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

October 22, 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 uses these student fixed effects to formulate the ability level of the peer group. Monte Carlo evidence shows that our algorithm performs well even when with a short panel. While it is generally thought that peer effect estimates are biased upward, we find strong evidence to the contrary for both our technique and the standard selection on observables approach. The bias due to measurement error in peer ability is downward and stronger than the upward bias associated with selection into the peer group. We demonstrate the technique using transcript data from University of Maryland undergraduates. We find statistically significant peer effects, particularly in courses of a collaborative nature.

1 Introduction

Within a community—whether it be a country, state, district, or individual school—one of the most important policy objectives is promoting a high quality of education. Yet, there

*We wish to thank participants at the 2004 Labour Econometrics Workshop held at the University of Auckland, 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.

is much debate on the most effective way to attain this goal. One policy lever highlighted in this debate is the distribution of students within a classroom, as students may affect the achievement of others. For example, the average ability student may have higher than average academic achievement if placed in a study session full of high ability students. Peer effects can also work in the opposite direction, where some students' academic achievement is harmed by being in contact with certain types of peers. Determining the existence of peer effects and measuring their magnitude are imperative for effective educational policy-making.

Measuring peer effects, however, has proved to be extremely difficult. There are at least two barriers that must be overcome when estimating peer effects in student achievement. First, the underlying abilities of the students are measured with error. Second, peer groups are in general not randomly assigned. 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 the peer group itself but rather from not measuring the ability of the individual accurately. On the other hand, having noisy measures of the true abilities of an individual's peers may bias the estimated peer effect downward. The use of test scores, such as the SAT or MCAT, as a measure 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.[1]

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. One set of papers uses proxy variables to break the link between unobserved and peer ability.[2] These papers rely on noisy measures of underlying peer ability. The proxy variables may then absorb some of the peer effect, biasing the estimated peer effect downward, or may not fully remove the correlation between unobserved ability and peer ability, biasing the estimated peer

[1] In the language of Manski (1995), this discussion has focused on distinguishing between exogenous and correlated effects. It may be possible to extend our models to endogenous effects as well, but we leave this to future research.

[2] 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.

effect upward. Another set of papers relies on some form of random assignment.[3] However, peer influence may be much stronger among individuals who choose their peers than among those who are randomly assigned. Finally, researchers have used to try to circumvent the endogeneity problem is instrumental variables.[4] This method, while attractive in theory, suffers from the drawback that the suggested instruments (traditionally, attributes of the entire population in an individual's neighborhood) may not be valid. Misspecification of the peer group (or, more generally, assumption of homogeneity of peer effects) is a blanket concern across most of the existing literature, regardless of identification strategy, and can result in added bias on conventional peer effects estimates.

Our estimation strategy allows us to both accurately measure peer ability and accurately estimate how peer ability affects the individual's outcome of interest without the requirement of random assignment. The only requirement of the method is variation in both the outcome measure and the composition of the peer group over time. We propose an iterative algorithm that controls for student fixed effects. Further, rather than using test scores as a noisy measure of peer ability, we measure peer ability using the individual fixed effects of the student's peers.

Monte Carlo results suggest that the algorithm performs well even when the number of outcomes per student is small. Further, the simulations suggest that the use of noisy measures of ability– both of the student and the student's peers– lead to biased downward estimates of peer effects. That is, the downward bias associated with measurement error is much stronger than the upward bias associated with selection into the peer group.

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 course type, and then with the full sample, where we conceive of grades in different course types as distinct outcomes. We find significant peer

[3] Sacerdote (2001), Zimmerman (2003), and Winston & Zimmerman (2003) rely on random assignment of roommates. This strategy limits a researcher since the random assignment of roommates occurs only in the first year, and might have limited influence on outcomes in subsequent years. 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 the individual's unobserved ability.

[4] The first paper to apply this technique in the peer effects context was Evans, Oates & Schwab (1992)

effects which vary by course type. A one point increase in peer ability is equivalent to between a 0.076 and a 0.165 increase in own ability depending upon the course type. Higher spillovers are found in vocational courses and courses in the social sciences with spillovers the smallest in the math and the 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 students choose their peer group. Section 3 describes the University of Maryland data. Section 4 presents the results. Section 5 shows the biases associated with using data only from observable ability measures. Section 6 concludes.

