regressing individual outcomes on group means. The presence of peer effects on student outcomes creates a wedge between the two estimators. However, importantly, Angrist (2014) notes that the two estimates might differ for several reasons other than peer effects possibly resulting in large positive or negative estimates of the peer effect in the original equation such as selection, differential measure error, or mechanical links. Using this framework, we discuss each of these possibilities and how they are addressed in our setting.

Reflection Problem and Common Shocks

An important and well documented concern in estimating peer effects is commonly referred to as the "reflection problem" due to Manski's (1993, 1995) influential work on the topic. The "reflection" arises because the individual's behavior is not only influenced by his/her peers, but also potentially influences the behavior of those peers. Not accounting for this simultaneity creates problems for the interpretation of estimates of the causal effects of peers on individual outcomes. The most common method in the empirical literature of addressing this concern is to include pre-determined peer measures in the regression instead of contemporaneous peer outcomes.[8] The estimated effects are sometimes called exogenous peer effects (Manski 1993), social returns (Angrist 2014) or contextual effects contrasting with endogenous outcome-on-outcome peer effects.[9] Equation (3.2) follows this approach by using pre-determined measures of student ability such as prior GPA.

Another concern are common shocks that affect peers similarly, such as having a bad instructor or an especially disruptive student that term. If these shocks are unobserved or not controlled for in the regression then they might be captured in the peer effect estimate. The focus on pre-determined peer effect measures addresses the bias created from randomly determined common shocks, but not when common shocks are correlated with the independent variable.

---

[8] It is possible, however, that lab partners interacted with each other in previous classes, and that the student had an effect on his/her peers' previous outcomes which are captured in pre-course GPA. These effects although impossible to rule out are likely to be very small because of the large class sizes of introductory courses and the wide range of majors taking Chemistry labs (as shown below).

[9] Angrist (2014) notes that pre-determined characteristics are more important for social planners for boosting achievement because they can only work with the information that they have prior to enrollment.

Random assignment to groups in our field experiment, however, removes the possibility of the common shock being correlated with the independent variables.

Self-Selection, Correlated Unobservables, and Balance/Sorting Checks

Perhaps the most important concern and most difficult problem to overcome in estimating peer effects is the potential bias from how groups are formed, especially if students can select into groups. In particular, positive group selection (and more generally any correlated unobservables) can lead to a positive bias in estimating peer effects even when one does not exist.[10] Estimating peer effects in college classrooms is especially difficult because there are very few scenarios where students cannot choose their classes or peers. The random assignment into lab pairs through our multi-year experiment removes these concerns regarding self-selection of peers and correlated unobservables.[11]

We use two methods to check for balance and sorting. First, we conduct a standard RCT style balance check. For the linear-in-means model, the experiment essentially randomizes treatment intensity, which is measured by the partner's baseline ability level. A balance check is performed for each student baseline characteristic by regressing that characteristic on the treatment intensity variable, which is partner's pre-determined ability. A_{-ig}^{PT}. In our case, we test the validity of random assignment of lab partners by estimating the following regression:

(3.3) $X_{ig} = \varphi_0 + \varphi_1 A_{-ig}^{PT} + \varphi_2 \bar{A}_{-is} + \lambda_s + \varepsilon_{ig}$,

where X_{ig} represents a baseline characteristic of the student (i.e. A_{ig} or a separate row of W_{ig}). When estimating peer effects students cannot be matched with themselves, and thus we are essentially sampling without replacement. To correct for the downward bias associated with this test of randomization we control for the average ability of the urn from which the partner was drawn (i.e. lab section) but leave out the student, $\bar{A}_{-is}$.[12] We discuss the bias and the inclusion of

[10] Continuing with the IV vs OLS comparison framework noted above, it is easy to see how positive group selection (and more generally any positively correlated unobservables) can lead to a positive bias in estimating peer effects (Angrist 2014). Positive selection into groups (such as high ability students forming pairs) will lead to larger relative estimates of ψ_IV because of the created correlation in the Xs within groups. The positive correlation in the Xs from selection, however, does not affect estimation of ψ_OLS resulting in ψ_IV>ψ_OLS.

[11] Randomization occurs within sections so we control for section fixed effects in all analyses, although our results without section fixed effects are similar.

[12] The bias is created by the difference between the unconditional and conditional expectations from sampling without replacement. The conditional expectation from not including the student is $\frac{n}{n-1}\mu - \frac{1}{n-1}x_j$, and creates a bias if

this variable in more detail in Section V, and we propose a new alternative estimation strategy to correct the problem.

This specification tests whether baseline student characteristics are uncorrelated (or "balanced") with treatment intensity (i.e. the partner's ability level). Table 1 reports estimates of the "balance" check. Importantly, we find the student's GPA is uncorrelated with the peer's or partner's GPA. We also find using an alternative measure of partner's GPA, the Chemistry 1A grade, that student's ability is uncorrelated with partner's ability. For all of the student characteristics we find "balance" with either measure of partner's ability.

X_{ig} is correlated with A_{-ig}^{PT}. Guryan et al. (2009) include the leave-me-out section mean in the related randomization check, which we also estimate below.

Table 1: Randomization Check of Experiment

	Var Mean	(1) Pre-GPA	(2) Chem 1A GPA	(3) Pre-GPA	(4) Chem 1A GPA
Own Pre-GPA	3.213	0.004	-0.003	-0.008	0.001
		(0.003)	(0.005)	(0.018)	(0.007)
Own Chem 1A Grade	2.906	0.018	0.000	0.014	-0.004
		(0.022)	(0.005)	(0.031)	(0.017)
Female	0.580	-0.011	-0.010	-0.009	-0.008
		(0.015)	(0.007)	(0.014)	(0.006)
EOP Student	0.337	-0.001	0.008	0.001	0.007
		(0.014)	(0.007)	(0.013)	(0.006)
Freshman	0.239	0.013	-0.000	-0.003	-0.007
		(0.012)	(0.005)	(0.013)	(0.005)
Sophomore	0.589	-0.008	0.002	0.008	0.008
		(0.014)	(0.007)	(0.014)	(0.006)
Junior	0.128	-0.000	-0.000	-0.002	-0.001
		(0.010)	(0.005)	(0.009)	(0.004)
Asian	0.316	-0.005	0.001	0.002	-0.002
		(0.014)	(0.007)	(0.013)	(0.006)
Black	0.020	0.002	-0.003	0.003	-0.002
		(0.004)	(0.002)	(0.004)	(0.002)
Latinx	0.274	-0.014	-0.003	-0.015	-0.002
		(0.012)	(0.006)	(0.012)	(0.006)
White	0.291	0.016	0.005	0.007	0.005
		(0.014)	(0.006)	(0.013)	(0.006)
Observations		6,334	5,913	6,334	6,116
FE		Section	Section	Course-Term	Course-Term
Additional Control		LMOUM	LMOUM	None	None

Notes: Each cell represents a separate regression of the specified student characteristic on the partner's ability measure. Columns 1 and 2 also include the leave-me-out urn mean (LMOUM) which is the average of ability for the other students in the lab section. Standard errors are clustered as the section level. *** p<0.01, ** p<0.05,* p<0.1.

Table 1 also reports specifications that do not include section fixed effects and instead include year-quarter-lab type fixed effects (Specifications 3 and 4). Randomization does not occur at this level but we continue to find balance. Baseline student characteristics are not correlated with either measure of treatment intensity at the section level.

The second method we use is a standard and related sorting test used when random assignment is not possible. The common check for the potential problem of sorting is to estimate the correlation in different baseline characteristics between the student and the partner (e.g. Sacerdote 2001; Guryon et al. 2009). The focus is on sorting based on observable baseline characteristics of both the student and their partner as a proxy for sorting based on unobservables. In this case, we estimate the following regression:

(3.4) $X_{ig} = \varphi_0 + \varphi_1 X_{-ig}^{PT} + \varphi_2 \bar{X}_{-is} + \lambda_s + \varepsilon_{ig}$

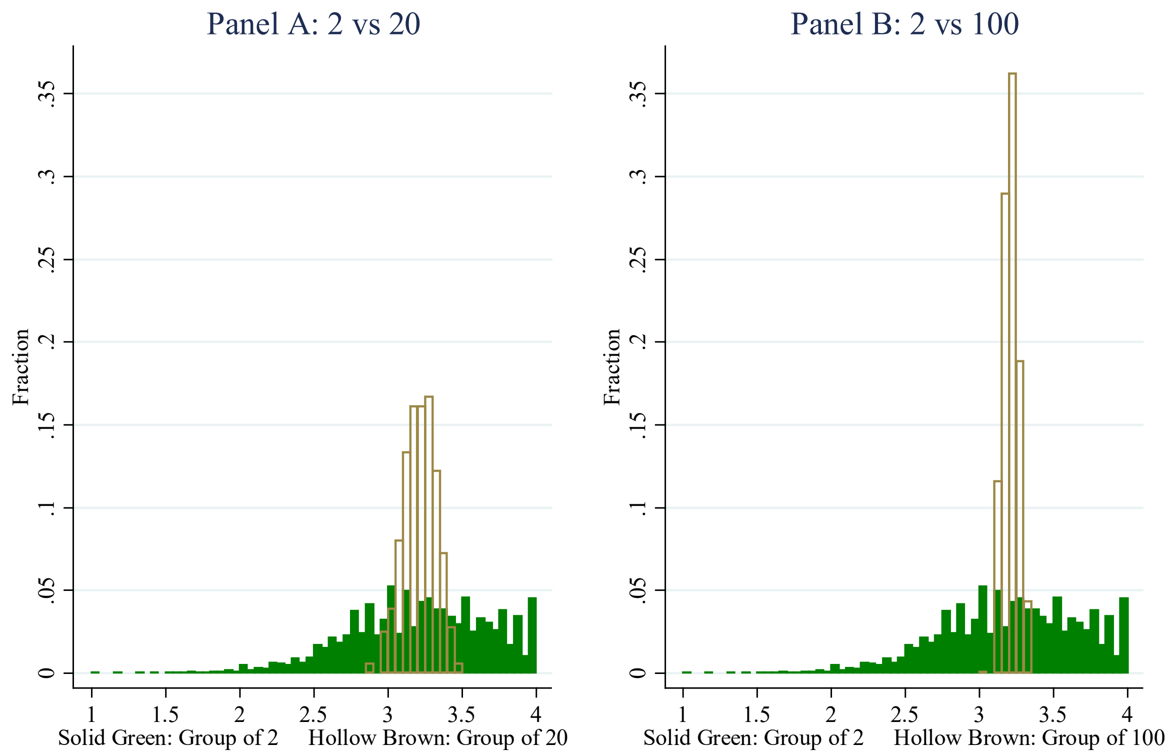
X_{-ig}^{PT} represents the corresponding baseline characteristics of the partner. We also include $\bar{X}_{-is}$ which controls for the average value of peer characteristics within each laboratory section for the same reasons as noted above. Appendix Table 1 reports estimates of (3.4). Each row represents a separate regression measuring potential sorting by students along that baseline characteristic. Of all the student characteristics we only find one that is statistically significant at the 0.10 level which is a negative correlation between Asian student and Asian partner.

