Classmate Peer Effects: Evidence from Core Courses at Three Colleges*

Jeffrey Parker, Reed College

James Grant, Lewis & Clark College

Jan Crouter, Whitman College

Jon Rivenburg, Reed College

Presented to Mellon Workshop on
Evaluating Teaching and Learning at Liberal Arts Colleges
Wellesley College
September 2008

Abstract

This paper takes advantage of random assignment of first-year students into core-course sections at three liberal-arts colleges to test for classmate peer effects. Our statistical analysis finds no evidence that more able core-course classmates affect students' success outside the core course. While there are several possible explanations for this, interviews with core-course instructors suggest that peers' attitudes and personalities have greater effects than their raw academic abilities.

*This paper grew out of an earlier study of classmate peer effects at Reed College, which was undertaken by Jessica Hoel with Parker and Rivenburg under the sponsorship of the Bernard Goldhammer Collaborative Grant Program. The current research is part of a project supported by an Outcomes and Assessments Grant from the Teagle Foundation. Nigel Nicholson, Ruth Russo, Ben Westervelt, Jay Beaman, and Neal Christopherson are members of the Teagle project research team and contributed, directly and indirectly, to this research. The project also benefited from outstanding research assistance from numerous students at all three colleges, including Nick Chandler-Klein, Kathryn Deane, Christa Gibbs, Ranganai Gwati, Joe Hand, Ben Keefer, David Kling, Joshua Lingerfelt, Morgan Matthews, Ramon Mitchell, Daniela Morar, Dylan Smith, Michael Turner, Izarra Varela, and Liza Wickline.

Introduction

Colleges and their applicants believe that the quality of one's classmates matters. Most colleges and universities strive to admit the highest-quality students they can attract; increasing selectivity is often an institutional goal. From a student's perspective, a better pool of students enables a more challenging curriculum and, so the story goes, a better education through interaction with better peers inside and outside the classroom. The theoretical analysis of Rothschild and White (1995) supports the use of "merit aid" to compensate students who contribute to the education of their peers. At schools that award merit aid, it is usually given to students with very high academic potential as demonstrated by the standard admission credentials: standardized test scores, high-school record, and information gleaned from application essays and interviews.

Applicants often choose to attend the "best" school to which they are admitted, provided it is financially feasible. While outstanding faculty instructors are surely part of the reason, high-quality peers are widely believed to be valuable as well—both for the positive educational effect they may have and also because identification with a school known for high-quality students may enhance a student's post-graduate opportunities.

In recent years, economists have devoted considerable attention to the question of how students' learning is influenced by their peers. Many peer-effects studies have focused on primary- or secondary-school classmate effects, for example Henderson, Miseszkowski, and Sauvegeau (1976), Epple and Romano (1998), Hoxby (2000), Zimmer and Toma (2000), Boozer and Cacciola (2001), Gaviria and Raphael (2001), Vandenberghe (2002), Hanushek et al. (2003), McEwan (2003), Rangvid (2003), Robertson and Symons (2003), Zimmer (2003), Angrist and Lang (2004), Dills (2005), Lefgren (2004), Ammermueller and Pischke (2006), Burke and Sass (2006), Ding and Lehrer (2006), and Kang (2007). Although the evidence is not unanimous, many studies find support for the hypothesis that young students learn more if they have peers of higher ability.

Studies of peer effects in higher education have been more limited. When students select their peers on the basis of variables we cannot observe, it becomes impossible to identify the effects of the student's own characteristics from the effects of peers.[1] Most of the work on peer effects in higher education has focused on roommate and dorm-mate effects because these peers are assigned by colleges either randomly or based on observable criteria. Studies based on residential peers include Sacerdote (2001), Zimmerman (2003), Winston and Zimmerman (2004), Stinebrickner and Stinebrickner (2006), Foster (2006), Lyle (2007), Kremer and Levy (2008), and Stinebrickner and Stinebrickner (2008). These studies have found some evidence that more able roommates are beneficial; the most recent studies have suggested that the mechanism of roommate peer effects may operate through study time and alcohol use, which are more closely related to the attitudes and personalities of roommates than to their innate academic ability.

There is a significant disconnect in the peer-effects literature between the primary and secondary studies, which focus for obvious reasons on classmates or schoolmates, and the higher-education studies, which look primarily at roommates.[2] A major reason for this is that college classmates are rarely assigned by random or observable methods. Since students enroll voluntarily in the majority of their classes and sections, they may choose their classmates either explicitly—say, to take classes with their friends—or implicitly as when students seeking a challenge choose a difficult course, an honors section, or a section taught by a notoriously difficult professor.

We explore classmate peer effects in higher education by taking advantage of quasi-random assignment of freshmen into core-course sections at three Northwest liberal-arts colleges. Our statistical analysis estimates the effects of various characteristics of the classmate-ability distribution on student outcomes. We find little evidence of significant peer effects for core-course classmates.

The statistical analysis reported here is one component of a larger research project on peer effects in core courses. A parallel line of research involved interviewing 30 instructors in these courses to elicit their observations about peer effects in their classrooms. The results of the interviews provide a context to help us interpret our statistical results. Based on the

[1] See Manski (1993).

[2] An exception is García-Diez (2000), though the central focus in this paper is on the effects of a curriculum reform rather than peer effects.

comments of the instructors, we conclude that the most likely reason for the absence of statistical peer effects is that we are measuring the wrong peer characteristics. Our peer measures are all based on potential academic ability (as represented, for example, by SAT scores) whereas the instructors were nearly unanimous in believing that students' attitudes and personalities have greater influence on their peers than their raw ability.

Peer Effects

We often think about college choice in motivating why peer effects may be important. One consideration in deciding whether to attend X College, Y University, or Z Community College is the nature of the fellow students with whom you would share classes and perhaps dorms. If your peers affect your education, then such considerations may be important alongside factors such as faculty and curriculum.

We might call these "cross-institution" peer effects, since they compare the outcomes resulting from having one college's student body as peers with those of another college. While they are probably the most important kind of peer effects, cross-institution peer effects can only be estimated if one can distinguish the effects of a college's peers from the effects of the college's other characteristics: faculty, curriculum, facilities, etc. This is generally impossible.

