Estimating Spillovers in the Classroom with Panel Data*

Peter Arcidiacono, Duke University

Gigi Foster, University of South Australia

Natalie Goodpaster, Duke University

Josh Kinsler, Duke University

December 30, 2004

Working Paper

Abstract

We develop a new strategy for estimating peer effects when there are multiple observations per person and the peer group varies across observations. This technique allows us to explicitly account for student fixed effects and to circumvent student selection into classes based on unobservables. Iteratively estimated student fixed effects are used to formulate the ability level of the peer group. Furthermore, we are able to quantify the impact of measurement error in peer ability on peer effect estimates using both our model and traditional approaches. Monte Carlo evidence shows that our algorithm performs well even given a short panel. We find statistically significant peer effects, particularly in courses of a collaborative nature. We are able to compare our methods with traditional approaches and quantify the biases associated with both selection and measurement error.

1 Introduction

Most educational institutions are highly interested in identifying and funding the inputs into educational production. Over the past ten to fifteen years, there has been increasing interest

*We wish to thank Joe Altonji, Pat Bayer, participants at the 2004 Labour Econometrics Workshop held at the University of Auckland, Duke University Colleagues, seminar participants at Yale University, and participants at the NBER's Higher Education Workshop for helpful comments, as well as the University of Maryland administrators who have provided us with the data used in this paper. Special thanks go to Chris Giordano and Bill Spann of the Maryland Office of Institutional Research and Planning.

in the possibility that peers may be one such input, particularly in the college classroom environment.[1] This basic question—do peers affect a student's achievement—underlies debates regarding the impact of affirmative action, school quality, and public school improvement initiatives such as school vouchers, and is central to more immediate concerns such as how best to group schoolchildren to maximize learning.

There are at least two barriers that must be overcome when estimating peer effects on student achievement. First, the underlying abilities of students are highly likely to be measured with error. Second, peer groups are in general not randomly assigned, and student selection into peer groups may be based on characteristics that affect performance but are unobservable to the researcher. The root cause of both of these problems is econometricians' inability to accurately measure each student's bundle of performance-relevant characteristics.[2]

When individuals choose their peer groups, high ability students may sort into groups with other high ability students. Positive estimated peer effects may then result not from true peer influence, but rather from not measuring the ability of the individual accurately. On the other hand, having noisy measures of true peer ability may bias the estimated peer effect downward, leaving the final direction of bias uncertain. The use of test scores, such as the SAT or MCAT, as measures of student ability is prevalent in much of the peer effects literature. These one-time events are typically inaccurate measures of prior student ability, and moreover, measuring peer ability using only test scores and other background ability measures (like high school GPA) ignores the probable strong contribution of contemporaneous peer effects.[3]

[1] While the concept of peer influence was around much earlier, only with the advent of advanced computing and new econometric methodologies have researchers been able to explore peer effects in a sophisticated and comprehensive manner.

[2] To facilitate discussion, we refer to this bundle as "ability" throughout much of the paper. However, it can be thought of as the set of everything that a student brings to the classroom that influences either his own or his fellow students' learning. One serious limitation plaguing empirical efforts toward identifying peer effects is that we do not know what this set includes. However, there is reason to believe that it could include effort, motivation, mental state, curiosity, observable and unobservable academic aptitude, race, gender, age, personality, self-esteem, respect for others, and much more. This state of affairs supports the use of some form of fixed-effects modeling, wherein all fixed person-specific performance-relevant attributes can be accounted for without having to be explicitly enumerated.

[3] In the language of Manski (1995), researchers have felt the need to distinguish between endogenous and contextual effects, as they have differing policy implications. The fixed effects method can be thought of as accommodating both types of effects. However, the impact of student-specific attributes that are also specific

Researchers have undertaken a variety of estimation strategies to try overcome these barriers, paying particular attention to the need to control for the endogeneity of the peer group due to student selection. One set of papers uses proxy variables to break the link between unobserved and peer ability.[4] Another set of papers relies on some form of random assignment.[5] Finally, researchers have tried to circumvent the endogeneity problem with instrumental variables.[6]

Each of these approaches is innovative and has significantly advanced the discipline's thinking about peer effects. However, there are significant empirical problems that linger, due both to persistent measurement error in ability and to the somewhat narrow focus on mimicking randomness. First, most prior papers rely on noisy measures of underlying peer ability, which unavoidably creates downward pressure on estimated effects, as previously discussed. Second, relying on random assignment or attempting to mimic it via a technique such as instrumental variables does not accommodate the possibility that peer influence may be much stronger among individuals who choose their peers than among those who are randomly assigned.

Our estimation strategy allows us both to improve significantly the measurement of peer ability, and to estimate accurately how peer ability affects the individual's outcome of interest, without reliance on random assignment. The only requirement of the method is variation in both the outcome measure and the composition of the peer group within students. Importantly, additional student-level information on demographics, prior achievement, or indeed anything except performance and peer group assignment is completely unnecessary for the method to be implemented. The method iteratively estimates student fixed effects, and rather than using test scores as a noisy measure of peer ability, we measure peer ability using the estimated individual fixed effects of the student's peers.

Monte Carlo results suggest that the algorithm performs well even when the number of outcome/peer group combinations per student is small. The estimator yields downwardly-

---

to a particular peer environment cannot be separately identified. We discuss this in more detail below.

[4]Among many others, Arcidiacono & Nicholson (2004) and Hanushek, Kain, Markman & Rivkin (2003) attempt to remove the correlation between the unobservables and peer ability through controlling for school fixed effects, with the latter removing individual effects through first differencing.

[5]Sacerdote (2001), Zimmerman (2003), and Winston & Zimmerman (2003) rely on random assignment of roommates. Hoxby (2001) uses the variation in gender and race composition of the classroom over time in her identification strategy, assuming that these variations are not associated with individuals' unobserved ability.

[6]See, for example, Evans, Oates & Schwab (1992).

biased estimates of the peer effect. The downward bias is more pronounced the larger the transitory component of performance, and the lower the variation in the ability of the peer group within students. The estimator actually performs better when there is some selection into the peer group, as random assignment yields small variation in peer ability.

We estimate the model using student-level data from the University of Maryland. Six semesters of transcript data are available covering the semesters from the spring of 1999 to the fall of 2001. We observe grades for every class the individual took over the course of this period as long as the individual lived on campus during any one of the six semesters. We estimate the model separately by each of four distinct course types, where we conceive of grades in different course types as distinct outcomes. We find significant peer effects which vary by course type. A one-point increase in peer ability yields similar returns to between a 0.04 and 0.15 increase in own ability, depending upon the course type. The highest spillovers are found in courses in the social sciences, with spillovers the smallest in mathematics and natural sciences.

The rest of the paper proceeds as follows. Section 2 presents the model and estimation method. In section 3, we provide evidence from Monte Carlo simulations that our estimator performs well even when only a short panel is available, and when students choose their peer group. Section 4 describes the University of Maryland data, and Section 5 presents the results. Section 6 illustrates the biases associated with using only observable ability measures to estimate peer effects, and addresses a few other points regarding the chosen estimation procedure. Section 7 concludes.