Random Assignment and Endogenous Sorting within Classes

Even within randomly assigned groups, if group sizes are larger than two, then interactions within groups may be endogenously determined. Students might only interact with a few other students in the classroom and likely choose those other students. Carrell, Sacerdote and West (2013), for example, devised an experiment in which students were assigned into class peer groups in a way that would improve either aggregate student academic performance or at least the performance of the lowest ability students. However, findings from the experiment indicated that the lowest ability students experienced a negative treatment effect whereas middle ability students experienced a positive treatment effect. Students appear to have avoided the peers with whom they were intended to interact with and instead formed more homogeneous subgroups. These types of sorting within classrooms, perhaps by ability, are avoided in our experiment because students are required to work together in one-to-one pairs the entire term.

A related problem that random assignment cannot address is the lack of variation in mean ability of peers for normal classroom sizes. With larger class sizes, for example, it is unlikely that average peer ability will vary much across classrooms. Even with smaller class sizes there might not be much variation in mean peer ability across classrooms with random assignment. In our experiment matching one-to-one peers, however, we get full range of distribution of peer groups. Figure 2 displays how the distribution of the peer group ability measure changes from our case (actual partner data from our experiment) and larger class sizes (which are created from randomly assigning students to "classrooms" of the specified size). The loss of variation moving from our lab partner case to a class size of 100, or even an unusually small class size in higher education of 20, is substantial.

Figure 2: Peer Ability Distribution by Group Size



IV. Peer Effect Estimates

Table 2 reports estimates of our basic regression model for peer effects (Equation 3.2). Specification 1 reports estimates for the numeric total score in the lab course. We rescale the underlying numeric continuous score (i.e. scale of 0-100) in the course by demeaning and dividing by the standard deviation. As expected, we find a large positive and statistically significant coefficient on the student's own GPA (measured prior to the current term).[13] Increasing a student's GPA by 1 unit on a scale of 0-4 is associated with 0.81 standard deviations higher total score in the lab course. Students, however, do not perform better when randomly

[13] Estimates are similar when we add the section leave-me-out mean. We discuss issues regarding these measures in Section VI.

partnered with higher ability partners. We find a coefficient estimate on partner's ability that is essentially zero and is estimated very precisely. The point estimate is -0.009 and the 95 percent confidence interval is [-0.06, 0.04], which rules out even small negative or positive effects on course scores. We can rule out positive peer effects larger than 0.04 of a standard deviation.

Table 2: Peer Effect Regressions for Course Outcomes

	(1) Numeric Score	(2) Grade (4-Point Scale)	(3) Dropped Course	(4) Passed Course
Own Pre-GPA	0.807***	0.282***	-0.041***	0.047***
	(0.038)	(0.017)	(0.008)	(0.008)
Partner's Pre-GPA	-0.009	-0.003	-0.008	0.006
	(0.024)	(0.011)	(0.007)	(0.007)
N	6,334	6,332	6,684	6,684
Dep. Var. Mean	0.003	3.827	0.051	0.947
Dep. Var. S.D.	1.000	0.440	0.220	0.225

Notes: Own Pre-GPA and partner's Pre-GPA are measured prior to term. Standard errors are clustered by course section. Additional controls include course-section fixed effects, indicators for race/ethnicity, gender, Educational Opportunity Program (EOP), the most common majors (proposed or declared) and undeclared major, and prior units accumulated. *** p<0.01, ** p<0.05,* p<0.1.

The results are similar when we examine additional measures of performance in Chemistry labs. Specification 2 reports estimates in which the dependent variable is the grade in the course on a 4-point scale. Using this alternative measure of course performance, the estimates also indicate that students are unaffected by the ability level of their randomly assigned partner.

Peer interactions might operate along different channels than course performance. For example, students might decide to drop Chemistry labs when randomly assigned to a weaker student before a score or grade is recorded in the system. Specification 3 reports regression estimates for whether the student dropped the course. We find that students are not more likely to drop the lab course when partnered with low ability students. The results are important because this is a behavioral variable too, which might be easier to influence than an ability/performance measure such as a grade or test score. But, we continue to find no effect. The drop decision, however, must be made in the first couple of lab meetings and thus represents more of a first impression than a true peer effect from interacting throughout the quarter.

Finally, Specification 4 reports peer-effect estimates for passing and completing the course. Failure to complete the required lab course could have subsequent consequences such as disrupting the student's trajectory in the major or even causing the student to lose interest in Science, pre-med or STEM more generally. We continue to find no evidence of peer effects.[14][15]

The null estimates are not likely due to attenuation bias from measurement error. In our setting, the peer's mean ability level is calculated directly from the microdata. All measures are from the same administrative data source and calculated in exactly the same way.[16]Although measurement error does not create a problem in this sense because of random assignment, finding the best measure of what peer characteristic affects student outcomes is difficult. We examine alternative measures of ability using pre-determined characteristics and outcome-on-outcome regressions.

Alternative Measures of Ability

Another concern especially in using the pre-determined model is peer effects might be underestimated if a good measure of student ability is unavailable. We estimate course outcome regressions using an alternative measure of student's academic ability prior to the course -- the student's and partner's grade in Chemistry 1A. Chemistry 1A is the first lecture course taken in the introductory sequence and is taken in a term prior to enrollment in the labs. Table 3 reports estimates of Equation (3.3) replacing own and partner's prior GPA with own and partner's Chem 1A grade. As expected, grades in Chemistry 1A are also a strong predictor of own performance in the lab. The estimates of peer effects are similar to those reported in Table 2 using prior GPA. For the dropped course specification, the point estimate is negative and statistically significant, but remains small (which is consistent with this variable capturing a first impression and not peer

[14] We also estimate the same set of regressions as reported in Table 2 adding the other section-mates' mean prior GPA. We find no evidence of a peer effect from the other students in the section. In this case, we replace section fixed effects with course-term fixed effects because of collinearity. If there was absolutely no variation in section sizes, for example, then it would be impossible to separately estimate the coefficients on own prior GPA and other section-mates' mean prior GPA (Guryan et al. 2009).

[15] We also estimate a model in which we measure partner's ability as whether the student's prior GPA is higher or lower than the partner's prior GPA (i.e. a rank measure, Elsner, Isphording and Zölitz 2021). We find no evidence of peer effects with or without a pair fixed effect.

[16] Differential measurement error could be caused by when the peer group mean is measured from a different source of data than the individual's own measure. For example, the peer group mean could come from administrative data from a school (with less measurement error) and the individual variable could come from a student survey (with more measurement error). Another example would be the case in which a student reports their own value of X and also reports the group mean of X on a survey (which they know less about). A third example is when the two measures come from the same source of data, but the peer mean is manipulated in some way (e.g. removing outliers that my reveal a student's identity) and thus differs from the raw average.

effects throughout the entire term). We return to focusing on prior GPA to avoid losing 6.6 percent of the sample and to capture more of the student's prior course performance.[17]

Table 3: Peer Effect Regressions for Course Outcomes using Partner's Prior Chem 1A Grade

	(1) Numeric Score	(2) Grade (4-Point Scale)	(3) Dropped Course	(4) Passed Course
Own Chem 1A Grade	0.262***	0.092***	-0.016***	0.019***
	(0.018)	(0.008)	(0.004)	(0.004)
Partner's Chem 1A Grade	-0.018	-0.007	-0.005*	0.004
	(0.013)	(0.007)	(0.003)	(0.003)
N	5,913	5,912	6,203	6,203
Dep. Var. Mean	0.011	3.830	0.045	0.952
Dep. Var. S.D.	1.000	0.440	0.208	0.214

Notes: Own Chem 1A Grade and partner's Chem 1A Grade are measured on 4-point scale and prior to term. Additional controls include course-section fixed effects, indicators for race/ethnicity, gender, Educational Opportunity Program (EOP), the most common majors (proposed or declared) and undeclared major, and prior units accumulated. *** p<0.01, ** p<0.05,* p<0.1.

Bounds on Course Scores

The negative coefficient on the peer effect in the dropped course regression suggests that there might be selection into who receives a grade in the class. If being paired with a low ability peer makes a student more likely to drop a course then the students who remain in the class and ultimately receive a grade might be less prone to peer effects. There is no way to directly test this assertion so we instead conduct a bounds analysis. Specifically, we impute grades for those students who dropped (5.3 percent of the sample) by using Equation (3.3) with different hypothesized values of the peer effect. We then add these imputed observations to the non-dropped sample and re-estimate the equation using the combined sample of dropped and non-dropped observations. Appendix Tables 2 and 3 report estimates. We simulate peer effects for dropped students of 0.03, 0.10 and 0.20 of a standard deviation from a 1-unit change in the peer's prior GPA or Chemistry 1A grade (e.g. equivalent of going from a B to an A). We view these as providing a wide range of large positive examples because a 1-unit change in the peer's prior GPA and Chemistry 1A grade only results in an estimated 0.03 standard deviation effect on dropping the course. The bounds analysis reveals that adding the dropped students back into the course score sample and assuming that those students would have had very large peer effects

[17] Using prior GPA we find a higher R^2 and higher t-statistic on the own ability measure coefficient.

continues to indicate null estimates of peer effects. The estimates also continue to rule out large positive peer effects.

We also examine bounded estimates assuming that there is negative selection into the sample of dropped students. Although students who dropped a lab course are lower ability on average than students who did not drop, the differences are very small (3.20 compared to 3.22 for prior GPA, and 2.80 compared with 2.91 for the Chem 1A grade). We assume an extremely high level of negative selection on unobservables that essentially adds a full -0.35 of a standard deviation to the imputed value of the course score in addition to the predicted value which already accounts for differences along observables. Column 5 of Appendix Tables 2 and 3 report estimates using this assumption in addition to the highest level of peer effect (0.20). Even with both of these extreme assumptions we continue to find small upper bounds on the peer effect for score in the lab course.

The lack of sensitivity in the bounds analysis might be due to only 5.3 percent of the sample dropping the lab course and the dropped sample and non-dropped sample do not differ in partner's ability. Overall, the bounds analysis does not reveal concerns that selection on dropping or not dropping the class is creating a bias in the lab score peer-effect estimates.