As in the other studies cited above, we examine within-institution peer effects: the effects of interacting in a given institutional setting (class, dorm, etc.) with a particular set of students at a single college. Within-institution peer effects are more feasible to estimate because one need not control for differences across colleges.

Tests of peer effects usually involve three categories of variables: outcomes, controls, and peer variables. The outcomes are usually measures of academic success within the college: grades, graduation, or avoidance of academic disciplinary actions. Controls are non-peer variables that affect the chosen academic outcome. Controls include both the student's measured individual characteristics (SAT scores and the like) and characteristics of the student's environment in college (for example, the courses and instructors taken or the dorm in which he or she lives). Peer characteristics are measures of the characteristics of particular subsets of the general student body with whom the student has close peer interactions, such as roommates, dorm-mates, or classmates.

A typical estimating equation for peer effects thus takes the form

$$y_i = X_i\alpha + P_i\beta + u_i, \tag{1}$$

where y_i is the measured academic outcome for student i, X_i is a row vector of controls for the ith student and P_i is a row vector of variables describing student i's peers, and u_i is an error term.

The most common hypothesis is that some or all students benefit from having more able peers. This is tested by including one or more measures of peer ability in P and examining whether the corresponding elements of β are positive. Other peer-effect hypotheses are less common but can be examined within the framework of equation (1). For example, by including a measure of the dispersion of peer abilities in P one can test whether students gain or lose from having a wider range of abilities among peers. By including a measure of the number (or share) of peers who have ability levels similar to student i, one can test the "like-me" hypothesis that students benefit from having peers that have a similar ability level to their own.

This study uses classmates in core courses as the peer group. Our outcome measures are grades in courses other than the core course, which is omitted because variations in peer quality are likely to bias the relationship between learning and core-course grades by affecting the instructor's curve. We control for admission credentials of the student and for some non-peer elements of the student's core-course experience—in particular, for core-course instructor and year. Our various measures of peer characteristics attempt to capture relevant properties of the distribution of classmate abilities. The variables are described in more detail below.

Colleges and Courses

The three colleges in our study—Lewis & Clark College, Reed College, and Whitman College—are all selective liberal-arts colleges located in the Pacific Northwest. They range in size (in 2007–08) from 1,365 to 1,917 full-paying equivalent students and the average entering SAT composite scores (during our 1988–2002 sample) ranged between 1,240 and 1,330.[3] All three teach on a semester system.

While the three schools share geographical proximity and have numerous cross-applications, there are distinct differences in the cultures of their student bodies. Each admissions office and administration also records a somewhat different set of data on entering students. Thus, while we feel comfortable combining them into a single study, we cannot pool data from the three colleges into a single sample.

Each college teaches a common-syllabus core course that is required for graduation and taken by all or nearly all first-year entering students.[4] Table 1 shows the basic characteristics of the core courses. Although the topics vary somewhat, all three are discussion-based courses with small sections and a humanities orientation. All three are "gateway" courses that introduce the first-year student to academic practices such as careful reading, analytical thinking, class participation, and critical writing. These courses are important in teaching new students both the process of being a college student and the content of the courses themselves. Both process and content learning from the core courses is built upon in many subsequent courses in each college's curriculum, especially in the humanistic disciplines.

Because they occur at the beginning of the student's college career, experiences in the core course often play a key role as students define their academic identities. Peers in these courses provide models of good (or bad) academic behavior that may influence students' decisions about how much effort to put into assigned reading, how (and whether) to participate in class discussions, and how seriously to take class writing assignments. Once formed, all of these habits of academic behavior are likely to carry over into the student's other courses, both in the first year and beyond.

---

[3] Detailed summary statistics are available from the authors by request. At the beginning of the study we promised each school to avoid making direct quantitative comparisons between the schools, hence we do not present them in the paper.

[4] Many transfer students take the core courses, though some are not required to do so based on prior course work. In unusual circumstances a student may postpone the core course to the second year (or retake it after failing it in the first year). We exclude transfer students and students who take the core course after the first year from the estimating sample, but include them as peers of the freshmen in their class sections.

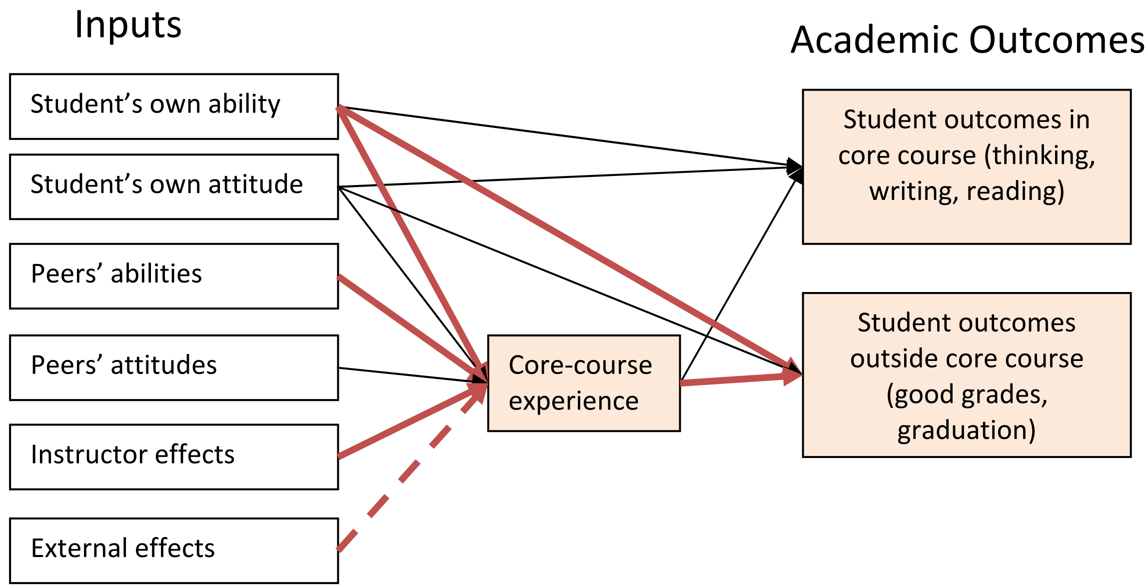

Figure 1. Schematic model of core-course peer effects

Figure 1 shows a simple schematic conceptualization of the role of core courses. The core-course experience consists of the discussions, reading and writing activities (and feedback on these from the instructor and/or peers), and other events that occur in relation to the core course. Any student's core-course experience depends not only on his or her own abilities and attitudes and the influences of the instructor, but also upon peers' abilities and attitudes and on external factors (sometimes called "common shocks") that influence the class.[5] Because of the centrality of these courses in the colleges' curricula, the core-course experience is likely to affect the student's academic achievement not only in the core course, but in other courses as well.