2 Model and Estimation

2.1 Background

In this section we present the model and estimation strategy. We being by writing down an outcome equation along the lines of Manski (1995) except now allowing for the peer group to change over time.[7] Throughout, we work with a linear model where peer effects operate

[7]Note that variance in the peer group across time is not specifically required by our model. In this exposition, we assume that across-time variation is the source of the variation in the peer group that the model does require. In our actual data, variance in the peer group results from multiple observations per student both across time

through averages of the relevant characteristics of the reference group. We further assume that there are no accumulation effects: neither outcomes nor peer groups today have a direct influence on outcomes tomorrow.[8] The most general case under these assumptions has an individual i's outcome at time t depending upon their own observed and unobserved characteristics, X_i and u_i, the average observed and unobserved characteristics of the peer group, $\overline{X}_{\sim it}$ and $\overline{u}_{\sim it}$, the average outcomes of the peer group, $\overline{Y}_{\sim it}$, environment-specific variables that affect the whole group, Z_t, and a transitory error ϵ_{it}:

$$Y_{it} = X_i\alpha_1 + u_{it}\alpha_2 + \overline{X}_{\sim it}\gamma_1 + \overline{u}_{\sim it}\gamma_2 + \overline{Y}_{\sim it}\gamma_3 + Z_t\delta_1 + \epsilon_{it} \quad (1)$$

In the language of Manski (1995), endogenous effects would operate through $Y_{\sim it}$, exogenous effects through $\overline{X}_{\sim it}$ and $\overline{u}_{\sim it}$, and correlated effects through Z_t. The u_i's can either be viewed as unobserved characteristics or as measurement error associated with only having noisy measures of an individual's characteristics in X_i.

Even if we assume that there are no endogenous or correlated effects (γ_3 and δ_1 both equal zero), estimating equation (1) is still problematic when peer groups are chosen by individuals. In particular, there may be correlation between u_i and $\overline{X}_{\sim it}$. Also, we will not be able to capture the peer influences from $\overline{u}_{\sim it}$. While random assignment can remove the correlation between u_i and $\overline{X}_{\sim it}$, the measurement error problem remains.

Throughout the rest of this section we will assume that there are no endogenous effects, or in other words that γ_3 is zero. Endogenous effects through the outcome measure itself do not seem plausible in the context of student achievement. It is not one's grade itself that may matter to others' performance, but the effort and abilities that are ultimately reflected in one's grade.[9] In our context, these are part of the X's and u's, and are assumed to be constant for a given individual. We show how to solve both the selection problem and the measurement error problem using panel data to capture the effects of both the X's and the u's when there are no correlated effects. Then, we extend the model to accommodate correlated effects as well.

and within a given semester.

[8]We intend to relax these assumptions in future work.

[9]Endogenous effects through the outcome measure make more sense in the context of smoking, where it is the smoking itself that may matter, as opposed to the root causes of the choice to smoke.

2.2 Baseline Method

Assuming that there are no endogenous or correlated effects, equation (1) reduces to:

$$Y_{it} = X_i\alpha_1 + u_{it}\alpha_2 + \overline{X}_{\sim it}\gamma_1 + \overline{u}_{\sim it}\gamma_2 + \epsilon_{it} \tag{2}$$

We next make one additional assumption: the relevance of peer characteristics is proportional to that of own characteristics:

$$\begin{aligned}\gamma_1 &= \gamma\alpha_1 \\ \gamma_2 &= \gamma\alpha_2\end{aligned}$$

This implies, for example, that if two dimensions of an individual's ability are equally important in their effect on Y_{it}, then those two dimensions of peer ability will also be equally important in determining Y_{it}. This is the same assumption used in Altonji, Elder & Taber (2004) in a static model.

Now define α_i as:

$$\alpha_i = X_i\alpha_1 + u_i\alpha_2$$

We can then rewrite equation (2) as:

$$Y_{it} = \alpha_i + \gamma\bar{\alpha}_{\sim it} + \epsilon_{it} \tag{3}$$

An individual's outcome is then a function of the individual's fixed effect plus the average fixed effect of the other members of the peer group. Using fixed effects in this way allows us to abstract from many other covariates that may affect student outcomes. All of the heterogeneity in fixed student characteristics that might affect student outcomes, such as age, sex, socioeconomic status, or race is captured with this one measure. Less tangible attributes that are also grade-relevant, such as effort level and motivation, are included as well.

While it is theoretically possible to estimate the set of fixed effects in one step, it is not computationally feasible. Instead, we pursue an iterative procedure. The algorithm begins by making an initial guess as to the values of the unobserved abilities. In practice, we use the average value of the outcome measure for each student as our initial guess of each α, although the algorithm is not sensitive to the starting values. Utilizing these initial values, we can calculate the average abilities of all other students in the peer group at time t. Our first estimation of γ given the α's can then be obtained by ordinary least squares. In particular, denote by α_i^0 the initial guess of student i's α. We estimate γ^0 by running OLS on the following:

$$Y_{it} - \alpha_i^0 = \gamma^0 \bar{\alpha}^0_{\sim it} + \epsilon^0_{it}$$

Once we estimate the initial γ^0, we can update our conjecture of the individual ability level for each student i by subtracting the fitted value of the previous regression from the student's grades and averaging:

$$\alpha_i^1 = \sum_{t=1}^{T} \frac{Y_{it} - \gamma^0 \bar{\alpha}^0_{\sim it}}{T}$$

Using these new abilities to calculate new average peer abilities for each student in each time period, we continue to iterate on this process by estimating a γ^n for the n^{th} iteration with:

$$Y_{it} - \alpha_i^n = \gamma^n \bar{\alpha}^n_{\sim it} + \epsilon^n_{it}$$

A student's new ability is re-calculated each iteration as

$$\alpha_i^{n+1} = \sum_{t=1}^{T} \frac{Y_{it} - \gamma^n \bar{\alpha}^n_{\sim it}}{T}$$

As long as the γ's remain less than one, the equilibrium is unique and the iterative procedure converges to the one-step maximum likelihood estimates of abilities and peer effects.

2.3 Adding Correlated Effects

If each individual peer group is exposed to a different environment, it is impossible to separate the correlated effects from the exogenous effects without further parameterizations. A restriction that is easily imposed for correlated effects in our context is that teachers have the same effect on their students across classes. In this case we could place teacher fixed effects into the regression above. Equation (3) then becomes:

$$Y_{it} = \alpha_i + \gamma \bar{\alpha}_{\sim it} + \delta_t + \epsilon_{it} \tag{4}$$

where the δ_t's are the teacher fixed effects.

Instead of estimating the teacher fixed effects themselves, we demean grades and control for the average of the all the α_i's in the course, $\overline{\alpha}_t$. After some algebra, we can rewrite (4) as:

$$Y_{it} - \overline{Y}_t = \alpha_i + \gamma \bar{\alpha}_{\sim it} + \delta \overline{\alpha}_t + \epsilon_{it} \tag{5}$$

where mechanically $\delta = -1 - \gamma$.

3 Monte Carlo Simulations

To investigate the properties of our estimator, we simulate the model under different assumptions regarding selection into classes, the number of observations per individual, and the noise present in our outcome measure. In each case, the model is simulated using 3,000 students assigned to classes of 15 students apiece. We simulate the model under various states of the world constructed by varying three dimensions of the problem:

1. *Observations per student-* The number of outcomes observed per student varies across simulations between 4, 10, and 20. 4 is the maximum number of observations a researcher may have when analyzing grade school or high school test score data, and 20 observations is more appropriate when analyzing grades achieved in university-level courses. More observations per student implies more accurate measurement of the α's.