V. Outcome-on-Outcome Models, Mechanical Problems, and New Leave-out-Observation Estimation Technique

In this section, we discuss estimation issues that partly depend on the type of peer effect being estimated. All of these problems apply to situations with or without the random assignment of peer groups.

Outcome-on-outcome estimates

We push the model further by turning to estimating outcome-on-outcome estimates of peer effects. The major advantage of including the partner's outcome on the right-hand side of the equation instead of the partner's previous GPA (or another measure) is that performance in the lab might more accurately capture the potential ability of the peer to affect the student's performance in the lab. The partner's pre-GPA or grade in the introductory lecture Chemistry course might be great proxies for ability, but might be less specific to ability in the Chemistry lab course. Thus, the null estimates that we find might be missing an actual peer effect.

Although random assignment in our field experiment addresses concerns over selection into groups it does not directly address problems with reflection and common shocks. Common shocks at the section or TA level are controlled for by the section specific fixed effect. Perhaps weaker common shocks might occur to the student pair, but it is difficult to come up with examples of potentially strong ones that would create substantial and even a detectable bias in estimates. Examples of potential problems such as breaking equipment, faulty equipment, or illness from working together are rare and probably not influential on average course performance. Even in cases where there are lots of small shocks over the quarter they would tend to cancel out (i.e. positive and negative shocks) and not all work in the same direction, and these common shocks would have to widespread across many student-pairs instead of just a few isolated pairs.[18] Thus, the reflection problem is likely to be the main concern in our application of the outcome-on-outcome model.

Following our previous specification (Equation 3.3) we estimate the following equation that replaces the partner's pre-determined ability measure, A_{-ig}^{PT} with the partner's outcome in the lab course, Y_{-ig}^{PT}:

(5.1) $\mathrm{Y}_{ig} = \beta_0 + \beta_1 A_{ig} + \gamma Y_{-ig}^{PT} + \theta' \mathrm{W}_{ig} + \lambda_s + \varepsilon_{ig}$.

The model is useful in that it can detect whether a peer effect exists using this measure of ability. Table 4 reports estimates for the same set of course outcomes as reported in Table 2. We denote these estimates as "uncorrected" as discussed in more detail below. We find some coefficients that are negative and statistically significant. But these estimates do not necessarily imply that we have negative outcome-on-outcome peer effects. We show below that these negative estimates are instead due to "mechanical problems" noted in Angrist (2014), and that the standard regression model provides biased estimates in this case (even in the absence of peer effects).

[18] We explore how common shocks affect our simulation estimates and find that common shocks need to be very large in magnitude, frequent, and line up in sign to have a discernable effect.

Table 4: Peer Effect Regressions for Course Outcomes using Partner's Course Outcome (Uncorrected Estimates)

	(1) Numeric Score	(2) Grade (4-Point Scale)	(3) Dropped Course	(4) Passed Course
Own GPA	0.804***	0.277***	-0.038***	0.045***
	(0.039)	(0.017)	(0.008)	(0.008)
Partner's Numeric Score	0.006			
	(0.019)			
Partner's Grade (4-Pt. Scale)		-0.009		
		(0.020)		
Partner Dropped Course			-0.055***	
			(0.019)	
Partner Passed Course				-0.049**
				(0.020)
N	6,072	6,071	6,755	6,755
Dep. Var. Mean	0.003	3.828	0.051	0.947
Dep. Var. S.D.	0.996	0.436	0.220	0.225

Notes: Regression estimates are not corrected for mechanical problems or endogeity. Own GPA is measured prior to term. Standard errors are clustered by course section. Additional controls include course-section fixed effects, indicators for race/ethnicity, gender, Educational Opportunity Program (EOP), the most common majors (proposed or declared) and undeclared major, and prior units accumulated. *** p<0.01, ** p<0.05,* p<0.1.

Exploring Problems with Mechanical Links in the Estimation of Peer Effects

We next explore potential problems arising because of the mechanical links associated with the inclusion of repeated observations for the student and their peers (Angrist 2014). We examine this problem in the context of estimating peer effects in our setting and explore three methods that attempt to correct the bias. We also discuss these options in the case of the main model, the outcome-on-outcome model, and related randomization/sorting checks.[19] Importantly, we propose a new estimation method that satisfies Angrist's (2014) suggestion that the best approach to estimating peer effect models is to make a "clear distinction between the subjects and their peers," which "eliminates mechanical links between own and peer characteristics." We propose a straightforward and novel technique where we include only one observation per peer group, which addresses the intuition of this suggestion.

Before returning to the outcome-on-outcome model and whether the coefficient estimates reported in Table 4 are biased, we follow Angrist (2014) by first demonstrating the bias in a peer effect model that does not include the student's own pre-course GPA. At first glance, given that we have random assignment, the exclusion of this variable should not affect estimation of the peer effect from the lab partner because the student's prior GPA is not correlated with the

[19] Guryan et al. (2009) note the problem in the context of a randomization/sorting check.

partner's prior GPA. Specifically, we estimate the following, simplified exogenous peer effect model:

(5.2) $Y_{ig} = \beta_2 A_{-ig}^{PT} + \lambda_s + \varepsilon_{ig}$.

Table 5 reports estimates of Equation (5.2) using the course numeric score variable (i.e. column 1 of Table 2). The coefficient on the partner's prior GPA is negative and statistically significant. This contrasts sharply with the precisely estimated zero coefficient reported in Column 1 of Table 2, and is due to the removal of student's own ability, A_{ig}, from the equation.

Table 5: Peer Effect Regressions and Simulations - No Correction for Mechanical Bias

	(1) Y on PA	(2) Y on PA	(3) Y on PA	(4) Y on PY	(5) Y on PY	(6) Y on PY	(7) X on PX	(8) X on PX
Data Source	Actual	Simulated	Simulated	Actual	Simulated	Simulated	Actual	Simulated
Target Value for Peer Effect if M.C. Simulation		0.08	0.00		0.25	0.00		0.00
Peer Pre-GPA (PX)	-0.088***	0.033	-0.047				-0.088***	-0.059
	(0.030)	(0.002)	(0.002)				(0.019)	(0.001)
Peer Numeric Score (PY)				-0.013	0.423	-0.060		
				(0.022)	(0.001)	(0.001)		
Total Observations	6334	1,800,000	1,800,000	6072	1,800,000	1,800,000	6334	1,800,000
Peer Groups	3,167	900,000	900,000	3,036	900,000	900,000	3,167	900,000
Urns (Sections)	421	101,245	101,245	422	101,245	101,245	421	101,245
Variation in Urn Size	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep. Var. Mean	0.003	0	0	0.001	0	0	3.215	3.21
Dep. Var. S.D.	1.000	1	1	1.000	1	1	0.463	0.42

Notes: The dependent variables are numeric score (Columns 1-6) and own Pre-GPA (Columns 7-8). All regressions include course-section fixed effects. Y on PA is the exogenous peer effect model, Y on PY is the outcome-on-outcome model, and X on PX is the randomization/sorting check for student and partner's ability. Standard errors are clustered by section. Significance levels are denoted only for regressions using actual data: *** p<0.01, ** p<0.05, * p<0.1.

The negative estimates reported in Table 5 indicate the inherent negative bias when regressing Y_{ig} on the peer group leave-out-mean, $\bar{A}_{-ig}$ (which is A_{-ig}^{PT} in our case where the peer group is the one partner). Intuitively, the problem is that mechanically when A_{ig} (and thus Y_{ig}) is relatively large $\bar{A}_{-ig}$ is more likely to be relatively small because A_{ig} is being drawn from the section without replacement. Conversely, when A_{ig} is relatively small then $\bar{A}_{-ig}$ is likely to be relatively large because that observation is not being replaced in the section, which is the "urn" from which both values are drawn.[20] This is due to the nature of peer effect estimation that necessitates sampling without replacement or conditional sampling – a student cannot be their own peer.[21] Thus, there is a mechanical negative correlation within peer groups.

We confirm the negative bias with a Monte Carlo simulation reported in Column 2 of Table 5. We create normally distributed random variables for each student's ability and error term. Y is a function of own ability, partner's ability, and the error term according to Equation (3.2). To demonstrate that the bias exists, even in the case of a peer effect, we create a positive peer effect (β_2) between the pairs in the simulation equal to 0.08 (which is 10 percent of the own effect). The mean section or "urn" size is 18, which is the lab size in our setting. We generate results for a total sample size of 1,800,000 observations with some variation in urn sizes. We confirm the negative bias on the peer effect coefficient. We find a coefficient of 0.033 which is smaller than the simulated 0.08 peer effect. In Monte Carlo simulations that assume zero peer effects (shown in Column 3) the simulated estimate is -0.047, which captures the same downward bias as Column 2.[22]

Mechanical Bias in the Monte Carlo Outcome-on-Outcome Model

We turn to exploring potential biases in the outcome-on-outcome model. In Table 5 using actual data we find a negative, but statistically insignificant point estimate for the model that regresses the student's course score on the partner's course score without controlling for own prior GPA (i.e. analogous to Equation 5.2). Because we do not know what the potential outcome-on-outcome peer effect is in our experiment, we cannot identify the negative bias component.

On the other hand, we can simulate a known relationship between student's outcome and peer's outcome for lab partners. These Monte Carlo simulation estimates are useful for demonstrating potential biases. For the outcome-on-outcome model we use the following structural equation:

[20] A simple illustrative example can be seen from the case of four students with the following prior GPAs: 3.0, 3.3, 3.7, and 4.0. The expected prior GPA for the partner of each student is 3.7, 3.6, 3.4, 3.3, respectively, clearly creating a negative relationship.
[21] Although there is substantial variation across sections potentially partly fixing the problem, the inclusion of section fixed effects soaks up all of the between-section variation.
[22] The estimated bias is consistent with the conditional expectation formula noted above where the own effect of ability on course scores is 0.8. If the relationship was one-to-one instead, the bias would be -0.059.

(5.3) $Y_{ig} = \beta_1 A_{ig} + \gamma Y_{-ig}^{PT} + \lambda_s + \varepsilon_{ig}$.

But, we cannot estimate this model directly because the student and the student partner's outcomes are interdependent, and thus not constant in the Monte Carlo simulations. To estimate the model, we first solve for the reduced form equation for Y_{ig}:

(5.4) $\mathrm{Y}'_{ig} = \frac{\beta_1 A_{ig} + \gamma \beta_1 A_{-ig}^{PT} + \gamma \varepsilon_{-ig}^{PT} + \lambda_s + \varepsilon_{ig}}{(1-\gamma)^2}$

We then use Equation (5.4) to create new values of Y for each observation in the data and include those new reduced form values to estimate:

(5.5) $\mathrm{Y}'_{ig} = \gamma Y'^{PT}_{-ig} + \lambda_s + \omega_{ig}$,

which will suffer from positive reflection bias if there is a peer effect, but nevertheless will illustrate other biases resulting from the peer effect data and estimation.