The heavy red arrows in Figure 1 show the channels through which the effects we measure are assumed to operate. On the input side, we have (noisy) measures of the "abilities" of both the student and his or her peers from their admission files. We use dummy

---

[5] Among the infinite variety of external factors that could be important, instructors we interviewed specifically identified the time of the class, the setup of the classroom (conference table vs. chairs in a circle vs. stadium seating), and external events that occurred during the semester, such as serious illness of a classmate or a major disruption on campus.

variables to capture the effect of each instructor in our samples and year dummies to capture some of the external effects.[6]

On the outcome side, it seems likely that the core-course experience would have its strongest effect on the grade in the core course itself. However, core-course grades at all three schools are assigned by the individual instructors; we have no independent measure of performance that is assigned on a scale common to all class sections and instructors. Many instructors are assumed to assign grades at least partially on a "curve," so the quality of a student's peers could affect his or her grade (perversely) through this "curve effect" apart from any peer effects on student learning. To overcome the curve effect, we restrict our outcome measures to grades in courses other than the core course, although the effect of the core-course experience on these grades, and hence our measurement of peer effects, is likely to be diluted.

We have no direct measure of the core-course experience itself.[7] Thus, our estimated models link non-core academic outcomes directly to the student's own quality measures, the quality measures of the student's core-course peers, and the dummy variables for instructor and other measurable core-course influences. For peer effects to occur in our model, not only must peer abilities affect a student's core-course experience, but that experience must influence the student's grade performance in other classes.

It is noteworthy that we have no measure of the attitude, personality, or other characteristics of either students or peers. When we interviewed core-course instructors at all three schools, we found a strong consensus that peers' attitudes toward the course and peers' classroom behavior—often reflecting their personalities—were very important. We must be very clear that the peer effects we measure are based solely on potential academic ability as

[6] We tested time and location dummies, but found none to be statistically significant. There are surely many unmeasured and unmeasurable external effects that end up in the statistical error term.
[7] One could imagine using student course evaluations as a rough measure of **Error! Reference source not found.**'s central box. However, these evaluations are aimed more at instructor quality than course (and specifically peer) quality, and are moreover unavailable to us.

measured by admission credentials. Any peer effects occurring through the influence of students' personalities or attitudes will not be captured by our coefficients.[8]

A final aspect of these core courses is highly relevant for our statistical work. As noted above, reliable identification of peer effects is facilitated by either random selection of peers or selection based on observable criteria. Nearly all first-year students take these courses, so the choice of peers involves assignment to course sections rather than choosing whether or not to take the course.[9]

Although the exact criteria for section assignment vary across schools, assignment is largely random. Registration occurs either during the summer before arriving on campus or shortly after arrival, so it is unlikely that self-selected groups of peers (e.g., friends) would choose a common section even when students choose their sections. Thus, any self-selection of peers would be the tendency of like individuals to choose a section meeting at a specific time or taught by a specific instructor.

At Whitman College, all core sections meet at a common time and assignments are done by the registrar using a random-number generator to assign sections with uniform numbers of males and females. At Reed College, students register for a "conference" section meeting at a time that fits their schedules. Sections are listed without instructors in the class schedule, so there is no basis for discriminating among multiple sections meeting at the same time. The registrar then reassigns students to sections as necessary to achieve a relatively similar male/female ratio. At Lewis & Clark College, sections are listed with instructors and students may submit a list of four preferred instructors. They are usually, but not always, assigned to a section taught by one of their four choices. Since we have dummy variables for instructor and time of meeting, we are able to control for the most obvious criteria on which such implicit self-selection of peers would occur.

---

[8] To the extent that this paper serves to test the rationale for ability-based merit aid, this shortcoming is minor. The vast majority of merit awards seem to be based on precisely the admission credentials that we use to measure peer quality, so the kind of peer effects used to justify merit aid should show up in our estimates, as long as they occur in the core courses and affect performance in other courses.

[9] The only "regular" exception of which we are aware is that at one school students with exceptional preparation in chemistry (usually at most one per year) may enroll in an advanced chemistry course that creates a time conflict with the core course.

Variables and Data Issues

Outcome measures

We examine five grade-point average measures, all of which exclude the core course to avoid confounding peer effects with the effects of the instructor's grading curve.[10] First we use overall cumulative GPA, which includes all courses (except core) taken at the college.[11] Next, since the effects of the core-course experience may be more important during the year in which the core course is taken and the year after, we examine first-year and first-and-second-year GPA.[12]

Finally, the core-course experience may have a stronger impact on courses in related subjects or that use related teaching methods than in other courses. We use a narrow and a broad version of a core-related GPA to measure success in these courses. The narrow measure consists of courses in fields directly related to the disciplines in the core course.[13] The broader measure includes all courses except those in the natural sciences and mathematics.

Student characteristics

The academic ability of the individual student is expected to be the most important influence on grades. We control for the student's own academic ability through the use of several sets of characteristics. First, we have standard measures of academic quality from the student's pre-college record: SAT verbal and math scores, high-school grade-point average,

---

[10] Note, however, that the "predicted GPA" variable that we use as a measure of incoming student ability is a prediction of total, cumulative GPA including core.

[11] All GPA measures exclude transfer work and work taken on overseas programs not taught by regular college faculty. Courses taken on a pass-fail basis and courses from which a student withdrew are also excluded. All courses are evaluated on a 4.0 scale with 0.3 added (deducted) for a plus (minus) grade, so an A– counts as 3.7 and a B+ as 3.3.

[12] The first year is taken to be the fall semester in which the core course is begun and the following spring semester. The second year comprises the fall and spring semesters of the next academic year. For a student who took a year off between freshman and sophomore year, only the freshman year would be counted in the first and second years. We chose this definition to maintain temporal proximity of the courses whose grades are measured to the core course.