2. *Selection into classes-* To show that our estimator solves the selection problem, we simulate the model under alternative assignment rules. Under random assignment, the average standard deviation of the α's within a peer group equals the standard deviation of α in the population. We also simulate the model with selection such that the average standard deviation of the α's within a peer group is 75%, 50%, and 25% of the population standard deviation. To put these levels of selection into perspective, the Maryland transcript data has average standard deviations of ability across peer groups between 78% and 85% of the standard deviation in the population, implying that even our low-selection simulations have much more selection than what we see in the data.

3. *Transitory component-* The noisier the outcome measure, the noisier the estimates of the α's will be. The distribution of the α's is set at N(0,1). The ϵ's are distributed with mean zero and standard deviations ranging from 0.35 to 2.

More details regarding how the Monte Carlo simulation were conducted, and in particular how students were assigned to classes, can be found in the appendix.

To first verify that the estimator does not give positive peer effects when no peer effects are present, Table 1 shows simulation results when the true peer effect, γ, is zero. In all cases, regardless of the number of observations per student, the level of selection, and the noisiness of the outcome measure, γ is precisely estimated around zero. There is no evidence of bias when the true peer effect is zero.

Table 2 investigates the case when the true value of γ is 0.15. The first panel shows the results for random assignment. Moving across the sets of rows within a selection level shows the results as the number of observations per student increases, while moving down rows within a selection level shows what happens as the outcome measure becomes less noisy. For random assignment, we see downwardly biased estimates when there are few observations per student and the outcome measure is particularly noisy. Typically when estimating a regression coefficient, the amount of variance in the dependent variable that goes unexplained does not affect the magnitudes of the coefficients. In our procedure, however, we are estimating not only the coefficients, but the regressors as well. In the worst case scenario— four observations per student with the standard deviation of the noise twice that of the fixed effect— the peer effect is estimated to be half that of the true value. The results improve substantially as the standard deviation of the noise is decreased and as the number of observations per student increases.

In the second panel, the data generating process yields standard deviations of ability in a peer group that are 75% of the standard deviation of ability in the population. Here the estimator performs even better than under random assignment. While this may seem counterintuitive at first, there is much more variation in peer ability when there is some selection. With this level of selection, the average variance in peer ability within students is .28 in the case of four observations per student, as compared to .08 for random assignment; a similar pattern obtains for the cases of ten and twenty observations per student. As with random assignment and throughout the rest of the Monte Carlo simulations, decreasing the standard deviation of the noise and increasing the number of observations per student improves the estimates.

The final two panels of Table 2 illustrate the model's performance when selection is extremely high: standard deviations of ability in a peer group are 50% and 25%, respectively, of the standard deviation of ability in the population. Again we see no evidence that the estimator is upwardly biased. Rather, as with the first two panels, downward-biased estimates are found when the noise term is large and the number of observations per student is small.

Finally, in Table 3 we repeat the analysis in Table 2 except that we now allow for correlated effects. Each of the classes are grouped such that there are 5 classes in a group, mimicking the relationship between sections and courses. Within each group, a common shock is added to the outcome. The method performs just as well as in Table 2, with the exact same patterns emerging. In Table 3, this common shock is not correlated with the abilities of the classes. However, in simulations not reported here, we found the same results when the common shock was correlated with the abilities of the group.

4 Data

The administrative data set used in this paper covers all undergraduates observed residing in University of Maryland on-campus housing during any of the following six academic semesters: spring 1999, fall 1999, spring 2000, fall 2000, spring 2001, or fall 2001. The data set includes students living off-campus in a given semester as long as they were observed living on campus during at least one of the six semesters. 90% of University of Maryland entering freshmen live on campus in their first semester,[10] so the data set includes at least 90% of the University of Maryland undergraduate population who began study sometime in the six-semester period. There is a less complete representation for upperclassmen, some of whom entered before our observation period and may not have lived on campus during the period. However, given that the identification of our section-level peer effect comes from large, multi-section courses in which lowerclassmen predominate, we feel that the sample is adequate for our purposes.[11]

The identification of the appropriate peer group for a specific course is facilitated by the use of sections for large university courses. Students often meet at least once a week with a portion of the entire class, called a section, where greater communication and interaction is expected. Our peer measure in this setting will capture how the ability of students in a common section impact each other's eventual grade in a class.

To generate the student-section-level sample used to estimate our models, two major restrictions were placed on the data set: students had to have valid A through F grade information for the given section, and they could not be the only student observed in the section

[10] This number is taken from publicly-available statistics posted on the University's web page.

[11] We did test to see if classes that were under-represented had lower estimated peer effects than those with a complete representation and found no meaningful differences.

that semester.[12] These restrictions yielded a sample of 300,640 student-section observations, representing 18,554 individual students. Sample sizes are provided in Table 4.

Students at the University of Maryland enroll in a wide range of courses. Performance in each of these courses will differ according to the particular student's strengths and weaknesses. Therefore, instead of encapsulating all the attributes of a student into one ability measure, we allow students to have separate ability measures for each course type in which they are enrolled. All courses are classified as belonging to one of the following course types: Humanities, Math and Science, Social Science, or Vocational. Each student will have an independent ability measure for each type of course, conditional on enrolling in at least one class within that course type.[13]

Under the assumption of multiple course type-specific ability measures, and given the structure of our data, we use the following model of a student's grade in a specific type of course, as previously discussed:

$$Y_{it} - \overline{Y}_t = \alpha_i + \gamma\bar{\alpha}_{\sim it} + \delta\tilde{\alpha}_t + \epsilon_{it} \tag{6}$$

All grades are demeaned at the course level, with corresponding controls included for the overall ability level in the course via the $\tilde{\alpha}$ term. This model is run through our algorithm separately for each type of course, yielding four section peer effect and class effect estimates, as well as multiple ability measures for each student who was observed in more than one course type.

The inclusion of the class ability term is similar to removing teacher fixed effects. Because grades are assigned at the course level, there is a relationship between students who share a course but are not in the same section that cannot be captured by the section peer effect, γ. We might expect, for example, that if the course is graded on a curve and the entire class

[12]Numeric grade equivalents were assigned as follows: A = 4; B = 3; C = 2; D = 1; and F = 0. Students who withdrew from a course, audited, or received a non-letter grade (such as Pass) were excluded from the sample due to concerns that they might not have been present during sections and classes. If two separate grades were recorded for the student for a given section, the highest grade was used.

[13]The four course types used in practice were based on the more detailed subgroups presented in Tables 13 through 15. As listed in those tables, the humanities course type contains the subgroups "ah" and "app;" the math and science course type contains the subgroups "cmps," "ls," and "se;" the social science course type contains the subgroups "bsos" and "sb;" and the vocational education course type contains the subgroups "anr," "ed," "j," "hhp," "is," "pa," and "ug."

is extremely able, a mediocre student's grade may suffer. Similarly, the teacher of a course may respond to higher average class ability by teaching in a way that enhances learning and therefore raises grades for at least some portion of the class. By demeaning at the course level and including the $\delta\tilde{\alpha}_t$ term, we can account for these potential class-level peer effects on a student's grade while making the outcome measure comparable across classes.