2 Model and Estimation

In this section we present the model and estimation strategy. The premise behind the model is that an individual's outcome has a component that is fixed over time: the individual's fixed effect. However, the outcome also depends upon the fixed effects of the other individuals in the peer group. We then show how an iterative estimation algorithm can be used to estimate the peer effect and the corresponding fixed effects for the individuals.

Student i's outcome at time t depends only on student i's own ability and the average ability of student i's peer group. Including unobserved ability in our model allows us to abstract from many other covariates typically found in a school outcome regression. All of the heterogeneity in student characteristics that might effect grades, 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. The baseline model is formulated as follows:

$$Y_{it} = \alpha_i + \gamma \bar{\alpha}_{-it} + \epsilon_{it} \tag{1}$$

where α_i is student i's innate ability and $\bar{\alpha}_{-it}$ is the average ability of all other students in the individual's peer group at time t. The last term, ϵ_{it}, is a random shock that is orthogonal to both peer and individual ability. The key parameter is γ, which measures the influence peer ability has on the student's outcome.

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.[5] Utilizing this initial guess, we can calculate the average abilities of all other students in the peer group at time t. Estimation of γ given the α's can then be obtained by ordinary least squares. In particular, denote α_i^0 the initial guess on the α's. We estimate γ^0 using the following:

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

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,

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

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 OLS on the following equation:

$$Y_{it} - \alpha_i^n = \gamma^n \bar{\alpha}_{-it}^n + \epsilon_{it}^n$$

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

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

As long as the γ's remain less than one, the equilibrium is unique and the iterative procedure converges to the true set of abilities and peer effects.

3 Monte Carlo Simulations

In this section we simulate the model under different assumptions regarding selection into classes, the number of observations per individual, and the degree of noise present in our measure of the student's grade. In each case, our algorithm never overestimates the peer effects. The algorithm may underestimate peer effects when the number of observations per student is small, the fixed effects explain little of the variation in the outcome measure, and

[5] In practice, we use the average value of the outcome measure as our initial guess of the α's, though the algorithm is not sensitive to the starting values.

peer groups are assigned randomly. The algorithm performs particularly well when there is some selection into the peer groups.

The number of outcomes observed per student varies across simulations between 4, 10, and 20. 4 is likely the maximum number of observations a researcher may have when analyzing grade school or high school test score data, and 20 observations reflects our university level data. As the number of observations per student increases, the accuracy with which we estimate unobserved ability will increase.

Three levels of selection are modeled: random assignment, low selection, and high selection. The statistic used to measure selection is the difference between the average within class standard deviation of abilities and the population standard deviation of student abilities. Student abilities within the population are all drawn from a N(0,1) distribution. Using a sorting algorithm that groups individuals with similar abilities together, students are assigned to classes. Sorting in this manner decreases the amount of variation in ability within a section. In the case of low selection, students are sorted such that the within-section standard deviation of abilities is approximately .5, half that of the population measure. The high selection sorting procedure reduces the within-section ability deviations even further to .25. Students are randomly assigned to classes under the no selection procedure, yielding a within-section variation of abilities that matches the variation in the population. Student outcomes can now be generated according to the baseline model by calculating the average ability of a student's classmates, and adding the students own ability and random error term. Note that in our data the average standard deviation of ability across peer group is 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.

The addition of the random noise term takes on added importance in our model. Typically when estimating a regression coefficient the amount of noise that goes unexplained does not impact the consistency of the estimate. In our case however, we are estimating not only the coefficients, but the regressors as well. Thus increasing the variation of the error term reduces our capacity to accurately estimate student abilities, and hence impacts the consistency of the peer effect estimate. Standard errors-in-variables theory suggests that our peer effect measure will likely be biased downward.

The results of our baseline simulations are presented in Table 1, where the true peer effect in all simulations is .15.[6] Notice that as we move across each row, decreasing the amount of noise in the model, the peer effect steadily improves, highlighting the importance of an accurate ability measure. In addition, as we increase the number of observations per student within each selection category, the peer effect estimate typically improves, again indicating the importance of measuring ability precisely. Note that the algorithm handles selection quite well and actually performs better in those cases where peer groups are not randomly assigned.