[13] We solicited the opinions of the core-course chairs at each campus in making these judgments. The courses deemed to be closely related were largely courses in history, literature, philosophy, religion, the arts, and related fields.

class rank (as a percentile) in high school, and ratings assigned by the schools' admission offices.[14]

Since all of these variables are measures of the student's academic ability, we expected to find extreme collinearity among them. However, as shown in Table 2, only high-school GPA and high-school rank are highly correlated. To reduce the threat of multicollinearity and to facilitate interpretation of the coefficients, the Reed and Lewis & Clark admission ratings are entered in the equation as residuals from a regression of admission rating on other student characteristics. Since admission ratings at these schools are affected by such variables as SAT scores, it is difficult to interpret the coefficients on SAT scores in a regression that includes the admission rating. Such a coefficient would measure the effect of an increase in SAT holding the admission rating constant (presumably by lowering some other variable), which is a highly unnatural experiment. By replacing the admission rating with the residual, which is by construction uncorrelated to SAT scores, the SAT coefficients have the natural interpretation and the coefficient on the admission-rating residual can be interpreted as measuring the effects of the aspects of the admission rating that are unrelated to SAT scores and other measured characteristics—aspects of the student's application such as essays, interviews, quality of high school, and extracurricular activities.

Other student variables for which we control are self-reported ethnicity, sex, status as an international student (a dummy based on citizenship), need-based and non-need-based financial-aid status, athlete status (at Lewis & Clark), and type of high school attended (public vs. parochial, vs. non-religious private).

Peer variables

The ideal peer variable would both measure the characteristics of individual peers that are relevant and would measure the relevant characteristics of the *distribution* of these relevant

[14] The admission office ratings vary by school. Reed and Lewis & Clark assign a comprehensive rating summarizing their assessment of the student's admissibility. Whitman has an "academic rating" that is a linear function of SAT scores and high-school GPA and an "affinity rating" that measures other factors in the application, such as essays, recommendations, and extracurricular activities. Due to changes in admission office practice and conversion of the college database system, this affinity rating is available on a consistent basis only after 1999.

characteristics. More specifically, focusing on peer ability, we need to measure each peer's ability appropriately and also find the properties of the distribution of abilities (mean, standard deviation, number of excellent students, etc.) that most strongly influences each student's core-course experience. Thus, definition of the peer variable implies both a choice of the individual-level measure of ability and the class-level distribution characteristics.

Most higher-education peer-effects studies have used either SAT scores or admission rating as the measure of peer ability. We utilize these measures below, but our preferred specification incorporates all relevant and available student-level information into a predicted cumulative GPA variable.[15] This is constructed as predicted values from a "student-achievement model" (SAM) regression of cumulative GPA on individual level variables, including admission credentials and demographic dummies.

We also use core-course instructor dummies, year dummies, and dummies for the student's area of major in the SAM to capture instructor effects in the core course, changes in grading or admission standards over time, and differences in GPA that arise due to differences in grading practices in the sets of courses taken by students with different majors.[16] These variables have statistically significant effects on cumulative GPA but they occur because of factors occurring after arrival at college; they are not measures of the ability that a student brings to college. To get a measure of student ability that puts all students on a level basis, we zero out these variables when calculating predicted GPA. This yields a prediction of the GPA the student *would have earned* if he or she had come in the reference year, taken core from the reference instructor, and chosen the reference major (where "reference" refers to the category omitted from the set of dummy variables).

[15] As discussed elsewhere, Whitman lacks a consistent and comprehensive admission rating and Lewis & Clark lacks SAT scores for a large and self-selected share of the student population. This makes it impossible to use either SAT or admission rating for all three schools, though both can be used for Reed.

[16] Students majoring in the sciences and mathematics earn the lowest grade-point averages (controlling for admission credentials) among declared majors at all three schools. Arts and humanities majors have the highest GPAs with social and behavioral science majors in between. Students who leave the institutions without declaring a major, not surprisingly, have much lower GPAs than those declaring a major.

Missing-data problems

College databases (and indeed colleges themselves) are not operated for the convenience of subsequent analysis by economists. One manifestation of this is the incomplete information available on many students. In some cases this occurs due to external constraints such as high schools that do not report their students' class ranks. Sometimes it results from institutional policies such as making submission of admission credentials such as SAT scores optional. In other cases it happens when students and applicants fail to answer questions about their ethnicity.

Approximately 70 percent of the students in our sample have complete data for the variables in which we are interested. Missing-data problems are not uncommon in economics and in many applications we can simply ignore the missing cases and estimate our regressions using the majority of the sample that is complete, provided that the selection of available cases did not introduce bias. The missing-data problem is more severe for us, however, because our goal is to characterize the distribution of student abilities in an entire class. If we have data on all but one or two of the students in a class section, our picture of the class distribution of abilities is likely to be quite accurate. But with one-third of the average class section missing, the description of the class ability distribution based solely on complete cases would be largely useless. In order to obtain a predicted GPA value for all or nearly all students in each class, we must employ one or more of several methods for estimating models with missing cases.

Statistical Models with Missing Data

Although not common in econometrics, missing-data models are frequently employed in epidemiological and other statistical applications.[17] The choice of model depends on the nature of the missing cases. In the simplest case, the data are "missing at random," meaning that the probability of a case being missing for a particular variable is independent of the

[17] An survey of methods is provided by Little (1992). Econometric applications are discussed by Brownstone and Valletta (2001).

value of that variable.[18] In the more difficult case, data are not missing at random, which leads to selection bias if not handled properly. While most of our missing variables can safely be assumed to be missing at random—high schools either report class rank or they do not, but we assume that they do not report it for some students and not for others—an important exception is missing SAT scores resulting from the SAT-optional admission policy at Lewis & Clark. In that case, it is very likely that the students whose SAT scores are missing would have lower scores (all else equal) than those who submit the scores. This is a perfect example of data that are not missing at random.

The missing-data literature is heavily focused on estimation of coefficients, but in our case the crucial application of missing-data methods is prediction.[19] We use two methods for estimation and prediction with missing data. To facilitate the explanation, assume that we have two regressors, one of which is missing for a subset of the observations.