Descriptive statistics by course type are given in Table 5. Science and math courses tend to draw both higher math ability students and higher verbal ability students. Table 5 also provides information on the level of selection in each course type. The numbers in italics give the average standard deviation of the variable across sections divided by the standard deviation of the variable in the population for that course type. The smaller these numbers, the tighter the distribution of ability within sections relative to the unsorted distribution, and therefore, the more ability-based selection into sections is evident. Not surprisingly, science and math classes have the most selection on math ability. However, in all cases, the average standard deviation of ability within a section divided by the standard deviation of ability in the population is greater than 0.75.

5 Estimates of Classroom Spillovers

Table 6 shows the results from estimating Equation 2 for each of the four type of courses. The results indicate positive and significant section peer effects in all cases. The weakest estimated section peer effect occurs in science and math courses, and the strongest in social science courses. This pattern may reflect the amount of collaborative work required in each course type. In addition, if the environment in course types such as social sciences is more conducive to communicating knowledge across peers, or if peer discussion is more conducive to learning the type of material in these courses than in math and science, larger peer effects may result. Note that the estimated coefficients on course peer ability are mechanically equal to $-1-\gamma$.

The magnitudes of the section-level peer effects show that a one point increase in peer ability is equivalent to a maximum .1478 increase in individual ability in social sciences classes and a minimum .0397 increase in individual ability in math and science classes. Scaling by standard deviation increases shows smaller relative peer effects, as the standard deviation of average peer ability is smaller than the standard deviation of ability itself. Table 7 displays the relevant standard deviations, along with the effect of a one standard deviation increase in peer ability divided by the effect of a one standard deviation increase in individual ability. The gap evident in the raw marginal effects between math and science and the other course types is somewhat mitigated. A one standard deviation increase in peer ability is equivalent, in terms of effects on grades, to a maximum of 6.5% of the effect of a one standard deviation increase in individual ability in the social sciences, or a minimum of 2.5% of the effect of a one standard deviation increase in individual ability in math and science classes. As we will show in more detail in a moment, this narrowing of the course type gap once we consider the relative contribution of changes in standard deviations is because math and science courses have much more selection into sections than other course types.

An important by-product of our procedure is course type-specific student ability estimates. Because the vast majority of students are observed in courses of multiple types during their tenure at Maryland, we obtain multiple estimates of ability for each student. Each estimated ability measure is specific to courses of a particular type, by virtue of the estimation procedure, and as such each can be assumed to reflect a skill set that is particularly in demand in that course type. Calculating the correlations between estimated ability levels illuminates to what extent good performance in each of the four course types is driven by similar student attributes as performance in the other types, and therefore provides an empirical index of the academic similarity of course types.

Table 8 shows the correlation coefficients among estimated ability levels across the four course types. These correlations are created using estimated abilities from students observed in all course types.[14] The highest correlation coefficient is for social sciences and math and science at .6842, followed closely by humanities and social sciences at .6826. This is as we would expect, since both social and hard sciences require aptitude in logical reasoning and numeracy. Similarly, a student who has high ability in humanities most likely has a high skill level in reading and writing papers, which is also required for success in the social sciences. The ability that determines performance in vocational courses and that determining performance in math and science courses exhibit the smallest correlation coefficient, at .5475.

[14]Correlations among estimated ability levels were also calculated for all students who were observed in each pair of course types. Similar coefficients resulted.

6 Comparing the Method to Selection on Observables

Next, we compare peer effect estimates from our method to those derived from other conventional methods. We explore to what extent the fixed effects from our iterative algorithm are predictable using the observed proxies for ability measures consistently used in related literature, and to what extent the peer effects estimated using the two methods differ. In order to facilitate this comparison, we use SAT scores and high school performance as regressors for a conglomerate observable measure of ability. For each course type, we regress our estimated student fixed effects on the student's SAT math score, SAT verbal score, and high school grade point average. We find that the R^2 for these regressions ranges from .14 to .33 (depending on the course type). We then include a set of grade-related controls such as gender, race, and participation in sports and honors programs as covariates in the ability regression. Even with these additional covariates, we can still only explain a maximum of 34% of the variation in estimated student ability.

Full results from regressing the student fixed effects on observable characteristics of the student are presented in Table 9. The statistical significance and magnitude of the coefficients on SAT math and SAT verbal scores vary across the four different course types. This seems quite intuitive. For example, SAT math scores are insignificant when explaining ability in humanities courses, but they are a better proxy for math and science ability, judging by the large and statistically significant coefficients.

Since SAT math, SAT verbal, and high school GPA scores only explain a small portion of our estimated ability measures, we suggest that these observable performance indicators, which are conventionally used to formulate peers' ability in peer effects research, are overall poor proxies for true student ability. The consequences of this measurement error in ability on peer effects estimates may be significant. As discussed in the introduction, we would expect measurement error in student ability to yield a downward bias on estimated peer effects. However, when measurement error in ability is combined with student selection into peer groups, the end result could be an upward bias on peer effect estimates given the two biases oppose each other in estimation.

To compare empirically the biases induced by measurement error and student self-selection on conventional peer effect estimates, we take the predicted values from the regressions run to produce Table 9 and call these our " predicted," course-type specific ability measures, in the sense that they represent the portion of our model's estimated student-level ability that is explained by observable student characteristics. We use these predicted values as multi-dimensional proxies for the observed ability measures typically used when estimating peer effects. We can now re-estimate Equation (5) via OLS using either the predicted fixed effects from Table 9, the true estimated fixed effects, or a combination of the two measures as the underlying ability measurement. Estimation of the model using only the predicted abilities will suffer from both selection and measurement error, as students may select based on unobserved portions of ability. Estimating the model using our estimated fixed effects as own ability, and the predicted values from Table 9 as peer ability, the model suffers only from measurement error in peer ability. Selection issues are eliminated by using the "true" own ability measure, thereby isolating the impact of measurement error.

The results of estimating Equation 2 using different proxies for own and peer ability are given in Table 10. For each course type, three rows of results are shows. Row (1) for each course type naturally displays identical results to those from our iterative procedure (shown in Table 6). Results in row (2) for each course type demonstrate the bias due only to measurement error in peer ability, since the predicted values from Table 9 are used to create the peer abilities for each student, but the true ability estimates are used as the individual ability measures. This leaves no possibility that students are sorting into peer groups based on an unobserved portion of grade-relevant ability. Finally, the specification in row (3) injects selection bias into the section peer effect estimate by using the predicted abilities from Table 9 to calculate both own and peer ability, thereby intentionally simulating the case where both are measured with error.

When we control for individual ability in the models but use noisy "observable" ability measures for peer ability (row (2) for each course type), the section peer effect is underestimated, as expected when measurement error in ability is present. The estimated peer effects are uniformly smaller than in row (1), and lose significance in two of the four course types. The results in this row elicit concern about researchers' attempts to eliminate student selection into peer groups while relying on noisy measures of peer ability. Note that these results are conservative, in the sense that they assume a precise measure of own ability, which is also often lacking.