Estimates are reported in Column 5 of Table 5. We find an outcome-on-outcome peer effect estimate of 0.423 which is larger than the simulated 0.25 peer effect. A downward bias from the mechanical problem, however, is concealed by the upward bias due to the reflection problem. In fact, in our simulation the upward bias dominates resulting in a net upward bias. We can demonstrate the negative bias from the mechanical problem by removing the reflection problem by setting peer effects to zero. Column 6 reports Monte Carlo estimates in the case of no peer effects. We find a negative bias: the coefficient estimate is -0.060.

Mechanical Bias in the Randomization Check

We can also demonstrate a downward bias in the randomization/sorting check as noted in Guryan et al. (2009). The sorting check involves a regression of own X on peer X where X could include ability or any other student characteristic in Equation (5.2). Column 7 of Table 5 reports estimates when we use own ability, A_{ig}, as the dependent variable and partner's ability, A_{-ig}^{PT}, as the independent variable for the sorting check. We find a negative and statistically significant coefficient estimate on the peer variable even though we have random assignment of lab partners. Monte Carlo simulations confirm the anticipated downward bias in the sorting test. Column 8 of Table 5 indicates a negative coefficient of -0.059.

Estimation Solutions to Correct Mechanical Bias

We explore three options to address the problem. First, we follow a similar suggestion as the one made by Guryan et al. (2009) for their randomization or sorting tests. Specifically, they argue for including the leave-me-out mean of the "urn" (LMOUM) from which individuals are randomly drawn to create pairs. We expand on their suggestion for sorting tests to examine how including the LMOUM regressions perform in all three of our applications. We estimate the peer effect on our outcome measure:

(5.6) $Y_{ig} = \beta_2 A_{-ig}^{PT} + \beta_3 \bar{A}_{-is} + \lambda_s + \varepsilon_{ig},$

where $\bar{A}_{-is}$ is the leave-me-out mean for the section or urn. Estimates using this correction are reported in Table 6. Using the actual data (reported in Column 1) we find a small and insignificant peer effect estimate. The inclusion of the leave-me-out mean of the section/urn eliminates the negative coefficient reported in Table 5 which might be due to removing the downward bias. Monte Carlo estimates reported in Column 2 confirm the result. The regression returns the simulated peer effect after including the leave-me-out section mean. The technique works in the case of regressing own Y on peer's ability.

Table 6: Peer Effect Regressions and Simulations - Correction for Mechanical Bias by Including Leave-Me-Out Urn Mean

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Y on PA	Y on PA	Y on PA	Y on PY	Y on PY	Y on PY	Y on PY	X on PX	X on PX	X on PX
Data Source	Actual	Simulated	Simulated	Actual	Simulated	Simulated	Simulated	Actual	Simulated	Simulated
Target Value for Peer Effect if M.C. Simulation		0.08	0.08		0.25	0.00	0.25		0.00	0.00
Peer Pre-GPA (PX)	-0.014	0.080	0.080					0.004	0.000	N/A
	(0.026)	(0.001)	(0.001)					(0.003)	(0.000)	
Peer Numeric Score (PY)				0.006	0.009	0.000	N/A			
				(0.019)	(0.000)	(0.000)				
Total Observations	6,334	1,800,000	1800000	6,072	1800000	1,800,000	1800000	6,334	1800000	1,800,000
Peer Groups	3,167	900,000	900,000	3,036	900,000	900,000	900,000	3,167	900,000	900,000
Urns (Sections)	421	101,245	100,000	422	101,245	101,245	100,000	421	101,245	100,000
Variation in Urn Size	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No
Dep. Var. Mean	0.003	0	0	0	0	0	0	3.215	3.21	3.21
Dep. Var. S.D.	1.000	1	1	1	1	1	1	0.463	0.42	0.42

Notes: The dependent variables are numeric score (Columns 1-7) and own Pre-GPA (Columns 8-10). All regressions include course-section fixed effects. Y on PA is the exogenous peer effect model, Y on PY is the outcome-on-outcome model, and X on PX is the randomization/sorting check for student and partner's ability. Standard errors are clustered by section. All regressions also include the leave-me-out urn mean which is the average of X or Y for the other students in the lab section. Significance levels are denoted only for regressions using actual data: *** p<0.01, ** p<0.05, * p<0.1.

Turning to the outcome-on-outcome model, we find a null estimate using the actual data (Column 4 of Table 6). When we estimate the outcome-on-outcome model using the simulated data we find that the estimates are biased with a simulated positive peer effect (Column 5) but not with a simulated zero peer effect (Column 6). Similar to Guryan et al. (2009) we find that including the leave-me-out urn mean removes the bias for the randomization/sorting test of regressing own X on partner's X (Column 9).

The Monte Carlo simulations, however, reveal different problems in estimation. The use of the leave-me-out urn mean relies on variation in urn sizes. If we had a fixed section size of 18 students, we would not be able to include the leave-me-out mean for the urn when we control for section fixed effects in the outcome-on-outcome model because of the created linear dependence (Column 7). Similarly, for the randomization/sorting test (i.e. X on PX regression), fixed section sizes cause problems for estimation using this method (Column 10).[23] In Guryan et al. (2009)'s original empirical application and Monte Carlo simulations, there was variation in urn sizes that addressed this problem.

For the randomization check using the actual data we find a null estimate of the relationship between own prior GPA and partner's GPA. But, it should be noted that there are many cases in which classrooms have fixed sizes or very little variation in classroom sizes (which is our case with the 18-student lab courses). In our setting a few lab courses have fewer than 18 students creating minimal variation but enough variation to include leave-me-out section means in the regressions.

Replacing the Leave-Me-Out Urn Mean with Own Ability

The second method that we explore is adding own ability, A_{ig}, to the equation instead of the LMOUM. For the Y on PA model this is the same as estimating Equation (3.4) and was reported in Specification 1 of Table 2 (although we remove the controls for student characteristics). We report the coefficient estimates for convenience in Column 1 of Table 7, which show null peer effects and positive own effects of ability. The Monte Carlo estimate is reported in Column 2. The model performs the same as the regression that includes the

[23] As noted in Guryan et al. (2009), the coefficient on the leave-me-out section (urn) mean is closer to $-N_u$ (N_u -1), where N_u is the average section (urn) size.

LMOUM.[24] We recover the target simulated value of the peer effect (and for the own ability coefficient). In Column 3 we also add the leave-me-out section mean to show that the coefficient on own ability is now slightly biased. The simulated value is 0.80 and we find an estimated coefficient of 0.781.

Table 7: Peer Effect Regressions and Simulations - Correction for Mechanical Bias by Including Own X

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Y on PA	Y on PA	Y on PA	Y on PY	Y on PY	Y on PY	X on PX	X on PX
Data Source	Actual	Simulated	Simulated	Actual	Simulated	Simulated	Actual	Simulated
Target Value for Peer Effect if Monte Carlo Simulation		0.08	0.08		0.25	0		0
Peer Pre-GPA (PX)	-0.017	0.080	0.080				NA	NA
	(0.025)	(0.001)	(0.001)					
Peer Numeric Score (PY)				0.004	0.397	-0.051		
				(0.019)	(0.001)	(0.001)		
Own Pre-GPA (X)	0.808***	0.799	0.781	0.804***	0.775	0.797	NA	NA
	(0.037)	(0.001)	(0.012)	(0.037)	(0.001)	(0.001)		
Total Observations	6,334	1,800,000	1,800,000	6,072	1800000	1800000	6,334	1,800,000
Peer Groups	3,327	900,000	900,000	3,327	900,000	900,000		900,000
Urns (Sections)	421	101,245	101,245	422	101,245	101,245	421	101,245
Variation in Urn Size	Yes	No	No	Yes	No	No	Yes	No
Dep. Var. Mean	0.003	0	0	0.001	0	0	3.215	3.21
Dep. Var. S.D.	1.000	1	1	1.000	1	1	0.463	0.42

Notes: The dependent variables are numeric score (Columns 1-6) and own Pre-GPA (Columns 7-8). All regressions include course-section fixed effects. Y on PA is the exogenous peer effect model, Y on PY is the outcome-on-outcome model, and X on PX is the randomization/sorting check for student and partner's ability. Standard errors are clustered by section. All regressions also include own Pre-GPA and Column 3 adds the leave-me-out urn mean. Significance levels are denoted only for regressions using actual data: *** p<0.01, ** p<0.05, * p<0.1.

For the outcome-on-outcome model the original estimate is reported in Colum 1 of Table 4 and repeated for convenience in Column 4 of Table 7. Columns 5 and 6 of Table 7 report the Monte Carlo estimates for different simulated values of peer effects. In both applications, the bias is downward. For the Monte Carlo simulation with zero peer effects (and thus no potential for reflection bias) the coefficient is -0.051. Including own ability does not address the mechanical bias from sampling without replacement because it does not match the measure of partner's ability in this equation. For the randomization check (i.e. regression of X on PX), we

[24] In the case of fixed urn sizes, the two methods provide identical estimates of the peer effect. Controlling for own ability and the section fixed effect is the same mathematically as controlling for the leave-me-out section/urn mean and the section fixed effect when there is no variation in section sizes.

cannot include own X because it is the dependent variable. Thus, this test is undefined, which is noted in Columns 7 and 8 to illustrate the inability to this correction here.

One-Observation per Peer Group Estimation Technique

Thus, we are left with models that remove the bias and are estimable in some cases, but not in others. Estimation of the outcome-on-outcome model is especially problematic across the different techniques, and as noted above it could reveal peer effects even when the exogenous model does not. We propose a new approach to correct for the mechanical link bias and a framework in which all three models can be estimated. We estimate Equation (3.4) randomly removing one student from each of the pairs in the sample. The removal of one observation from each pair eliminates the mechanical negative correlation within groups. To implement the procedure, we first estimate the regression randomly drawing half the full sample (one from each pair). We repeat the process randomly choosing one of the two students from each pair 1,000 times. We then calculate the average of both the coefficient estimate and standard error over these 1,000 iterations (which represents a conservative estimate of the standard error). This new approach that we take of removing observations is intuitive because it essentially mimics a random experiment in which one half of all Chemistry lab students are randomly chosen as participants in the experiment. This "experimental" or estimation sample of students then are randomly assigned varying levels of treatment intensity as measured by their partner's pre-course GPA. Student A gets randomly assigned a partner (Student B) with low to high ability. Student B's ability is the intervention in the experiment, and thus this student does not get included in the estimation sample. Progressing to the next student, Student C then gets randomly assigned a partner (Student D) with low to high ability. Student D does not get included in sample because that student's ability is the intervention.