The first method, which works effectively even when data are not missing at random, is to essentially split the sample into two sub-samples: one with complete data and the other comprising the observations for which one regressor is missing. We use the sub-sample method for the Lewis & Clark sample to overcome the problem of missing SAT scores. We essentially treat the students who submitted SAT scores as being drawn from a different population than those who did not. Conditional on SAT scores being present or absent, the prediction of the relevant regression model is efficient.

The sub-sample model is convenient when missingness is confined to one or two variables and the number of observations in each sub-sample is large. When the number of partially missing variables is larger and the sub-samples with a particular pattern of missingness are small, then this method becomes unwieldy.

An alternative that can be used when data are missing at random is multiple stochastic imputation, which is described in detail by Rubin (1996) and implemented in the Stata procedures **ice**, which calculates imputed values for missing cases, and **mim** (formerly

[18] A stronger condition is "missing completely at random," in which the probability that a variable is missing for a variable is independent of *all* variables. For the methods we use, the distinction between missing at random and missing completely at random is unimportant.

[19] Issues of prediction in missing-data models are explored in a Monte Carlo analysis by Gwati (2007).

micombine, which performs regressions on the imputed samples and combines them) by Patrick Royston (2005). Multiple imputation consists of estimating the conditional probability distribution for each missing variable conditional on the other variables of the model, then taking *m* (ten in our case) random draws from this conditional distribution. This leads to *m* imputed samples in which missing data points are replaced with imputed values.

The desired regression is run independently for each of the *m* imputed samples. The estimates are then combined with the estimated coefficient variances taking account not only of the estimated variances of each imputed sample but also of the variation in estimated coefficients across samples.

Our main interest is in prediction—construction of a predicted GPA for peers with incomplete data. Since there will be *m* different imputations of the missing variables, there will be *m* distinct predicted GPA values for students with missing data. This leads to *m* different values for each of the peer variables. We carry the multiple imputations through to the stage of estimating peer-effects equation (1), estimating *m* separate regressions for the imputed samples, then combining them to form the final estimates.[20]

Predicting GPA

Because multiple-imputation models are not standard econometric tools, we wanted to assure that using imputed observations did not radically alter the basic regression results. Table 3 shows the estimated coefficients and standard errors for admission and demographic variables from the student-achievement models. The corresponding multiple-imputation estimates are shown in Table 4. In both tables, the coefficients on admission variables are allowed to be different for the sub-sample blocks that are missing SAT for Lewis & Clark and

[20] We explored two methods of estimating the final peer-effects equation: using both complete and imputed cases and using only complete cases. Even using complete cases, the final equation must be estimated *m* times and combined because even students with complete data will have *m* different imputations of the peer variables (assuming that at least one classmate had missing data). The results of the two methods were very similar and the choice between them has no effect on our qualitative conclusions.

affinity rating for Whitman.[21] Although the coefficient estimates of the SAM vary somewhat between the complete-case and multiple-imputation models, the qualitative results are very similar. We explain almost 40 percent of the variance in college GPA at Lewis & Clark and Whitman and about 25 percent at Reed.[22]

There is no particular reason to expect that the relationship between GPA and admission variables such as SAT scores should be linear.[23] We use polynomial specifications for SAT scores, high-school GPA, and high-school class percentile when higher-order terms are statistically significant. In all cases, the bivariate relationship slopes upward except occasionally at the very tails of the sample distribution.

As expected, higher SAT scores, high-school GPA, and high-school rank are all associated with higher college GPA. Members of some minority groups perform worse than would be expected based on their admission credentials whereas females and international students perform somewhat better. Year dummies are statistically significant at all three schools and core-instructor dummies have strong effects at Lewis & Clark and Reed.

The multiple-imputation SAM estimates of Table 4 are the basis for the predicted GPA variable that is our preferred measure of student quality. As discussed above, the estimated equations include controls for effects on GPA occurring after arrival at college such as core-course instructor and choice of major. In forming the predicted GPA variable, these variables are set to zero.

---

[21] This avoids what is essentially an omitted-variable bias on coefficients of variables that are correlated with the variable that is omitted from part of the sample. See Jones (1996).

[22] There are several idiosyncrasies of Reed that could explain the lower R-square. Reed students are not told their grades either during or after their courses (although they can obtain them by asking), which tends to de-emphasize grades in the student culture. In addition, bad grades at Reed are more commonly associated with low effort or external influences than with low ability.

[23] Indeed, it is not clear that SAT scores are really an "interval" variable in the sense that the difference between scores of 750 and 700 is quantitatively the same as the difference between 650 and 600.

Tests for Peer Effects

Are core-course classmates' academic outcomes correlated?

A basic test for peer effects is to examine whether being in the core class together leads to positive correlation between the outcomes (other grades) of a pair of students. If we find correlated outcomes among classmates, the cause could be either peer effects or exogenous effects experienced in common by all members of the core class—what Lyle (2007) calls "common shocks."

Table 5 shows the probability values associated with standard analysis-of-variance F tests of whether the core section affects various GPA outcome measures. The test statistic indicates the significance of the correlation in grade outcomes among members of core-course sections. We find statistically significant correlation for all measures at Whitman College and for first-year grades at each school. This suggests that *something* about the core-course experience matters for performance in *some* non-core courses.

However, a positive within-section correlation does not necessarily require the presence of peer effects, since this could result from the effects of the instructor, the year, or common external shocks or from systematic difference across sections in student abilities. We can control for the effects of instructor, year, and student abilities by regressing the outcome measures on student characteristics, instructor dummies, and year dummies, and then testing whether dummy variables for core section have marginally significant effects.

The results of F tests on core-section dummies are shown in Table 6. One of 15 F statistics is significant at the 0.10 level, which is below the expectation of $0.10 \times 15 = 1.5$ significant values through random chance. Thus, we find no evidence of correlation among classmates' academic outcomes when controlling for the effects of students' own abilities and core-course instructors.

Do academically stronger peers improve classmates' outcomes?

The most common peer-effects hypothesis is that students gain by having more able classmates. Academically excellent classmates elevate the quality of the class discussion and provide positive academic role models. Other effects may occur as well, which could reinforce or mitigate the positive influence of able peers. Strong students might spur one another to greater achievement if they kindle healthy competition. However, they could intimidate weaker students or, by dominating class "air time," discourage them from getting fully involved in the class.