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,[7] 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 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.

Estimation of the models discussed above required the conversion of grade information into numerical scores.

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 sub-section of the entire class, 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.

[6] We experimented with different values for γ and in all cases the algorithm performed quite well. We were particularly concerned that the algorithm did not estimate positive peer effects when the in fact the true coefficient was zero. The simulations when for $\gamma = 0$ yielded extremely small and insignificant peer effect estimates in all cases of selection, fit, and number of student observations.

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

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 that semester.[8] These restrictions yielded a sample of 300,640 student-section observations, representing 18,554 individual students. Sample sizes are provided in Table 2; some descriptive information regarding the sample's diversity is provided in Table 3.

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 one of the following course types: Humanities, Science and Math, Social Science, and Vocational. Each student will have an independent ability measure for each type of course, conditional on enrolling in at least one class within a course type. [9]

Under the assumption of multiple ability measures, the model we believe yields the most reliable estimate of a student's grade *in a specific type of course* is:

$$Y_{it} = \alpha_i + \gamma\bar{\alpha}_{-it} + \delta\tilde{\alpha}_{-it} + \epsilon_{it} \tag{2}$$

The above model will be estimated separately for each type of course, yielding four section peer effect and class effect estimates, as well as multiple ability measures for each student.

The applied model of student grades contains one additional term that the baseline model does not, namely $\delta\tilde{\alpha}_{-it}$. This term is included to capture the hierarchical nature of our data. $\tilde{\alpha}_{-it}$ represents the average ability of all of the students in a particular course. A student's grade and peer group are defined at the section level, but each section also belongs to a course.

[8] 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.

[9] The four course types used in practice were based on the more detailed subgroups presented in Tables XX through XX. The humanities course type contains the subgroups "ah" and "app;" the science and math 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."

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 is extremely able, a mediocre student's grade may suffer. By including $\delta\tilde{\alpha}_{-it}$, we can account for these potential class-level effects on a student's grade. In order to eliminate any effects of heterogenous grading practices, we normalize the average grade within each course to zero before estimating our model.

5 Estimates of Classroom Spillovers

Table 4 shows the results from estimating Equation 2 for all four types of courses. The results indicate positive and significant section peer effects in all cases. It is interesting to note the increase in magnitude of the section peer effect as we move from the math and science courses to the social sciences and vocational courses. This result may reflect the amount of collaborative work required in each course type. In addition, if the the environment in course types such as social sciences and vocational is more conducive to communicating knowledge across peers, larger peer effects may result in these courses. The negative coefficients that are close to one on the course peer ability result mechanically from normalizing the mean grade in a course to zero.

The magnitudes of the section level peer effects show that a one point increase in peer ability is equivalent to a high of 0.165 increase in individual ability in vocational classes and a low of 0.076 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 5 displays the relevant standard deviations and the effect of a one standard deviation increase in peer ability over the effect of a one standard deviation in individual ability. The gap between math and science and the other course types shrink as the standard deviation of peer ability relative to the population standard deviation is much larger. As we will show in a moment, this is because and science courses have much more selection into sections than other course types. A one standard deviation increase in peer ability is equivalent to a high of 0.065 increase in individual ability in the social sciences to a low of 0.043 increase in individual ability in math and natural sciences.

An important byproduct of our estimation is course type specific student abilities. Table 6 shows the correlation coefficients among estimated ability levels across course types. These correlations are created using estimated abilities from students observed in all course types.[10] The highest correlation coefficient is for humanities and social sciences at 0.67. This is as we would expect since a student who has high ability in humanities most likely has a high skill level in reading and writing papers. These skills translate most into skills needed for social sciences. Vocational and math and science show the smallest correlation coefficient at 0.49.