Column 1 of Table 8 reports estimates from actual data for the exogenous peer effect model using the proposed include one-observation per peer group technique (OOPG). We find a null estimate for main peer effect that are similar to the estimate reported in Specification 1 of Table 2. Monte Carlo simulations that take a similar approach of including one observation from each pair recover the peer effect estimate in both the case of variation in urn sizes (Column 2) and the case of no variation in urn sizes (Column 3).

Table 8: Peer Effect Regressions and Simulations - Correction for Mechanical Bias by using One-Observation per Peer Group Method

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Y on PA	Y on PA	Y on PA	Y on PY	Y on PY	Y on PY	Y on PY	X on PX	X on PX	X on PX
Data Source	Actual	Simulated	Simulated	Actual	Simulated	Simulated	Simulated	Actual	Simulated	Simulated
Target value for Peer Effect if M.C. Simulation		0.08	0.08		0.25	0	0.25		0	0
Peer Pre-GPA (PX)	-0.035	0.080	0.080					-0.021	0.001	0.001
	(0.022)	(0.002)	(0.002)					(0.020)	(0.001)	(0.001)
Peer Course Score (PY)				0.060***	0.470	0.000	0.470			
				(0.022)	(0.001)	(0.001)	(0.001)			
Total Observations	6,334	1,800,000	1,800,000	6072	1800000	1800000	1,800,000	6,334	1800000	1,800,000
Peer Groups	3167	900,000	900,000	3036	900,000	900,000	900,000	3167	900,000	900,000
Urns (Sections)	421	101,245	100,000	422	101,245	101,245	100,000	421	101,245	100,000
Variation in Urn Size	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No
Dep. Var. Mean	0.003	0	0	0.001	0	0	0	3.215	3.21	3.21
Dep. Var. S.D.	1.000	1	1	1.000	1	1	1	0.463	0.42	0.42

Notes: The dependent variables are numeric score (Columns 1-7) and own Pre-GPA (Columns 8-10). All regressions include course-section fixed effects. Y on PA is the exogenous peer effect model, Y on PY is the outcome-on-outcome model, and X on PX is the randomization/sorting check for student and partner's ability. Standard errors are clustered by section. Reported estimates are mean values of coefficient estimates and standard errors from 1,000 regression iterations. Each regression iteration leaves out one randomly-chosen observation per peer group. Significance levels are denoted only for regressions using actual data: *** p<0.01, ** p<0.05, * p<0.1.

For the outcome-on-outcome model we find a small, positive and marginally statistically significant coefficient on the peer effect using actual data (Column 4). After removing the bias from mechanical links we find evidence of a positive effect but note that this estimate is likely upward biased from the reflection problem. The Monte Carlo simulations confirm this suspicion. In the case, with a positive peer effect (Column 5) we find a peer effect estimate of 0.47 which is larger than the target value of 0.25.[25] When we simulate no peer effect we find no peer effect in the outcome-on-outcome regression (Column 6). The OOPG technique removes the downward bias in the outcome-on-outcome regression. Also, we should note that we do not need to create variation in section/urn sizes for this technique to work (see Column 7) as was necessary using the leave-me-out urn mean regressions. But, the OOPG technique does not fix the reflection problem. We must continue to view this estimate as an upward bound estimate of peer effects, or at least proof that as evidence that there is some peer effect present (otherwise the simulation shows that it should be zero). We return to this discussion in the next subsection.

Table 8 also reports estimates of the randomization/sorting tests regressing X on PX. The OOPG technique using actual data finds a null relationship between student's own ability and

[25] Using an instrumental variable model in which we use A_{-ig}^{PT} as an instrument for Y'^{PT}_{-ig} we recover the 0.25 estimate confirming the downward mechanical bias in this specification. But as noted above the IV approach does not help us in the outcome-on-outcome model with the actual data because it simply rescales the coefficient from the exogenous peer effect model and does not provide anything new.

partner's ability which is consistent with random assignment. The Monte Carlo simulations also recover a null relationship (Column 9) and work even without variation in section sizes (Column 10) unlike the leave-me-out urn mean regressions.

This leave-out-one-observation procedure satisfies Angrist (2014)'s suggestion of distinguishing between the subjects of the study (i.e. the treatment group here) and the peers (i.e. the left-out partners). He notes that the best approach to estimating peer effect models is to make a "clear distinction between the subjects and their peers," which "eliminates mechanical links between own and peer characteristics." Each iteration of the model removes the peer from the sample thus creating a clear distinction between the student and the experimentally varied peer.

The technique that we propose is useful in applications in which the leave-me-out mean of X for the entire urn is not available, and only a subsample of observations randomly drawn from that urn are available. In this case including both observations for the pair in the sample is problematic and cannot be fixed because the urn leave-me-out mean is not available. Another case is where the urn sizes are constant with no variation in size in the full sample. In this case the section or urn leave-me-out mean cannot identify the peer effect. Finally, OOPG could provide a robustness check for estimates that control for the leave-me-out urn mean.

Estimating Outcome-on-Outcome Model with One-Observation per Peer Group Technique

Returning to actual data and multiple outcomes, Table 9 reports estimates of the outcome-on-outcome model using the OOPG technique (i.e. the same specifications as those reported in Table 4). The coefficients on the peer effect estimates change. For the total numeric score and grade point we now find a positive and statistically significant estimate of the peer effect. For dropping and passing the course, however, we do not find evidence of a peer effect which might be due to the limited amount of time of exposure for dropping out of the course. After correcting for the potential downward bias from mechanical problems, the outcome-on-outcome regressions now provide evidence of positive peer effects. Although the estimates are likely upward biased from the reflection problem there is at least some evidence that a peer effect exists albeit a small one.[26] As shown above when there is no simulated peer effect in the outcome-on-outcome model the OOPG technique correctly recovers a peer effect estimate of zero (see Table 8, Column 6). In fact, the new OOPG technique that we introduce here is the

[26] As noted above, students are not receiving the same grades on any of the subgroups of assignments in the course.

only technique that can verify the null peer effect in the Monte Carlo simulations. Thus, we find some evidence of small positive peer effect. The point estimate is modest, implying that a 1 standard deviation increase in the partner's score in the course increases the student's score by only 0.066 standard deviations (and this is an upper-bound estimate). The precision of the estimate also allows us to rule out large effects. We can rule out effect sizes as large as 0.106 standard deviations.

Table 9: Peer Effect Regressions for Course Outcomes using Partner's Course Outcome - Correction for Mechanical Bias by using One-Observation per Peer Group Method

	(1) Numeric Score	(2) Grade (4-Pont Scale)	(3) Dropped Course	(4) Passed Course
Own Pre-GPA	0.807***	0.278***	-0.039***	0.046***
	(0.054)	(0.024)	(0.011)	(0.011)
Partner's Numeric Score	0.066***			
	(0.020)			
Partner's Grade (4-Pt. Scale)		0.057***		
		(0.021)		
Partner Dropped Course			0.010	
			(0.020)	
Partner Passed Course				0.015
				(0.020)
N	6,072	6,071	6,755	6,755
N (per Regression Iteration)	3036	3035.5	3377.5	3377.5
Dep. Var. Mean	0.003	3.828	0.051	0.947
Dep. Var. S.D.	0.996	0.436	0.220	0.225

Note: Standard errors are clustered by section. Reported estimates are mean values of coefficient estimates and standard errors from 1,000 regression iterations. Each regression iteration leaves out one randomly-chosen observation per peer group. Additional controls include lab-section fixed effects, race/ethnicity, gender, Educational Opportunity Program (EOP) status, prior units accumulated, the most common majors (proposed or declared), and undeclared major. *** p<0.01, ** p<0.05,* p<0.1.

The OOPG outcome-on-outcome estimates provide an upper-bound estimate of the peer effect. This is because the technique cannot remove the potential reflection problem. It is useful to note that an instrumental variables approach does not help. The main instrument we have for PY is partner's ability (PA). Thus, in the OOPG estimation of Y on PY we are basically trying to focus on the part of PY that is independent of PA which would capture something about how a student could be good at a lab or not good at a lab independent of what is predicted by their prior GPA or Chem1A grade. The IV estimate using PA simply scales up the coefficient that we find in the Y on PA regressions.[27]

[27] For example, we find an IV estimate of -0.14 which is essentially the exogenous peer effect estimate from Table 2 (-0.009) scaled up by the regression of PY on PA which is similar to the own GPA coefficient estimate from Table 2 (0.807).

Summary of Estimation Strategies

We briefly summarize the applicability of the different estimation strategies in each application. The inclusion of section-leave-out means works in only some applications and is problematic if the urn sizes are fixed. Including own A adjusts for the problem with sampling without replacement in the exogenous peer effect model, but not in the outcome-on-outcome model. In contrast to these limitations, the leave-out-observation technique is estimable in all applications even in the case of a fixed urn size.

	Regression in Context of No Peer Effects and Fixed Urn Size		
Method	Y on PA	Y on PY	X on PX
1. Leave-me-out urn mean (LMOUM)	Estimable	Unestimable	Unestimable
2. Control for own A (or X)	Estimable	Problematic	Unestimable
3. One-observation per Peer Group (OOPG)	Estimable	Estimable	Estimable

VI. More Evidence on the Reflection Problem, Timing, Course Assignments, and Longer-Term Outcomes

In this section, we use the detailed administrative information on grades and other academic outcomes to explore the timing of peer effects, course assignments, and longer-term outcomes. We estimate both exogenous and outcome-on-outcome peer effect models.

Timing of Peer Effects and Course Assignments

We address the reflection problem by taking advantage of rich data from coursebooks that include weekly scores on assignments.[28] Using information on weekly assignments we split the term into the first half and the second half, and calculate separate scores for each half (and separately normalize each score to have mean zero). Scores on assignments and exams from the second half of the quarter can be regressed on scores from the first half of the quarter because

[28] We lose some observations because complete information on assignment scores are not always entered in the gradebooks even when a final total score for the course is entered.

earlier scores cannot be reflected back on later scores. But, the advantage here is that we have a measure of a partial outcome in the course for the partner's peer effect instead of the partner's GPA. The disadvantage, however, is that learning from the first half of the term might directly impact performance in the second half of the term and the first half score might suffer from reflection. Table 10 reports separate estimates of Equation (3.2) using the score from the second half of the quarter as the dependent variable, and own and partner's scores from the first half of the quarter as the key independent variables.[29] We estimate the model using the full sample and do not use the OOPG technique because including own first-half score corrects for the downward bias of sampling without replacement as noted above. We find a large, positive coefficient estimate on the student's own score for the first half of the course but no evidence of a peer effect from the partner's first half score in the course.