We test for the effect of average core-course classmate quality by estimating an equation similar to the student-achievement model of Table 4 with a measure of average peer ability included.[24] For comparability with other students, we use the peer-mean composite SAT score (for Reed and Whitman only) and the peer-mean admission rating (for Lewis & Clark and Reed only) as measures of peer ability, alongside our preferred specification: the peer-mean predicted GPA.

Table 7 shows the estimated effects of peer-mean ability measured by SAT scores and admission ratings. No coefficient in the table approaches conventional levels of statistical significance. The results are similar when peer-mean predicted GPA is used to measure peer ability, as shown in Table 8.

Earlier work with a shorter sample for Reed College suggested that both peer mean and class standard deviation of predicted GPA matter, but with a highly significant cross-product term that allows the effect of mean to depend on the standard deviation and vice versa.[25] Both mean and standard deviation had positive coefficients with the cross-product term being negative. This result means that students in high-mean-low-variance and low-mean-high-variance sections overachieve while those in high-mean-high-variance and low-mean-low-variance sections underachieve.

Table 9 shows that this specification is still supported for Reed, but is barely statistically significant for only two of the three outcome measures shown. The other colleges show no support for this result—in fact, the signs are opposite in every case and are significant at the 10 percent level for first-year GPA at Lewis & Clark.

---

[24] The other difference between the student-achievement model and the peer-effects regressions is the dependent variable. The basic SAM uses total cumulative GPA whereas the GPA measures in the peer-effects regressions exclude the core-course grade.

[25] See Hoel, Parker, and Rivenburg (2006).

The evidence presented from our mean-ability-based regressions offers no support whatever for the hypothesis that students in core courses benefit from more able peers. We consider possible interpretations of this result in the next section of the paper. First, we use our data to examine one final peer-effects hypothesis.

Do students gain from having classmates like themselves?

Mean and standard deviation are not the only measures of the ability distribution of classmates. Since these measures are sensitive to outliers in the distribution, we examined a non-parametric measure as well. Table 10 reports the estimated coefficients on the share of students in the core-class section who are in each quartile of the overall student distribution. Because the shares in all quartiles add to one, the share in the bottom quartile is omitted. The dependent variable in all regressions is GPA in all courses except core.

The rows headed "All" in Table 10 are full-sample regressions. The coefficients on the three quartile shares never approach statistical significance.

The remaining rows divide the sample by quartile, performing separate regressions for the sub-sample of students in each quartile. This allows us to examine the hypothesis that students gain from having more students in the core-course section who have ability levels near their own. If that hypothesis were true, then the coefficient on the share of top-quartile students should be positive for the sub-sample of students in the top quartile, the share of the medium-high quartile should have a positive effect for the medium-high sub-sample, and so on. There is no indication of such "like me" peer effects in Table 10. The "diagonal" elements of the table are no more positive than the others and never approach statistical significance.

Discussion

The statistical results reported in the previous section provide no empirical support for the presence of peer effects in core courses at Lewis & Clark, Reed, and Whitman Colleges. Because there are several possible explanations for this lack of statistically measurable effects, alternative kinds of evidence are useful in interpreting the results and drawing conclusions.

As a companion study to the statistical analysis above, members of our project team talked with about ten faculty members at each institution who have taught the core courses frequently. Core-course instructors, of course, have their own biases—for example, they probably believe strongly in the importance of the core course—and are not always privy to students' influences on one another. Nonetheless, the ability to combine statistical analysis with direct observation provides a unique opportunity to assess peer effects from multiple perspectives.

We asked instructors for their observations of peer effects in a loosely structured one-hour interview. Among the specific topics we discussed were:

- Characterization of good (and bad) peers and how they aided (hindered) their classmates' learning. (Focus on individual students.)
- Composition of successful (and unsuccessful) core-class sections: What makes the good class work and what inhibits the weak class. (Focus on collective characteristics of classes.)
- Roles of ability, effort, attitude, and personality in peer influence.

About a year after the interviews, we brought interviewees together (along with core-course instructors from other colleges) at a conference for follow-up discussions and to review our conclusions from the interviews. While anecdotal in nature, this large body of evidence provides us with a rich background of insider detail with which to interpret our statistical results. On many fundamental questions there was near-universal agreement among the instructors, lending confidence to the generality of their observations.

Why don't we find statistical evidence of peer effects?

Our simple representation in Figure 1 suggests that there are two possible reasons for the lack of statistically measurable effects. The peer effects we attempt to measure run through the red arrows from peers' abilities to the core-course experience and on through the red arrows to student outcomes outside the core course. If the effect represented by either of these arrows is weak or non-existent, then our statistical analysis would reveal no effect: if peers' abilities do not affect the core-course experience or if the core-course experience has no effect on grades in other courses.

Our interviews give us an instructor's-eye view of these links. The interviewees believe, perhaps not surprisingly, that the experience that students have in the core courses affects their academic success in other classes. They emphasized that these courses teach important learning skills as well as academic content. Students whose experience in the core course is successful will learn skills of critical reading, discussion, and writing that apply to a broad collection of courses across the curriculum.

In terms of peer effects, the instructors believed that students have very strong effects on the quality of discussions in the core course. A student with a class full of good peers has the opportunity to participate in enlightening classroom interaction that one with a less effective group of peers would miss. However, two points raised by the interviewees may help explain the lack of strong statistical effects.

First, and most important, the instructors we interviewed stressed that the most important characteristics of a good peer depended more on the student's attitude and personality than on pure academic ability. A good class needed a sufficient number of students who were engaged with and well prepared for the class, mature and intellectually curious, respectful of their peers' contributions, and willing to speak in class. This reinforces the finding of Stinebrickner and Stinebrickner (2008) that roommates affect student success primarily through their effect on time spent studying.

After hearing their descriptions of good-peer characteristics, we probed the connection between these characteristics and academic ability as measured by traditional admission measures. According to the instructors, outstanding peers who elevated the quality of the class were never low-ability students, but neither were they always high-ability students. We heard many stories about brilliant students who elevated the quality of the class, but equally many about students of average ability whose personality and attitude improved the course.