The most dramatic results, however, are in the third row for each course type. Here we use the predicted values from the regressions reported in Table 9 as proxies for both own and peers' ability. This simulates the estimation of peer effects when peer ability is still poorly measured, and in addition, students can select into peer groups based on the unobserved portion of their ability. In all course types, peer effects estimated in this manner not only exceed the exclusively downwardly-biased estimates in row (2), but also exceed the actual effects shown in row (1). Bearing in mind that in our data, student selection into classes is relatively weak, these results provide evidence that even minimal student selection can result in dramatically overestimated peer effects. Moreover, this selection cannot be adequately controlled for using standard observable characteristics.

We would logically expect to find a higher peer effect estimate under selection in those course types where selection by ability is most important. Table 10 provides indirect evidence that selection into peer groups is more important for math and science and social science sections, as the peer effect estimates in row (3) for these course types are the highest amongst all four course types. Table 11 presents additional evidence that this is the case, especially for math and science courses. The first column shows the average standard deviation of our estimated student ability measures within a section in each course type, while the second column shows the corresponding average standard deviation of ability across the entire population of students observed in courses of that type. The more selection into course types, the smaller the first column will be relative to the second column. The ratio of these two numbers is presented in the third column. The third column shows that the standard deviation of ability within a section is between 79% and 90% of the standard deviation in the population, with the most selection occurring in math and science courses.[15]

Finally, we examine the relative importance of the observable and unobservable portions of student ability in determining the level of student selection into sections. Using the regression partition from column (2) of Table 9 for each course type, we divide our fixed effects into

[15]Note, as observed previously, that this level of selection is small compared to the selection imposed in our Monte Carlo experiments. It is likely that the population of students observed at the university has already been selected, and is therefore more homogeneous than would be found in the entire population of college-age individuals. Grade school children may therefore be likely to experience greater levels of selection, which may cause the error from selection bias to dominate measurement error even more dramatically in a primary or even secondary education context.

"predicted" and "residual" components. We then take each component of ability in turn, and calculate the selection ratio as calculated in Table 11, but for each component of ability separately. This yields two selection ratios for each course type. We can then compare the ratio calculated from the observed part of the fixed effect to the ratio from the unobserved part of the fixed effect.

Table 12 shows the results from this procedure. The first through third columns of the table show the standard deviations and selection ratio from the predicted part of the fixed effect, while the fourth through sixth columns show the same for the residual part of the fixed effect estimates. Each row of the table represents a different course type. In general, the selection on observables is of the same order of magnitude as the selection on unobservables.[16] Notice, however, that the relative selection ratios do vary across course types. In math and science and vocational courses, the selection ratio is higher for the predicted part of ability than for the residual part of ability. This implies that students in humanities and social sciences courses are relatively more likely to select into sections based on unobservable characteristics than based on observable characteristics, while students in math and science and vocational courses are relatively more likely to select based on observables. This may be partly explained by differences in personality attributes common among students enrolled in the different course types, and/or by differences in the type of work required in the different course types. In particular, students taking math and science courses may be more independent, and/or less motivated to be in sections with students who have similar characteristics unobserved to the econometrician. However, in courses that are more conducive to group work and feedback from peers, it may be more important for a student to select into sections full of students with whom they match on other levels than academic ability.

7 Conclusion

We present a new iterative method for estimating educational peer effects that controls for an individual's own ability and accurately measures the ability of the peer group. Conditional

[16]The results in Table 12 are consistent with recent estimation methods by Altonji et al. (2004) that rely on the assumption that the amount of selection on observable characteristics is the same as the amount of selection on unobservable characteristics.

on our definition of the peer group, selection into the peer group is no longer an obstacle, and indeed may assist in accurate measurement of the peer effect. In order to use this method, while observable student-level controls are unnecessary, there must be multiple outcomes and multiple peer groups observed for each student. We model outcomes as dependent upon individual fixed effects and of the average fixed effects of members of the peer group. Monte Carlo results show that the model performs quite well even when the number of observations per student is small. In our tests, the method never yielded upwardly-biased estimates of the peer effect.

We test the model on transcript data from the University of Maryland. Small but significant peer effects are found, with evidence of heterogeneity by course type. Social science courses show the largest peer effects, whereas grades in math and science courses rely least on peer ability and most heavily on a student's own fixed effect.

Our estimates are higher than corresponding estimates where selection is eliminated but peer ability is measured with error. However, we provide evidence that estimates derived from methods that do not control for student selection may dramatically overestimate peer effects, even in the presence of measurement error in peer ability.

There are many avenues to be explored in future research. First, we have assumed that all individuals in the classroom influence each other equally. Allowing the peer effect to vary across gender or race lines may yield even higher peer effect estimates as we move to better measures of peer group composition. Similarly, we can examine whether high or low ability individuals are affected most by the abilities of others, and whether students with particularly high or particularly low ability influence their peers more.

Second, future research will relax the assumption that an individual's ability to help others is proportional to an individual's own ability to perform well. The individual who asks the clarifying question may be more useful to others than a smarter individual who remains quiet. It is possible to extend the model such that spillover ability is treated differently from the ability to perform well for oneself.

Third, peer effects here are purely transitory. Future work will relax this assumption by allowing the effect of peer ability in a particular class to decay over time, as well as to influence performance in other contemporaneous classes. This will help us determine the long-run impact of peer groups on educational achievement, and may result in higher peer effect estimates as we account for spillovers beyond the classroom.

Finally, rather than separately estimating ability by each course type, we can estimate models with a factor structure on ability and allow the returns to the different abilities to vary by course type. This will allow better exploitation of our data, as we will use information on a student's performance in all course types collectively in order to estimate individual ability levels.

A Appendix: Monte Carlo simulation method

Each Monte Carlo simulation consists of a two step procedure: generating student data, and running the generated data through our iterative algorithm. Since the details of the iterative algorithm are discussed in the paper, the focus here will be on explaining how the sample data are created.

Because of the many permutations of selection levels, number of observations per student, and random noise that we wish to simulate, numerous student samples are generated. In each simulated state of the world, 3,000 unique students are created. Each student is assigned an ID and ability level drawn from a $N(0,1)$ distribution. Each student's record is then replicated 4, 10, or 20 times. Class structure is identical in every simulated state of the world. Each course is composed of 5 sections, each containing 15 students. The total number of courses in each sample depends on the number of observations per student. For example, in the case of 4 observations per student, there are 160 courses and 800 associated sections.

The next step in creating the simulated data is to assign students to sections. Each student receives a sorting shock, drawn from a $N(0,1)$ distribution. This shock is independent of the student's own ability measure. In the case of random assignment, students are sorted by this shock, and placed into sections in this (random) order. This yields a within-section distribution of abilities matching the population distribution of abilities, since sorting into sections is entirely random. To generate selection into sections, the sorting shock is weighted and added to the individual's ability level to create a composite sorting score:

$$\text{CompositeScore}_i = \alpha_i + \text{weight} * \text{SortShock}_i. \tag{7}$$

Individuals are then sorted according to their composite score and placed into sections in order. The level of selection will depend on the weight assigned to the sorting shock. As the weight gets larger, the level of selection will decrease until there is essentially no selection on ability. Perfect selection would result if the weight of the sort shock were zero.

To avoid homogeneous ability levels within a course, we randomly assign sections to courses. This results in courses with varied ability levels overall, but students grouped by ability in sections. The procedure of assigning students to sections and sections to courses is repeated until the desired number of observations per student is reached.