6 Comparing the Method to Selection on Observables

Next, we can analyze the difference between student ability measured using our iterative algorithm versus using standard background ability measures. We then compare estimates between our method with conventional peer effect methods. In order to facilitate this comparison we need a measure of both observed peer and individual ability. We therefore first create a conglomerate measure of ability by regressing the estimated individual abilities (the individual fixed effects) on the two components of SAT score (math and verbal) and high school grade point average at by course type at the student level. The R^2 for these regressions range from .11 to .24, indicating that using only background ability measures are not a good proxy for what we argue to be true student ability. Even including a set of grade-related controls such as gender, race, and participation in sports and honors programs, we can still only explain 26% of the variation in section grade.

Results from this first step are presented in Table 7. The statistical significance and magnitude of the SAT coefficients vary across course type as expected. For example, SAT-Math scores are insignificant when explaining Humanities ability, but are large and significant when predicting math and science ability.

Although the three measures of background ability are generally positive and significant in predicting unobserved abilities, there remains a large portion of the fixed effects to be explained. This result underlines the degree to which a student's own and peer ability are

[10] Correlations among estimated ability levels were also calculated including all students with valid observations. Similar coefficients resulted.

measured with error in conventional peer effects studies. A primary concern in this paper is the consequences of ability mismeasurement on peer effects estimates. As discussed in the introduction, we would expect this measurement error to result in downward pressure on peer effects estimates. However, when ability mismeasurement is combined with student selection into peer groups, the end result could be an upward bias on peer effect estimates.

To empirically test the impact of measurement error and selection on conventional peer effect estimates, we generate a set of predicted, course-type specific ability measures using the models highlighted in Table 7. These abilities are a proxy for the observed ability measure typically used when estimating peer effects. Then, we can re-estimate Equation 2 using either the predicted ability, our own estimated ability, or a combination of the two measures. Estimation of the model using only the predicted abilities will suffer from both selection and measurement error, while running the model using our estimated fixed effects as own ability, and the predicted values from Table 7 as peer ability, the model suffers only from measurement error. 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 listed in Table 8. For each course type, three rows of results are shows. Row (1) reflects the results from our iterative procedure (also shown in Table 4). The results listed in row (2) highlight the bias due to measurement error since the predicted values from Table 7 are used to create the peer abilities for each student. In this case, the section peer effect is dramatically underestimated. Finally, the specification in row (3) injects selection bias into the section peer effect estimate by using the predicted abilities from Table 7 to calculate own and peer ability.

Controlling for individual ability but using noisy ability measures for peer ability (row 2 for each course type) resulted in peer effects that were very small and sometimes even negative. The only positive and significant peer effect was found in the social sciences and here the coefficient was a quarter of what we estimated with our method.

The most surprising results, however, are in the third row for each course type. Here we find the biases associated with measurement error and selection effectively cancel out in the social sciences as well as in math and the natural sciences. For humanities and vocational, however, the peer effect estimates using noisy measures of individual and peer ability are downward biased. For our data, it appears as though measurement error is more important than the selection problem. Hence, not controlling actually moves us closer to the true peer effect.

Why are the estimates higher for math and science and for social sciences when we control selection using selection on observables? We expect this to be the case if selection into peer groups is more important for math and science and social science sections— the upward bias from selection into peer groups would then be stronger. Table 9 presents some evidence that this is the case. The first column shows the average standard deviation of ability within a section, while the second column shows the corresponding average standard deviation of ability across the population. 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 78% and 85% of the standard deviation in the population, with the most selection occurring in the math and science courses.[11]

7 Conclusion

We present a new iterative method that controls for an individual's own ability and accurately measures the ability of the peer group. Conditional on our definition of the peer group, selection into the peer group is no longer an issue . In order to estimate this method, there must be panel data on outcomes where the peer group changes over time. These outcomes depend upon an the fixed effects of the individual and 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. The method never yielded upward 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 the peer ability and heaviest on a student's own fixed effect.

[11]Note that this level of selection is small level of compared to the selection imposed in our monte carlo experiments. It is possible that the population of students at the university has already been selected and is likely more homogenous than would be found in the population of college age individuals. Grade school children are likely to experience greater levels of selection, which may cause the error from selection bias to dominate measurement error.