To an even greater degree, detrimental peer behavior depended on a student's attitude rather than ability. Low-ability students are not necessarily destructive peers and can be a useful part of the classroom mix if they participate with a positive attitude. In fact, many of the "problem cases" that our instructors reported were brilliant students who behaved in ways that detracted from their peers' learning: displaying a dismissive attitude, making comments to "score points" rather than further discussion, or belittling comments made by other students. Thus, even if peer effects in core courses are important, we may not have measured the characteristics of peers that are most important to classmates' learning.

With respect to the importance of core-course peers for measured outcomes such as grades, some interviewees drew an important distinction between the discussion environment of the core course, where peer influence is strong, and the writing component that weighs most heavily in grading. They argued that although peer interaction is crucial in discussion, classmates may have a limited impact on the quality of a student's writing.[26] This might limit the strength of peers' influence on grades in the core course, and also on grades in subsequent courses in which class discussion is not heavily weighted.

Another explanation for our finding is that student quality does not vary greatly within a selective liberal-arts college. Compared with a large public university or a non-selective college, our students all tend to have good scholarly credentials. Instructors may have found attitude and personality to be the important peer characteristics precisely because nearly everyone in their classes was sufficiently *able* to contribute positively. Not only is the range of abilities of individual students somewhat limited, but the core-section averages vary even less. For Reed College, the standard deviation of undergraduate GPA is 0.61, the standard deviation of predicted GPA is 0.26, and the standard deviation of peer average predicted GPA is only 0.07.

If low variation in peer-mean abilities explains our empirical results, then it may be important not to generalize them to situations in which variation is greater. In particular, concluding that relative core-course classmate abilities within an institution like Reed College do not seem to affect student outcomes does not imply that we could replace a student's Reed classmates with a group of low-ability community-college students without affecting the student's learning.

[26] A notable exception is that some core-course instructors assign students to do "peer-editing" of writing assignments. In these situations, students often improve their writing by reading others' assignments and by receiving comments from those who read their own.

Measurement of inter-institutional peer effects is difficult—likely impossible. Absent a controlled experiment in which a group of students is transplanted from one college to another with instructors and other institutional characteristics held constant, it is impossible to distinguish the effects of having, say, "Whittie" classmates from the effects of having a Whitman College instructor and the Whitman College curriculum.

Indeed, one can argue that the greatest peer effects across institutions occur in the design of the curriculum: classes can be more challenging at a college where bright and hard-working students are the norm. One gains from attending a highly selective college not only because of the explicit interaction with other bright students, but also (and perhaps to an even greater extent) because joining a highly able peer group usually means taking classes that tackle more difficult material and have higher expectations of student achievement. Anecdotal evidence from students at our schools who take classes at less-selective state universities suggests that they are often disappointed in two aspects of their transferred classes: the low quality of peers and the low quality of course content.

Finally, when the instructors convened together to discuss peer effects, we asked them to comment on the degree to which our conclusions were specific to core courses or generalizable to other courses as well. There was a general recognition that peer effects would differ depending on the nature of the course. Discussion-based peer effects of the kind emphasized in core courses might be quite different from the kinds of peer effects we would see in a mathematics (or economics) course where students work together on problem sets or in a science course where they perform lab activities and write lab reports together.

Conclusion

Our search for classmate peer effects in core courses has found no evidence that the ability distribution of their peers influences students' success in other classes. Based on interviews with instructors in these courses, we believe that the primary reason for the absence of peer effects is that the most relevant peer characteristics are based not on potential academic aptitude but on attitude and personality.

However, even if personality-based peer effects are very important, the results of this study still may provide evidence against the policy of awarding merit aid to students based on their contribution to the education of their peers. Colleges that give merit aid seem to base the awards entirely on perceived academic aptitude. The results of our study provide no support for using peer effects as justification for such awards.[27]

De Long and Lang (1992), in a paper whose title is only somewhat whimsical, suggest that the only econometric studies that get published despite failing to reject the null hypothesis are those in which the null hypothesis is widely believed to be false. This paper adds to a long list of econometric studies—many of them published in prestigious journals—that find weak or non-existent peer effects in higher education. That we continue to hunt for empirically elusive peer effects probably reflects strong prior beliefs that such effects exist, and we propose several further steps that may prove more fruitful in finding them.

First, given the testimony of the core-course instructors that we interviewed, it may be beneficial to attempt to quantify some of the personality and attitude characteristics of peers. This could be done by surveys of entering students, either using instruments specially designed for this purpose or perhaps extracting relevant information from the freshman survey designed by the Cooperative Institutional Research Program (CIRP) and administered by many colleges.

The peer effects estimated in the present study combine two separate effects, either of which may be weak: the effect of peers on the core-course experience and the effect of that experience on grades in subsequent courses. For purposes of understanding peer effect in core courses, it would be useful to attempt to measure the quality of the core-course experience directly—quantifying the central box in Figure 1. This would allow direct and separate examination of both the left-hand arrows linking peer characteristics to core-course effectiveness and the right-hand arrow representing the influence of the core course on subsequent academic work.

---

[27] However, as noted above, the absence of within-institution peer effects does not necessarily imply that there are no peer effects operating across institutions through channels such as difference in the quality of the curriculum that are driven by differences in average student ability.

Another way that the peer-effects link could be decoupled from the course-to-course influence link would be to use grades in the core-course itself as an outcome measure. We have avoided that in the present study because of concerns about the "curve effect" of classmate peers. It is likely that veteran instructors are able to identify strong and weak cohorts—our interviews clearly reveal that they *think* they can—and thus are able to average their grades over many cohorts of students. We plan to look at core-course grades as an outcome measure using a sample restricted to sections taught by professors who have taught the core course frequently. These "within-course" peer effects should be stronger than the "across-course" effects we examine in this paper.

The ideal test for within-course peer effects is a setting with both random assignment of students to sections and a common grading scale across sections. In related research, we are examining one such course at Reed College in which three sections of introductory political science students each fall semester rotated among the same trio of instructors and where all three sections were graded jointly on a common scale.