The final step of the data generating process is to create a grade for each student-section observation. Grades are generated as follows:

$$Y_{ijt} = \alpha_i + \gamma\bar{\alpha}_{\sim ijt} + \epsilon_{ijt} \tag{8}$$

The α's represent individual ability levels, γ is the peer effect, and ϵ_{it} is a normally distributed shock. Once gamma is set and a vector of shocks has been drawn, all of the terms in the above equation are known, and a grade is calculated. To allow for course-specific shocks, or correlated effects at the course level, we simply append a second shock that is identical for all students in course j at time t as follows:

$$Y_{ijt} = \alpha_i + \gamma\bar{\alpha}_{\sim ijt} + \epsilon_{jt} + \epsilon_{ijt} \tag{9}$$

Finally, we perform our iterative algorithm on the generated sample of student grades and peer group assignments.

References

Altonji, J. G., Elder, T. E. & Taber, C. R. (2004), 'Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools', *Journal of Political Economy* . Forthcoming.

Arcidiacono, P. & Nicholson, S. (2004), 'Peer effects in medical school', *Journal of Public Economics* . Forthcoming.

Evans, W. N., Oates, W. E. & Schwab, R. M. (1992), 'Measuring peer group effects: A study of teenage behavior', *Journal of Political Economy* **100**(5), 966–991.

Hanushek, E. A., Kain, J. F., Markman, J. M. & Rivkin, S. G. (2003), 'Does peer ability affect student achievement?', *Journal of Applied Econometrics* **18**(5), 527–544.

Hoxby, C. (2001), Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper 7867.

Manski, C. (1995), *Identification Problems in the Social Sciences*, Harvard University Press, Cambridge, Massachusetts.

Sacerdote, B. I. (2001), 'Peer effects with random assignment: Results for Dartmouth roommates', *Quarterly Journal of Economics* **116**(2), 681–704.

Winston, G. C. & Zimmerman, D. J. (2003), Peer effects in higher education, *in* C. Hoxby, ed., 'College Decisions: How Students Actually Make Them and How They Could', University of Chicago Press.

Zimmerman, D. J. (2003), 'Peer effects in higher education: Evidence from a natural experiment', *Review of Economics and Statistics* **85**(1), 9–23.

Table 1: Monte Carlo Simulations: True Gamma = 0

Selection Level	Obs [a]	γ	R^2 [b]	Obs	γ	R^2	Obs	γ	R^2
Random	**4**	.0016	.40	**10**	-.0005	.28	**20**	-.0008	.25
		(.0447)			(.0384)			(.0292)	
		.0014	.54		.0029	.46		.0000	.43
		(.0331)			(.0265)			(.0192)	
		.0014	.77		.0003	.74		.0006	.73
		(.0192)			(.0145)			(.0103)	
		-.0016	.92		-.0004	.90		-.0001	.90
		(.0107)			(.0079)			(.0056)	
75% [c]	**4**	-.0013	.40	**10**	.0002	.26	**20**	.0004	.23
		(.0241)			(.0186)			(.0131)	
		.0008	.54		.0001	.42		.0003	.41
		(.0158)			(.0119)			(.0083)	
		-.0001	.77		-.0005	.71		-.0003	.71
		(.0084)			(.0062)			(.0043)	
		.0004	.91		-.0000	.89		-.0001	.89
		(.0045)			(.0034)			(.0023)	
50%	**4**	.0005	.38	**10**	-.0004	.26	**20**	.0003	.22
		(.0200)			(.0136)			(.0101)	
		.0005	.51		.0022	.43		.0004	.39
		(.0128)			(.0086)			(.0064)	
		.0009	.75		-.0002	.71		-.0001	.69
		(.0068)			(.0045)			(.0033)	
		-.0003	.90		.0001	.89		-.0001	.88
		(.0037)			(.0024)			(.0018)	
25%	**4**	.0049	.38	**10**	-.0019	.26	**20**	-.0007	.21
		(.0176)			(.0123)			(.0090)	
		.0014	.51		.0010	.41		-.0006	.38
		(.0112)			(.0078)			(.0056)	
		.0012	.74		.0013	.70		.0004	.68
		(.0059)			(.0041)			(.0029)	
		-.0000	.90		.0000	.88		-.0003	.88
		(.0032)			(.0022)			(.0016)	

[a] In this table, this abbreviation denotes the number of observations per student used in the given simulation.

[b] The R-squared values reported in this table are those pertaining to the regression of grades onto the constructed fixed effect values. We alter the random error added on to the constructed grade for each student in order to manipulate the amount of variation in performance that is explained by the ability measure. See the Appendix for more detail on the Monte Carlo methodology used.

[c] This percentage is 100 times the ratio of the average section-wide standard deviation of ability to the standard deviation of ability across the whole student population. Random assignment is equivalent to 100%; a lower number indicates more selection.

Table 2: Monte Carlo Simulations: True Gamma = .15

Selection Level	Obs [a]	γ	R^2 [b]	Obs	γ	R^2	Obs	γ	R^2
Random	**4**	.0809	.40	**10**	.1108	.28	**20**	.1255	.24
		(.0412)			(.0343)			(.0268)	
		.1127	.54		.1330	.46		.1391	.42
		(.0307)			(.0235)			(.0177)	
		.1377	.78		.1446	.74		.1473	.72
		(.0178)			(.0129)			(.0094)	
		.1456	.92		.1489	.90		.1493	.90
		(.0099)			(.0070)			(.0051)	
75% [c]	**4**	.1206	.41	**10**	.1325	.29	**20**	.1414	.25
		(.0223)			(.0165)			(.0122)	
		.1350	.56		.1442	.47		.1465	.44
		(.0146)			(.0106)			(.0077)	
		.1458	.79		.1474	.75		.1488	.73
		(.0078)			(.0056)			(.0041)	
		.1489	.93		.1495	.91		.1498	.90
		(.0042)			(.0030)			(.0022)	
50%	**4**	.1149	.43	**10**	.1298	.31	**20**	.1411	.27
		(.0172)			(.0123)			(.0091)	
		.1325	.58		.1433	.50		.1454	.47
		(.0111)			(.0078)			(.0057)	
		.1447	.81		.1479	.77		.1486	.76
		(.0059)			(.0041)			(.0030)	
		.1476	.93		.1494	.92		.1497	.91
		(.0032)			(.0022)			(.0016)	
25%	**4**	.0721	.43	**10**	.1030	.32	**20**	.1204	.28
		(.0150)			(.0109)			(.0081)	
		.1040	.58		.1263	.50		.1368	.47
		(.0098)			(.0070)			(.0051)	
		.1340	.81		.1417	.77		.1473	.76
		(.0053)			(.0037)			(.0027)	
		.1454	.93		.1478	.92		.1495	.92
		(.0029)			(.0020)			(.0015)	

[a] In this table, this abbreviation denotes the number of observations per student used in the given simulation.

[b] The R-squared values reported in this table are those pertaining to the regression of grades onto the constructed fixed effect values. We alter the random error added on to the constructed grade for each student in order to manipulate the amount of variation in performance that is explained by the ability measure. See the Appendix for more detail on the Monte Carlo methodology used.

[c] This percentage is 100 times the ratio of the average section-wide standard deviation of ability to the standard deviation of ability across the whole student population. Random assignment is equivalent to 100%; a lower number indicates more selection.