Our estimates are higher than the corresponding estimates using selection on observables. This suggests that mismeasuring peer ability— which results in a downward bias of peer effect estimates— may sometimes be more important than the upward bias associated with selection. Indeed, the selection on observables method produced results closest to our method for the the course type with the most selection.

There are many avenues to be explored in future research. First, we have assumed that all individuals in the discussion section influence each other equally. Allowing the peer effect to vary across same gender or same race lines may yield even higher peer effect estimates as we move to better measures of who is in the peer group. Similarly, we can examine whether high or low ability individuals are affected most by the abilities of others. Second, we have assumed that peer effects do not persist but only help in the particular course. Future work will extend the model to include allow the effect of peer ability in a particular class to decay — or not decay— over time. 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.

References

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

Selection Level	γ	R^2	γ	R^2	γ	R^2	γ	R^2
No Selection								
4 obs/student	.0892	.3451	.1206	.4997	.1428	.7560	.1485	.9102
	(.1141)		(.0830)		(.0469)		(.0259)	
10 obs/student	.1019	.2800	.1303	.4504	.1479	.7303	.1498	.9102
	(.0982)		(.0676)		(.0372)		(.0203)	
20 obs/student	.1210	.2400	.1396	.4206	.1480	.7176	.1497	.8960
	(.0770)		(.0508)		(.0271)		(.0147)	
Low Selection								
4 obs/student	.1296	.4258	.1345	.5792	.1498	.8069	.1498	.9315
	(.0501)		(.0321)		(.0171)		(.0092)	
10 obs/student	.1360	.3095	.1383	.4940	.1482	.7672	.1491	.9174
	(.0358)		(.0226)		(.0199)		(.0064)	
20 obs/student	.1475	.2715	.1496	.4650	.1483	.7543	.1501	.9127
	(.0264)		(.0166)		(.0087)		(.0047)	
High Selection								
4 obs/student	.1018	.4332	.1242	.5884	.1442	.8149	.1499	.9350
	(.0433)		(.0281)		(.0150)		(.0081)	
10 obs/student	.1065	.3880	.1160	.5557	.1416	.8000	.1479	.9296
	(.0384)		(.0246)		(.0131)		(.0071)	
20 obs/student	.1311	.2832	.1437	.4808	.1465	.7659	.1504	.9177
	(.0288)		(.0145)		(.0076)		(.0041)	

Table 2: 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 required by the single-ability models. The data set for the two-ability models contains a total of 298,667 student-section observations. 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 3: Means and standard deviations of demographics and achievement

	Base sample
Percent black[a]	**13.79**
Percent Asian	**12.15**
Percent female	**47.85**
Mean high school GPA	**3.61 (.48)**
Mean SAT score	**1229 (143)**
N	**18,554**

[a]Calculations are performed at the student level. Standard deviations are in parentheses.

Table 4: Peer effect results by Coursetype

Dep. Var: Section grade - demeaned	Humanities	SocSci	MathSci	Vocational
Indep. Var.				
Section peer ability	.1320	.1609	.0765	.1651
	(.0103)	(.0116)	(.0091)	(.0178)
Class peer ability	-.9925	-1.0355	-.9601	-1.0178
	(.0103)	(.0116)	(.0091)	(.0178)
N	86,844	77,312	82,675	53,809
R^2	.5534	.5669	.5780	.5400

Table 5: Standard Deviations and Marginal Effects

Course Type	Section stddev	Population stddev	Marginal Effect Ratio[a]
Humanities	0.0888	0.2221	0.0528
Social Science	0.0945	0.2347	0.0648
Science and Math	0.1925	0.3408	0.0432
Vocational Education	0.0636	0.1696	0.0619

[a]This is the Marginal effect of the peers ability divided by the marginal effect of own ability.