Table 10: Peer Effect Regressions for First- and Second-Half of Quarter Course Scores

	(1) Second-Half Score	(2) First-Half Score	(3) Second-Half Score	(4) Second-Half Score
Own First-Half Score	0.432***			0.430***
	(0.044)			(0.058)
Partner's First-Half Score	-0.013	0.095***		
	(0.014)	(0.021)		
Partner's Second-Half Score			0.037	0.022
			(0.032)	(0.027)
Own Pre-GPA	0.344*	0.725***	0.555***	0.243***
	(0.205)	(0.056)	(0.060)	(0.054)
Technique	LMOUM	OOPG	OOPG	OOPG
N	5,156	5,156	5,156	5,156
Dep. Var. Mean	0.001	0.001	0.001	0.001
Dep. Var. S.D.	1.006	1.006	1.006	1.006

Notes: Standard errors are clustered by section. Additional controls include course-section fixed effects, indicators for race/ethnicity, gender, Educational Opportunity Program (EOP), the most common majors (proposed or declared) and undeclared major, and prior units accumulated. *** p<0.01, ** p<0.05,* p<0.1.

We also estimate separate outcome-on-outcome models for peer effects using first-half scores and second-half scores (reported in Columns 2 and 3, respectively). Peer effects might differ between the first half of the term when students do not know each other as well and the second half of the term after students have known each other for at least five weeks. In this

[29] We further investigate these issues by looking at specific groups of assignments. These assignments are not necessarily concentrated over any part of the term. Similar groupings as Figure 2 above. We find no evidence of exogenous peer effects on any of the components to the total score in the course (see Appendix Table 4).

model we use the OOPG technique and include the student's own GPA. We find evidence of positive peer effects using either the first half scores or the second half scores. In both cases, the estimates do not imply large effects and likely capture some reflection bias but provide evidence of the existence of peer effects over both parts of the term. In Column 4, we estimate the outcome-on-outcome model for the second half scores and include the student's first half score as an additional variable. In this case, the student's GPA and first half score predict success in the second half of the course, but there continues to be an influence of the student's partner in the course. We continue to find evidence of a small, positive peer effect in this outcome-on-outcome model.

Longer-Term Outcomes

We run regressions for several additional outcome measures in addition to the main course outcome measures. Peer interactions may have longer-term consequences such as dissuading students from continuing in Chemistry and STEM. Estimates of the base model are reported for four additional outcomes in Table 11. First, we estimate whether low-ability partners negatively influence subsequent course taking in Chemistry and STEM courses by students. Specification (1) reports estimates of peer effects when the dependent variable is future enrollment in the secondary sequence in organic chemistry. Second, we switch the focus to major declaration. Specification (2) explores declaring a major in Chemistry, and Specification (3) explores declaring a major in any STEM field. Specification (4) explores the impacts on grades in the large lecture Chemistry courses taken concurrently with the lab. The partner in the lab course might help the student in other Chemistry courses that term. For all of these measures of longer-term interest in continuing in STEM, we do not find evidence that students are positively affected when partnered with high-ability students. Only the point estimate on partner's GPA in the takes organic chemistry regression is marginally significant at the 0.10 level but is negative and small in magnitude. All of these estimates of peer interactions with partners in the labs on longer-term outcomes are consistent with the finding for immediate outcomes such as course scores, grades and dropping the course.

Table 11: Peer Effect Regressions for Longer-Term Outcomes

	(1) Takes Organic Chemistry	(2) Declared Chemistry	(3) Declared STEM	(4) Chemistry Lecture Course Grade
Own Pre-GPA	0.125***	0.042***	0.075***	1.059***
	(0.013)	(0.009)	(0.013)	(0.026)
Partner's Pre-GPA	-0.022*	-0.006	-0.016	-0.022
	(0.012)	(0.008)	(0.013)	(0.022)
N	6,684	5,562	5,562	5,447
Dep. Var. Mean	0.590	0.102	0.703	2.948
Dep. Var. S.D.	0.492	0.303	0.457	0.851

Notes: The dependent variables are future enrollment in the secondary sequence in organic chemistry (Column 1), declaring a major in Chemistry (Column 2), declaring a major in any STEM field (Column 3), and grade in a large lecture Chemistry course taken concurrently with the lab (Column 4). Standard errors are clustered by course section. Additional controls include course-section fixed effects, indicators for race/ethnicity, gender, Educational Opportunity Program (EOP), the most common majors (proposed or declared) and undeclared major, and prior units accumulated. *** p<0.01, ** p<0.05,* p<0.1.

VII. Exploring the Linear-in-Means Model, Quantile Treatment Effects, Heterogeneity by Own Ability, Non-Linear Effects, and Influential Peers

All of the estimates discussed thus far are from the workhorse linear-in-means model of peer effects. Peer effects are modeled as linear in partner's ability and capture mean impacts instead of differential effects across the post-treatment outcome distribution or the pre-treatment ability distribution. An additional commonly-made assumption in estimating peer effects is that the mean ability of the peer group (instead of the 75th percentile ability or the mean of the 75-99th percentile, for example) is the appropriate measure to estimate the peer group effect. Finally, it is commonly assumed that students interact with all of their potential peers in the classroom with equal weights, and there is no sorting into subgroups within the classroom. [30] We explore each of these assumptions.

We first explore whether there are differential treatment effects across the post-treatment grade distribution that could be hidden by focusing only on mean impacts. Peers could shift the grade distribution at other points in the distribution, and the finding of a zero average treatment effect could mask offsetting effects at the bottom and top of the grade distribution. We address this concern by estimating quantile treatment effects. Appendix Figure 2 displays estimates throughout the distribution. There is no evidence of peer effects at different points of the distribution.

Second, there might be heterogeneity in peer or "treatment" effects by the academic ability of the student. For example, low-ability students might benefit the most from peer effects,

[30] For example, in a classroom of 30 students the underlying assumption for using mean ability is that every peer student has an equal, but 1/29 effect on the student.

whereas high-ability students might be unaffected by their lab partner's ability because they will perform well regardless of who they are teamed up with in the class. We estimate different peer effects by the pre-treatment GPA distribution for students. The following regression is estimated:

(7.1) $\mathrm{Y}_{ig} = \beta_0 + \beta_1 A_{ig} + \sum \beta_2^p D_{ig}^p A_{-ig}^{PT} + \sum \beta_3^p D_{ig}^p + \theta' \mathrm{W}_{ig} + \lambda_s + \varepsilon_{ig}$,

where D_{ig}^p is an indicator for whether student i is in the pth percentile of the pre-course GPA distribution. Percentiles are restricted to five different categories (i.e. quintiles). Thus, β_2^p are estimates of peer effects on students of different ability levels.

Table 12 reports estimates of Equation (7.1). Column 1 reports estimates for the exogenous peer effect regression and Column 2 reports estimates for the outcome-on-outcome peer effect regression. Consistent with the main results for peer effects in the exogenous model we do not find evidence that any student ability group is affected by peer's ability as measured by prior GPA. For the outcome-on-outcome peer effect model, however, we find evidence of positive peer effects for most of the student ability groups. The highest ability group appears to be unaffected by their peer's ability as measured endogenously using the partner's score in the course.

Table 12: Peer Effect Regressions: Heterogenous Effects by Own Ability and Non-Linear Peer Effects

	(1) Using Pre-GPA	(2) Using Numeric Score from Course	(3) Using Pre-GPA	(4) Using Numeric Score from Course
Own Pre-GPA	0.919***	0.808***	0.807***	0.806***
	(0.157)	(0.054)	(0.038)	(0.054)
Partner's Peer Measure X Own Pre-GPA Quintile 1	0.020	0.129**		
	(0.039)	(0.050)		
Partner's Peer Measure X Own Pre-GPA Quintile 2	-0.009	0.071		
	(0.029)	(0.046)		
Partner's Peer Measure X Own Pre-GPA Quintile 3	-0.000	0.062		
	(0.025)	(0.038)		
Partner's Peer Measure X Own Pre-GPA Quintile 4	-0.024	0.102**		
	(0.028)	(0.045)		
Partner's Peer Measure X Own Pre-GPA Quintile 5	-0.031	-0.008		
	(0.037)	(0.023)		
Partner's Peer Measure is Low			-0.024	-0.118***
			(0.028)	(0.043)
Partner's Peer Measure is High			-0.039	0.016
			(0.028)	(0.043)
Technique	Standard	OOPG	Standard	OOPG
N	6,334	6072	6,334	6072
Dep. Var. Mean	0.003	0.003	0.003	-0.009
Dep. Var. S.D.	1.000	0.996	1.000	0.989

Notes: The dependent variable in all specifications is student's numeric score. Own Pre-GPA and partner's Pre-GPA are measured prior to term. Standard errors are clustered by course section. Additional controls include course-section fixed effects, indicators for race/ethnicity, gender, Educational Opportunity Program (EOP), the most common majors (proposed or declared) and undeclared major, and prior units accumulated. *** p<0.01, ** p<0.05,* p<0.1.

Third, there might be non-linearities in peer effects based on the academic ability of the peer. For example, there is the possibility that medium-ability peers have a positive effect relative to low-ability peers, whereas high-ability peers have no or even a negative effect relative to medium-ability peers (thus resulting in an inverted U-shaped relationship). We estimate the following regression:

$$(7.2)\ \mathrm{Y}_{ig} = \beta_0 + \beta_1 A_{ig} + \sum \delta^p A_{-ig}^{p,PT} + \theta'\mathrm{W}_{ig} + \lambda_s + \varepsilon_{ig}$$

where $A_{-ig}^{p,PT}$ is an indicator variable for whether the lab partner is in the *p*th percentile of the pre-course GPA distribution. Percentiles are restricted to 3 different groups. Thus, δ^p are estimates of peer effects from lab partners of different ability levels (leaving out the medium-ability group). The model has a simplified structure that is similar to a random experiment with three treatment arms and discrete treatments. We can think of students as being assigned to the control group (medium ability partner), treatment 1 (low-ability partner), and treatment 2 (high-ability partner).

Table 12 reports estimates of Equation (7.2) in Columns 3 and 4. In the outcome-on-outcome model we find some evidence that having a lower ability peer has a negative effect on student performance in the class relative to having a medium ability peer. We do not find evidence of a negative effect of high ability peers in this model.

Combining the previous two models, we take an additional approach that allows for interactions between students' and their lab partners' ability levels. To simplify we estimate a three-by-three matrix of peer effect estimates using the following equation:

$$(7.3)\ \mathrm{Y}_{ig} = \beta_0 + \beta_1 A_{ig} + \sum \delta^{p1,p2} D_{ig}^{p1} A_{-ig}^{p2,PT} + \theta'\mathrm{W}_{ig} + \lambda_s + \varepsilon_{ig}.$$

Each group of three ability levels for students is allowed to have its own peer effect measured as having a low-ability peer or a high-ability peer.