Table 3: Monte Carlo Simulations: True Gamma = .15, With Correlated Effects

Selection Level	Obs [a]	γ	R^2 [b]	Obs	γ	R^2	Obs	γ	R^2
Random	**4**	.0809	.40	**10**	.1072	.28	**20**	.1266	.25
		(.0447)			(.0384)			(.0292)	
		.1131	.54		.1301	.45		.1389	.43
		(.0332)			(.0265)			(.0192)	
		.1377	.77		.1435	.74		.1468	.73
		(.0192)			(.0145)			(.0103)	
		.1456	.92		.1482	.90		.1490	.90
		(.0107)			(.0079)			(.0056)	
75% [c]	**4**	.1194	.41	**10**	.1359	.28	**20**	.1441	.25
		(.0240)			(.0185)			(.0131)	
		.1371	.55		.1453	.45		.1473	.43
		(.0157)			(.0119)			(.0083)	
		.1472	.79		.1485	.73		.1494	.73
		(.0084)			(.0062)			(.0043)	
		.1493	.92		.1498	.90		.1499	.90
		(.0045)			(.0034)			(.0023)	
50%	**4**	.1189	.40	**10**	.1354	.29	**20**	.1412	.25
		(.0196)			(.0135)			(.0101)	
		.1352	.54		.1434	.47		.1464	.43
		(.0127)			(.0085)			(.0064)	
		.1462	.78		.1489	.75		.1492	.73
		(.0068)			(.0045)			(.0033)	
		.1489	.92		.1496	.91		.1496	.90
		(.0037)			(.0024)			(.0018)	
25%	**4**	.0807	.41	**10**	.1051	.29	**20**	.1248	.25
		(.0166)			(.0119)			(.0088)	
		.1071	.55		.1274	.47		.1387	.44
		(.0108)			(.0076)			(.0056)	
		.1359	.79		.1446	.75		.1468	.73
		(.0058)			(.0040)			(.0029)	
		.1478	.92		.1480	.91		.1490	.90
		(.0032)			(.0022)			(.0016)	

[a] In this table, this abbreviation denotes the number of observations per student used in the given simulation.

[b] The R-squared values reported in this table are those pertaining to the regression of grades onto the constructed fixed effect values. We alter the random error added on to the constructed grade for each student in order to manipulate the amount of variation in performance that is explained by the ability measure. See the Appendix for more detail on the Monte Carlo methodology used.

[c] This percentage is 100 times the ratio of the average section-wide standard deviation of ability to the standard deviation of ability across the whole student population. Random assignment is equivalent to 100%; a lower number indicates more selection.

Table 4: Sample sizes

	S99	F99	S00	F00	S01	F01	TOTAL
(1) Student-Sections	33,617	46,127	44,530	56,416	54,096	65,854	300,640
(2) Students	7193	9725	9568	11,936	11,606	13,908	63,936
(3) Unique Sections	3595	3990	3931	4270	4107	4464	24,357
(4) Unique Courses [a]	1352	1394	1501	1517	1587	1612	8963

[a]Figures represent the data set after applying the restrictions noted in the text. The unrestricted data set contained 351,940 student-section observations. Rows 3 and 4 show the total number of unique sections and courses, respectively, in which anyone in the sample during the given semester was observed.

Table 5: Means, standard deviations, and selection factors by course type: Observable achievement

	Humanities	Social Science	Science and Math	Vocational Education
SAT Math	613 (74)[a] *0.91*[b]	616 (74) *0.92*	649 (65) *0.77*	617 (70) *0.81*
SAT Verbal	603 (71) *0.90*	597 (70) *0.93*	616 (72) *0.88*	602 (65) *0.81*
High School GPA	3.57 (0.43) *0.92*	3.55 (0.44) *0.94*	3.70 (0.42) *0.89*	3.63 (0.41) *0.84*
Number of Observations	86,844	77,312	82,675	53,809
Number of Sections	8,217	5,199	7,400	3,541

[a]The numbers in parentheses give the average standard deviation of the variable, at the student level, within a section of the given course type. Theses figures are weighted by section size.

[b]The numbers in italics give the average standard deviation of the variable in a section divided by the standard deviation of the variable across all students in the course type.

Table 6: Peer effect results by Course type

Dep. Var: Section grade (demeaned)	Humanities	SocSci	MathSci	Vocational
Indep. Var.				
Section peer ability	.1158	.1478	.0397	.1380
	(.0096)	(.0110)	(.0083)	(.0154)
Class peer ability	-1.1158	-1.1478	-1.0397	-1.1380
	(.0139)	(.0138)	(.0093)	(.0185)
N	86,844	77,312	82,675	53,809
R^2	.5586	.5729	.6059	.5563

Table 7: Standard Deviations of Estimated Ability and Marginal Effects

Course Type	Section stddev[a]	Population stddev	Marginal Effect Ratio
Humanities	0.2836	0.6801	0.0483
Social Science	0.3123	0.7130	0.0647
Science and Math	0.5951	0.9497	0.0249
Vocational Education	0.2406	0.5704	0.0582

[a] "Section stddev" is the standard deviation of average section-level peer ability for students in sections of a particular course type. "Population stddev" is the standard deviation of individual ability in the population. These calculations are both based on the student fixed effects estimated by our model. "Marginal effect ratio" shows the ratio of (1) the effect on grades from a one-standard-deviation increase in peer ability to (2) the effect on grades from a one-standard-deviation increase in own ability.

Table 8: Correlations of estimated abilities among course types

Course type[a]	Humanities	SocSci	MathSci	Vocational
Humanities	1.0000			
SocSci	.6826*	1.0000		
MathSci	.6512*	.6842*	1.0000	
Vocational	.5906*	.5604*	.5475*	1.0000

[a] The estimated abilities used in this correlation matrix are those of the 10,722 students who took classes in all four course types.

Table 9: Regression of unobserved ability on observed ability

Dep. Var: Unobserved ability [a]

	Humanities		SocSci		MathSci		Vocational	
Indep. Var.	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Female	–	.26**	–	.17**	–	.15**	–	.21**
		(.01)		(.01)		(.01)		(.01)
Black	–	-.27**	–	-.23**	–	-.18**	–	-.16**
		(.02)		(.02)		(.02)		(.02)
Asian	–	-.12**	–	-.13**	–	-.10**	–	-.06**
		(.02)		(.02)		(.02)		(.02)
Hispanic	–	-.15**	–	-.18**	–	-.17**	–	-.09**
		(.03)		(.04)		(.04)		(.03)
Honors	–	.14**	–	.15**	–	.17**	–	.04*
		(.02)		(.02)		(.02)		(.01)
Sports	–	-.07**	–	-.14**	–	.01	–	-.04
		(.02)		(.03)		(.03)		(.02)
In-state	–	.05	–	.09*	–	.12**	–	-.02
		(.03)		(.03)		(.03)		(.03)
SAT - Math	-.01	.01	.11**	.11**	.36**	.36**	.05**	.07**
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
SAT - Verbal	.14**	.08**	.17**	.11**	.05**	.00	.04**	.01
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
High school GPA	.61**	.49**	.59**	.49**	.82**	.73**	.44**	.37**
	(.01)	(.01)	(.01)	(.02)	(.02)	(.02)	(.01)	(.01)
N	17,333	17,332	15,265	15,264	16,078	16,077	15,315	15,315
R^2	.19	.22	.22	.24	.33	.34	.14	.17

[a] The dependent variable is the student-level fixed effects estimated in model 2. The variables "Honors" and "Sports" are the average across the student's tenure at the University of dummy variables for honors and sports program participation each semester, respectively. The "In-state" variable is the average across the student's tenure of a variable taking the value 1 if the student was resident in-state during that semester. The excluded racial/ethnic category is white; racial/ethnic categories are mutually exclusive.