Table 6: Correlations of estimated abilities among coursetypes

Coursetype[a]	Humanities	SocSci	MathSci	Vocational
Humanities	1.0000			
SocSci	.6705*	1.0000		
MathSci	.5997*	.6153*	1.0000	
Vocational	.5667*	.5361*	.4863*	1.0000

[a]The abilities used in this correlation matrix are from observations who took classes in all four coursetypes

Table 7: Regression of unobserved ability on observed ability - Humanities - Demeaned

Dep. Var: Unobserved ability [a]								
	Humanities		SocSci		MathSci		Vocational	
Indep. Var.	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Female	–	.08**	–	.06**	–	.09**	–	.06**
		(.00)		(.00)		(.01)		(.00)
Black	–	-.09**	–	-.07**	–	-.12**	–	-.05**
		(.01)		(.02)		(.01)		(.01)
Asian	–	-.04**	–	-.05**	–	-.06**	–	-.02**
		(.01)		(.01)		(.01)		(.00)
Hispanic	–	-.05**	–	-.06**	–	-.10**	–	-.03**
		(.01)		(.01)		(.02)		(.01)
Honors	–	.05**	–	.05**	–	.06**	–	.00
		(.01)		(.01)		(.01)		(.00)
Sports	–	-.02**	–	-.04**	–	-.00	–	-.02*
		(.01)		(.01)		(.02)		(.01)
In-state	–	.01	–	.03*	–	.03	–	-.00
		(.01)		(.01)		(.02)		(.01)
SAT - Math	-.00	.00	.04**	.04**	.11**	.11**	.01**	.02**
	(.00)	(.00)	(.00)	(.00)	(.00)	(.01)	(.00)	(.00)
SAT - Verbal	.04**	.02**	.06**	.04**	.02**	-.01	.01**	.00
	(.00)	(.00)	(.00)	(.00)	(.00)	(.01)	(.00)	(.00)
High school GPA	.20**	.16**	.20**	.17**	.32**	.27**	.12**	.10**
	(.00)	(.00)	(.01)	(.01)	(.01)	(.01)	(.00)	(.00)
N	17,333	17,332	15,265	15,264	16,078	16,077	15,315	15,315
R^2	.18	.21	.22	.24	.24	.26	.11	.14

[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 8: Conventional peer effect results using estimated abilities - By Coursetype - Demeaned[a]

Indep. Var.	Own ability	Section peers' ability	Course peers' ability	Own ability type	Peers' ability type	N	R^2
Humanities	1	.13**	-.99**	Est.	Est.	86,844	.55
	(–)	(.01)	(.01)				
	.99**	-.07**	-.87**	Est.	Pred.	86,844	.53
	(.00)	(.02)	(.03)				
	.90**	.03	-.83**	Pred.	Pred.	86,844	.11
	(.01)	(.03)	(.04)				
Social Sciences	1	.16**	-1.04**	Est.	Est.	77,312	.57
	(–)	(.01)	(.01)				
	.97**	.04*	-1.15**	Est.	Pred.	77,312	.54
	(.00)	(.02)	(.03)				
	.82**	.18**	-.93**	Pred.	Pred.	77,312	.12
	(.01)	(.03)	(.04)				
Math Sciences	1	.08**	-.96**	Est.	Est.	82,675	.58
	(–)	(.01)	(.01)				
	.94**	.01	-1.09**	Est.	Pred.	82,675	.54
	(.00)	(.01)	(.02)				
	.78**	.10**	-.84**	Pred.	Pred.	82,675	.13
	(.01)	(.02)	(.02)				
Vocational	1	.17**	-1.02**	Est.	Est.	53,809 .54	
	(–)	(.02)	(.02)				
	.98**	-.06	-.78**	Est.	Pred.	53,809	.51
	(.00)	(.05)	(.12)				
	1.07**	-.15*	-.91**	Pred.	Pred.	53,809	.07
	(.02)	(.07)	(.17)				

[a]Section grade is demeaned at the course level. To test the effects of measurement error in both own and peers' ability, predicted values from the equation run to produce the preceding table are used in some regressions to form either the 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 under the columns (1), (2), and (3), respectively.

Table 9: Selection Into Peer Groups

Course Type	Section stddev	Population stddev	Selection Ratio[a]
Humanities	.1891	.2221	.8514
Social Science	.1999	.2347	.8517
Science and Math	.2658	.3408	.7799
Vocational Education	.1423	.1696	.8390

[a]This is the standard deviation of ability within a section divided by the population standard deviation for each course type.

Table 10: 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 11: 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 12: 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.