Table 13 reports estimates of Equation (7.3). Exogenous peer effect estimates are reported in Columns 1 and 3, and outcome-on-outcome estimates are reported in Columns 2 and 4. We find evidence from the outcome-on-outcome model that assignment to a low-ability peer has a negative effect on the performance of low-ability and middle-ability students. Low-ability peers do not affect the performance of high-ability students. Being matched with a high-ability peer does not increase or decrease course scores for any ability students. The positive peer effect found earlier in the outcome-on-outcome model and reported in Table 9 using the OOPG model appears to be driven primarily by the effects of being paired with low-ability peers. And, high-ability students are not affected by these low-ability peers and do well anyway in the course.

Table 13: Peer Effect Regressions: Matrix of Heterogenous and Non-Linear Peer Effects

	(1)	(2)	(3)	(4)
	Using Pre-GPA	Using Numeric Score from Course	Using Pre-GPA	Using Numeric Score from Course
Own Pre-GPA	0.811***	0.787***		
	(0.046)	(0.068)		
Own Pre-GPA Low X Partner's Measure Low	-0.068	-0.241**	-0.028	-0.200*
	(0.063)	(0.108)	(0.070)	(0.117)
Own Pre-GPA Middle X Partner's Measure Low	0.009	-0.108**	-0.012	-0.131**
	(0.035)	(0.053)	(0.040)	(0.056)
Own Pre-GPA High X Partner's Measure Low	-0.054	-0.022	-0.025	-0.005
	(0.040)	(0.064)	(0.038)	(0.062)
Own Pre-GPA Low X Partner's Measure High	-0.012	0.071	0.008	0.085
	(0.074)	(0.089)	(0.089)	(0.107)
Own Pre-GPA Medium X Partner's Measure High	-0.049	0.016	-0.065*	0.002
	(0.031)	(0.052)	(0.036)	(0.056)
Own Pre-GPA High X Partner's Measure High	-0.048	-0.029	-0.032	-0.007
	(0.037)	(0.056)	(0.035)	(0.054)
Own Pre-GPA is Low			-0.544***	-0.523***
			(0.047)	(0.066)
Own Pre-GPA is High			0.406***	0.388***
			(0.032)	(0.044)
Technique	Standard	OOPG	Standard	OOPG
N	6,334	6072	6,334	6072
Dep. Var. Mean	0.003	-0.009	0.003	-0.009
Dep. Var. S.D.	1.000	0.989	1.000	0.989

Standard errors are clustered by course section. Additional controls include course-section fixed effects, indicators for race/ethnicity, gender, Educational Opportunity Program (EOP), the most common majors (proposed or declared) and undeclared major, and prior units accumulated. *** p<0.01, ** p<0.05,* p<0.1.

VIII. Conclusion

Peer effects are one of the most studied topics in the social sciences but estimation is complicated by several threats to validity. Among the most notorious problems are reflection, and self-selection into peer groups, but problems also result from repeated observations per peer group, exogenous vs outcome-on-outcome peer effects, identifying influential peers in the classroom, isolating student-to-student peer effects, limited variation in classroom mean ability of peers, and choice of functional form. We address these estimation problems with the combination of our experiment that randomly assigns students to one-to-one partnerships in a gateway science course and our new OOPG estimation technique. We first explore exogenous peer effects in which we use different measures of peer ability and find no evidence of peer effects. We find no evidence of exogenous positive peer effects on course scores, grades, dropouts, pass rates, taking additional courses in Chemistry, and several long-term outcomes.

Drilling down further in the search for evidence of peer effects, we next estimate outcome-on-outcome models which sometimes reveal peer effects when exogenous models do not. Outcome-on-outcome models do not rely on proxies for course-specific as exogenous models do.[31] These models can be viewed as upper-bound estimates of peer effects but have the advantage over pre-determined measures of ability because they directly measure the peer's ability for the outcome of interest. Furthermore, in this setting IV estimates do not detect peer effects because they simply rescale the exogenous peer effect estimates. Uncorrected or naïve outcome-on-outcome models reveal null or even negative estimates. We propose a new estimation strategy in which we include one observation for each peer group (OOPG) and find small, positive and statistically significant peer effects. We validate through Monte Carlo simulations that: (i) the OOPG estimation strategy corrects the negative bias due to the "mechanical problems" noted by Angrist (2014) from repeated observations within peer groups, (ii) the outcome-on-outcome model detects positive peer effects if and only if they exist, (iii) the OOPG technique recovers correct exogenous peer effect estimates and randomization check estimates. Thus, our experiment reveals small positive peer effects from the OOPG outcome-on-outcome model which would have been missed otherwise.

Furthermore, not only does the OOPG technique allow for estimation of upper-bound estimates using the outcome-on-outcome model but it is also not restricted to cases in which there is variation in urn sizes. The leave-me-out urn mean model relies on having variation in urn sizes because of collinearity between the LMOUM variable and the urn (section) fixed effect. In some educational settings including lab courses there is either no variation or very little variation in class sizes. The OOPG technique works in all cases and is thus valuable for future research on peer effects in a wider range of applications.

We also explore beyond the workhorse, linear-in-means model of estimating peer effects. Estimating various forms of non-linear effects and allowing for heterogeneous effects we mostly find null results. The main exception is that we find some evidence that low-ability peers "pull down" the student's performance in the course and no evidence that high-ability peers help or

[31] Although estimates from outcome-on-outcome models are less intuitive for potential policy interventions to sort students by ability than pre-determined ability measures, estimates from these models shed light on whether sorting by ability has an usefulness in influencing student outcomes for specific courses.

hurt the student's performance in the course. We do not find evidence of an inverted U-shape relationship for non-linear peer effects.

The mostly null results are not due to measurement issues, functional form, or lack of peer ability variation. We estimate models with several different outcome measures and different measures of peer ability. More importantly because we have one-to-one student matching we do not have problems with not knowing the inner workings or interactions in the classroom which could be related to the endogenous sorting of students into smaller groups. In the case, for example, where only 5 out of 30 students have a peer effect on the student it will be very difficult to detect a peer effect using the mean ability of the entire class.[32] We also have the ability to easily examine non-linear effects because we do not have to make choices between using points on a distribution (e.g. 75th percentile) or means within parts of the distribution to measure non-linear effects. Having one peer makes these choices one in the same. We also do not have problems due to a lack of variation in peer ability. We have the maximum possible variation with one peer instead of using the mean of numerous classmates. The null estimates are also not due to lack of peer interaction. The two students interact with each other closely the entire term in the Chemistry lab. Perhaps peer effects are not strong in higher education classrooms and students are mostly going to perform at the level of their own ability. In all estimates we find very strong and robust estimates of own ability on performance in the course. More research is needed on in-class interactions in higher education.

Overall, the findings in this setting of intensive interactions in a gateway STEM course that are randomly assigned and our comprehensive approach to estimation that includes outcome-on-outcome models provide evidence of, at most, small positive peer effects contrasting with the findings from many previous studies of peer effects in education. Estimates from the outcome-on-outcome model provide evidence of positive effects but even these upper bound estimates can rule out large peer effects. We can rule out peer effects that are larger than roughly 0.1 standard deviations on the course score in the class from a one standard deviation increase in peer ability. Finally, our new OOPG technique captures the small positive, peer effect which would have otherwise been missed with conventional models.

[32] Mean classroom ability essentially provides equal weight for each and every student in the classroom on estimating peer effects.

References

Angrist, Joshua D. 2014. "The perils of peer effects." *Labour Economics* 30: 98-108.

Beattie, Graham, and Fulya Ersoy. "Effects of Group Work on Attitudes, Trust, and Performance." (2022). AEA Conference paper https://www.aeaweb.org/conference/2023/program/paper/tB8z68BF

Babcock, Philip, Kelly Bedard, Stefanie Fischer, and John Hartman. 2020. "Coordination and contagion: Individual connections and peer mechanisms in a randomized field experiment." *Journal of Public Economics*, 185.

Booij, Adam S., Edwin Leuven, and Hessel Oosterbeek. 2017. "Ability peer effects in university: Evidence from a randomized experiment." *The Review of Economic Studies* 84.2: 547-578.

Carrell, Scott E., Richard L. Fullerton, and James E. West. 2009. "Does your cohort matter? Measuring peer effects in college achievement." *Journal of Labor Economics* 27.3: 439-464.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. 2013. "From natural variation to optimal policy? The importance of endogenous peer group formation." *Econometrica* 81.3: 855-882.

Elsner, Benjamin, Ingo E. Isphording, and Ulf Zölitz. 2021. "Achievement rank affects performance and major choices in college." *The Economic Journal* 131.640: 3182-3206.

Feld, Jan, and Ulf Zölitz. 2017. "Understanding peer effects: On the nature, estimation, and channels of peer effects." *Journal of Labor Economics* 35, no. 2: 387-428.

Garlick, Robert. 2018. "Academic peer effects with different group assignment policies: Residential tracking versus random assignment." *American Economic Journal: Applied Economics* 10, no. 3 (2018): 345-69.

Golsteyn, Bart HH, Arjan Non, and Ulf Zölitz. 2021. "The impact of peer personality on academic achievement." *Journal of Political Economy* 129, no. 4: 1052-1099.

Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo. 2009. "Peer effects in the workplace: Evidence from random groupings in professional golf tournaments." *American Economic Journal: Applied Economics* 1.4: 34-68.

Han, Li, and Tao Li. 2009. "The gender difference of peer influence in higher education." *Economics of Education review* 28.1: 129-134.

Kamei, Kenju, and John Ashworth. 2023. "Peer learning in teams and work performance: Evidence from a randomized field experiment." *Journal of Economic Behavior & Organization* 207: 413-432.

Kimbrough, Erik O., Andrew D. McGee, and Hitoshi Shigeoka. 2022. "How do peers impact learning? An experimental investigation of peer-to-peer teaching and ability tracking." *Journal of Human Resources* 57, no. 1: 304-339.

Manski, Charles F. 1993. "Identification of endogenous social effects: The reflection problem." *The review of economic studies* 60, no. 3: 531-542.

Manski, Charles F. 1995. *Identification problems in the social sciences*. Harvard University Press.

Oosterbeek, Hessel, and Reyn Van Ewijk. 2014. "Gender peer effects in university: Evidence from a randomized experiment." *Economics of Education Review* 38: 51-63.

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

Sacerdote, Bruce. 2001. "Peer effects with random assignment: Results for Dartmouth roommates." *The Quarterly Journal of Economics* 116.2: 681-704.