Table 10: Conventional peer effect results using estimated abilities - By Course type

Indep. Var.[a]	Own ability	Section peers' ability	Effect of 1-stddev chg in peer ability	Own ability type	Peers' ability type	N	R^2
Humanities	1	.12**	.03	Est.	Est.	86,844	.58
	(–)	(.01)					
	1.01**	.04	.01	Est.	Pred.	86,838	.58
	(.00)	(.01)					
	.92**	.19**	.03	Pred.	Pred.	86,838	.15
	(.01)	(.02)					
Social Sciences	1	.15**	.05	Est.	Est.	77,312	.61
	(–)	(.01)					
	1.01**	.09**	.02	Est.	Pred.	77,307	.61
	(.00)	(.01)					
	.85**	.40**	.07	Pred.	Pred.	77,307	.16
	(.01)	(.02)					
Math Sciences	1	.04**	.02	Est.	Est.	82,675	.70
	(–)	(.00)					
	1.01**	.02	.01	Est.	Pred.	82,674	.70
	(.00)	(.01)					
	.80**	.58**	.22	Pred.	Pred.	82,674	.30
	(.01)	(.01)					
Vocational	1	.14**	.03	Est.	Est.	53,809	.60
	(–)	(.01)					
	1.01**	.11**	.02	Est.	Pred.	53,809	.60
	(.00)	(.01)					
	.94**	.20**	.03	Pred.	Pred.	53,809	.14
	(.01)	(.02)					

[a]Section grade is demeaned at the course level, and then added to the course mean grade effect estimated in our model multiplied by the average of estimated abilities in that course. This procedure simply adjusts the dependent variable to be comparable across classes before the regressions are run, so that we need not estimate a new course ability effect in each model that could interfere with our examination of the effects of selection and measurement error on the peer effect estimates in this table. To test the effects of measurement error in both own and peers' ability, predicted values from the equation run to produce column 2 under each course type in the preceding table are used in some regressions to form either, or both, of own or peers' ability estimates. Regressions are performed separately under three alternative schemes: (1) actual estimated true ability and actual estimated peers' ability; (2) actual estimated own ability and predicted peers' ability; and (3) predicted own and peers' ability. Parameter estimates from each of these schemes are presented in rows (1), (2), and (3), respectively, along with the predicted increase in grade of a one-standard deviation rise in peer ability.

Table 11: Selection Into Peer Groups

Course Type	Section stddev	Population stddev	Selection Ratio[a]
Humanities	.6143	.6801	.9032
Social Science	.6361	.7130	.8921
Science and Math	.7474	.9497	.7870
Vocational Education	.4954	.5704	.8685

[a]This is the average standard deviation of estimated ability within a section, weighted by section size, divided by the standard deviation of ability in the population, separately for each course type.

Table 12: Selection on Observables vs Selection on Unobservables

	Predicted part of Alpha (Observed)[a]			Residual part of Alpha (Unobserved)		
Course type	Section Level[b] (avg stdev)	Student Level (stdev)	Selection	Section Level (avg stdev)	Student Level (stdev)	Selection
Humanities	.3383	.3719	**.9095**	.5374	.5867	**.9160**
SocSci	.3764	.4114	**.9151**	.5659	.6187	**.9148**
MathSci	.5038	.6174	**.8159**	.6664	.7462	**.8930**
Vocational	.2333	.2817	**.8280**	.4537	.5035	**.9012**

[a]In order to obtain a measurement of selection, we first broke the estimated fixed effects into the sum of two parts. The first part is the predicted value from a regression of the fixed effect onto the observed variables for each student shown in column 2 for each course type in Table 9. The second part is the residual from that regression. Then, we took the average standard deviation of each part of our fixed effect estimate at the section level, weighted by section size, and divided it by the standard deviation of that same part of the fixed effect at the student level, analogous to what was done in the previous table. This yields the selection ratios shown in the table.

[b]Refer to previous tables for sample sizes.

Table 13: Discipline classification scheme

Group[a]	Included Disciplines:
bsos	African American Studies, Anthropology, Behavioral and Social Sciences, Criminology and Criminal Justice, Economics, Geography, Government and Politics, Hearing and Speech Sciences, Psychology, Sociology, Survey Methodology
ed	Curriculum and Instruction, Education Counseling and Personnel Services, Human Development Education, Measurement, Statistics, and Evaluation, Education Policy and Leadership, Special Education, Education
hhp	Family Studies, Health and Human Performance, Kinesiology
is	Library Science
j	Journalism, NEC Courses
pa	Public Affairs
sb	Business and Management
se	Aerospace Engineering, Civil Engineering, Chemical Engineering, Cooperative Education Engineering, Electrical and Computer Engineering, Engineering Science, Fire Protection Engineering, Materials Engineering, Mechanical Engineering, Nuclear Engineering, Reliability Engineering, Systems Engineering, Telecommunications, Gemstone Honors Program Courses, Cooperative Education Program

[a]This taxonomy is based on the University of Maryland's system of discipline groupings.

Table 14: Discipline classification scheme

Group[a]	Included disciplines:
ah	American Studies, Arabic, Arts and Humanities, Art History and Archaeology, Art Studio, Chinese, Classics, Comparative Literature, Communication, Dance, East Asian Languages and Literatures, English, French, German, Greek, Hebrew, History, Italian, Japanese, Jewish Studies, Korean, Latin American Studies, Latin, Linguistics, Music Education, Ethnomusicology, Music, Music Performance, Philosophy, Portuguese, Russian, Slavic, Spanish, Theatre, Women's Studies, Study Abroad
anr	Agriculture and Natural Resources, Animal Science, Agricultural and Resource Economics, Biometrics, Biological Resources Engineering, Environmental Science and Policy, Landscape Architecture, Nutrition and Food Science, Natural Resources Management, Natural Resource Sciences, Plant Sciences, Veterinary Medical Sciences
app	Architecture, Historic Preservation, Urban Studies and Planning
ug	Asian American Studies, Air Science, College Park Scholars Honors Program Courses, Honors Program Courses, Individual Studies Program, University Courses

[a]See footnote from previous table.

Table 15: Discipline classification scheme

Group[a]	Included disciplines:
cmps	Applied Mathematics and Scientific Computation, Astronomy, Computer, Mathematical and Physical Sciences, Computer Science, Geology, Mathematics, Meteorology, Physics, Statistics and Probability
ls	Biochemistry, Biology, Biological Sciences Program, Chemistry, Sustainable Development and Conservation Biology, Entomology, Marine-Estuarine-Environmental Sciences, Microbiology, Molecular and Cell Biology, Neuroscience and Cognitive Science, Plant Biology

[a]See footnote from previous table.