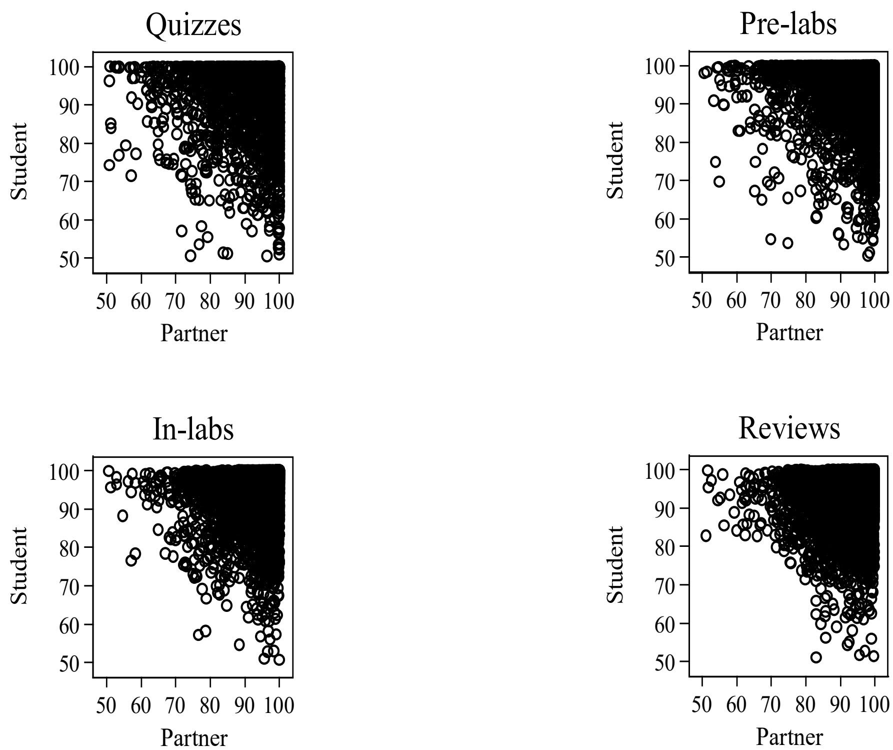
Sacerdote, Bruce. 2011. "Peer effects in education: How might they work, how big are they and how much do we know thus far?" In *Handbook of the Economics of Education*, vol. 3, pp. 249-277. Elsevier.

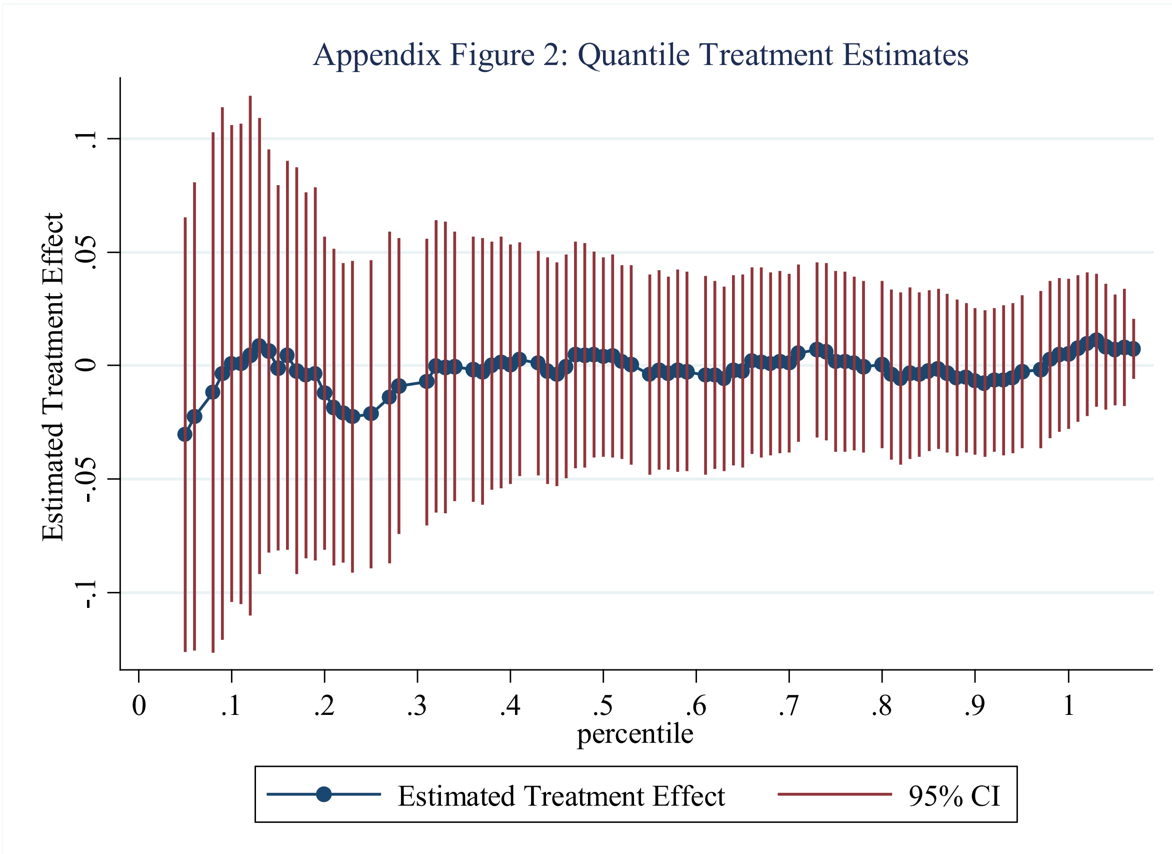
Sacerdote, Bruce. 2014. "Experimental and quasi-experimental analysis of peer effects: two steps forward?" *Annual Review Economics* 6.1: 253-272.

Zárate, Román Andrés. 2023. "Uncovering peer effects in social and academic skills." American Economic Journal: Applied Economics, 15(3): 35-79.

Zimmerman, David J. 2003. "Peer effects in academic outcomes: Evidence from a natural experiment." *Review of Economics and Statistics* 85(1): 9-23.

Appendix Figure 1: Student-Partner Scores on Different Types of Course Assignments




Appendix Figure 2: Quantile Treatment Estimates
Estimated Treatment Effect
-.1
-.05
0
.05
.1
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
percentile
Estimated Treatment Effect
95% CI

Appendix Table 1: Sorting Test by Student Characteristics

	Partner Baseline Characteristic										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Pre-GPA	Chem 1A Grade	Female	EOP	Freshman	Sophomore	Junior	Asian	Black	Hispanic	White
Individual Baseline Characteristc	0.004	0.000	0.001	-0.003	-0.006	0.005	-0.001	-0.006*	0.002	-0.005	0.002
	(0.003)	(0.005)	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)
N	6,334	5,930	6,470	6,466	6,466	6,466	6,466	6,470	6,470	6,470	6,470
Dep. Var. Mean	3.223	2.921	0.581	0.335	0.236	0.596	0.125	0.320	0.0201	0.273	0.288

Notes: Each cell represents a separate regression of the specified student characteristic on the partner's characteristic. All regressions include section fixed effects and the leave-me-out urn mean (LMOUM) which is the average of the student characteristic for the other students in the lab section. Standard errors are clustered as the section level. *** p<0.01, ** p<0.05,* p<0.1.

Appendix Table 2: Bounds Analysis for Lab Course Score adjusted for Drops (Partner's GPA)

	(1) Dropped Course	(2) Numeric Score	(3) bounds1	(4) bounds2	(5) bounds3	(6) bounds4
Peer Effect for Drops		N/A	0.030	0.100	0.200	0.200
Selection into Drops		N/A	0.000	0.000	0.000	-0.300
Own Pre-GPA	-0.041***	0.807***	0.805***	0.796***	0.783***	0.800***
	(0.008)	(0.038)	(0.035)	(0.035)	(0.036)	(0.035)
Partner's Pre-GPA	-0.008	-0.009	-0.010	-0.008	-0.004	-0.001
	(0.007)	(0.024)	(0.023)	(0.023)	(0.023)	(0.023)
Observations	6,684	6,334	6,684	6,684	6,684	6,684

Notes: Own Pre-GPA and partner's Pre-GPA are measured prior to term. Standard errors are clustered by course section. Additional controls include course-section fixed effects, indicators for race/ethnicity, gender, Educational Opportunity Program (EOP), the most common majors (proposed or declared) and undeclared major, and prior units accumulated. Dropped course observations are imputed using specified assumptions for peer effect and selection magnitudes. *** p<0.01, ** p<0.05,* p<0.1.

Appendix Table 3: Bounds Analysis for Lab Course Score adjusted for Drops (Partner's GPA)

	(1)	(2)	(3)	(4)	(5)	(6)
	Dropped Course	Numeric Score	bounds1	bounds2	bounds3	bounds4
Peer Effect for Drops		N/A	0.030	0.100	0.200	0.200
Selection into Drops		N/A	0.000	0.000	0.000	-0.300
Own Chem 1A Grade (4-Pt. Scale)	-0.016***	0.262***	0.261***	0.258***	0.253***	0.257***
	(0.004)	(0.018)	(0.017)	(0.017)	(0.017)	(0.017)
Partner's Chem 1A Grade (4-Pt. Scale	-0.005*	-0.018	-0.017	-0.015	-0.011	-0.010
	(0.003)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Observations	6,203	5,913	6,203	6,203	6,203	6,203

Notes: Own Chem 1A Grade and partner's Chem 1A Grade are measured on 4-point scale and prior to term. Standard errors are clustered by course section. Additional controls include course-section fixed effects, indicators for race/ethnicity, gender, Educational Opportunity Program (EOP), the most common majors (proposed or declared) and undeclared major, and prior units accumulated. Dropped course observations are imputed using specified assumptions for peer effect and selection magnitudes. *** p<0.01, ** p<0.05,* p<0.1.

Appendix Table 4: Peer Effect Regressions for Course Grade Components

	(1)	(2)	(3)	(4)
	Quizzes	Pre-Labs	In-Labs	Reviews
Own Pre-GPA	0.545***	0.632***	0.724***	0.678***
	(0.042)	(0.035)	(0.039)	(0.036)
Partner's Pre-GPA	-0.016	0.002	-0.014	-0.018
	(0.035)	(0.027)	(0.025)	(0.025)
N	3,581	6,334	6,334	6,334
Dep. Var. Mean	0.007	0.001	0.005	0.008
Dep. Var. S.D.	0.993	1.003	0.996	0.998

Notes: Own Pre-GPA and partner's Pre-GPA are measured prior to term. Standard errors are clustered by course section. Additional controls include course-section fixed effects, indicators for race/ethnicity, gender, Educational Opportunity Program (EOP), the most common majors (proposed or declared) and undeclared major, and prior units accumulated. *** p<0.01, ** p<0.05,* p<0.1.