Finally, it may be very productive to approach research on peer effects from an interdisciplinary perspective. Educational psychologists have studied peers and learning processes in children, but extensive searches of psychology and education indexes failed to reveal a literature on college-level peer effects. Combining detailed psychological analysis—such as measurement of personality attributes—with the statistical methods applied by economists may yield fruitful results.

Tables

Table 1. Descriptions of core courses

	Lewis & Clark	Reed	Whitman
Title of course	Inventing America	Classical Humanities	Antiquity and Modernity
Duration and syllabus	Two semester courses: Fall is common syllabus; students choose different topic-based sections in spring	Year-long course; few students change sections	Year-long course; very few students change sections
Student credit	4 semester hours each semester	6 semester hours (1.5 Reed units) for each semester	4 semester hours each semester
Contact hours per week	3 hours of discussion; occasional optional lectures	3 hours of discussion; 3 hours of required lecture	3 hours of discussion; occasional optional lectures
Content	Fall: Classical to modern, emphasizing political philosophy relating to founding of U. S. and to groups left out of founding process	Classical philosophy, history, and literature; Greece in fall and Rome in spring	Philosophy, history, and literature, focusing on classical in the fall and early modern/modern in the spring
Section size	Discussion sections of 18–19 (slightly larger in early years).	Discussion sections of 14–17.	Discussion sections of 16–18.
Scheduling of sections	Most at common afternoon time. A few at morning time for in-season athletes and students with afternoon labs.	Sections are offered at many times and days.	All sections meet at common morning time.
Section choice	Students rank-order four preferred sections/instructors.	Students register for section/time without knowing instructor.	Students are randomly assigned to sections by registrar.
Criteria for evaluation	Writing, class participation, and exams.	Writing, class participation, and exams. Single grade for full year.	Writing, oral reports, discussion participation, exams. Semesters graded separately.

Table 2. Correlations among admission variables

Lewis & Clark	SAT math	SAT verbal	High-school GPA	High-school percentile	Admission rating
SAT math	1.000				
SAT verbal	0.400	1.000			
High-school GPA	0.292	0.274	1.000		
High-school percentile	0.274	0.225	0.833	1.000	
Admission rating	0.524	0.477	0.663	0.642	1.000

Reed	SAT math	SAT verbal	High-school GPA	High-school percentile	Admission rating
SAT math	1.000				
SAT verbal	0.349	1.000			
High-school GPA	0.214	0.143	1.000		
High-school percentile	0.151	0.092	0.706	1.000	
Admission rating	0.310	0.306	0.518	0.447	1.000

Whitman	SAT math	SAT verbal	High-school GPA	High-school percentile	Affinity rating
SAT math	1.000				
SAT verbal	0.432	1.000			
High-school GPA	0.303	0.259	1.000		
High-school percentile	0.263	0.184	0.865	1.000	
Affinity rating	–0.024	0.071	0.047	0.050	1.000

Table 3. Complete-case student achievement model (SAM) estimates

	Lewis & Clark		Reed	Whitman	
	With SAT	No SAT	Full sample	With affinity	No affinity
Admission variables					
SAT math / 100	0.0425** (0.0206)		-0.618*** (0.191)	0.0608*** (0.0197)	0.0368** (0.0158)
(SAT math / 100) ^ 2			0.0532*** (0.0148)		
SAT verbal / 100	0.086*** (0.018)		0.461** (0.224)	0.106*** (0.0173)	0.0702*** (0.0152)
(SAT verbal / 100) ^ 2			-0.0283* (0.0167)		
High-school GPA	0.347*** (0.063)	0.493*** (0.090)	0.396*** (0.0426)	13.93*** (5.399)	11.52** (5.743)
(High-school GPA) ^ 2				-4.860*** (1.553)	-3.539** (1.715)
(High-school GPA) ^ 3				0.553*** (0.151)	0.381** (0.169)
High-school percentile / 100	-0.970* (0.570)	-0.326 (0.676)	0.385*** (0.113)	0.000641 (0.00193)	-0.000476 (0.00150)
(High-school prctile / 100) ^ 2	1.064*** (0.374)	0.409 (0.447)			
Admission rating residual	0.0152* (0.0092)	0.0365*** (0.0127)	0.266*** (0.0254)		
Affinity rating				0.0321*** (0.0116)	
Demographic variables					
Female	0.126*** (0.023)	0.089** (0.045)	0.0719*** (0.0218)	0.104*** (0.0229)	0.0538*** (0.0202)
Foreign student	0.311** (0.132)	0.130 (0.269)	0.0857 (0.0584)	0.0650 (0.178)	0.319** (0.136)
Black	-0.344*** (0.088)	-0.177* (0.099)	-0.459*** (0.167)	0.145 (0.115)	-0.00754 (0.0837)
Asian/Pacific Islander	-0.160** (0.066)	-0.339*** (0.115)	-0.149*** (0.0511)	-0.120*** (0.0410)	-0.0246 (0.0364)
Hispanic	-0.198** (0.084)	-0.231*** (0.082)	-0.143** (0.0679)	-0.267*** (0.0638)	0.0394 (0.0618)
Native American	-0.091 (0.088)	-0.799*** (0.249)	-0.097 (0.0990)	-0.116 (0.109)	-0.0675 (0.0958)
Regression statistics					
Number of observations	1823	776	2707	1148	1441
R-square	0.400		0.255	0.402	
Year dummies (p value)	0.000		0.033	0.024	
Faculty dummies (p value)	0.007		0.000	0.419	

Regressions also include financial-aid and high-school type dummies and dummies for core-course instructor, entry year, and area of major. Regressions of sub-samples were done jointly constraining dummies to be the same for both sub-samples. Standard errors are shown in parenthesis below coefficients. Statistical significance at the 10% (5%, 1%) level is shown by * (**, ***).

Table 4. Multiple-imputation student-achievement model (SAM) estimates

	Lewis & Clark		Reed	Whitman	
	With SAT	No SAT	Full sample	With affinity	No affinity
Admission variables					
SAT math / 100	0.0261 (0.0171)		-0.253 (0.155)	0.0442*** (0.0157)	0.0349** (0.0157)
(SAT math / 100) ^ 2			0.0268** (0.0121)		
SAT verbal / 100	0.0875*** (0.0154)		0.301 (0.185)	0.0884*** (0.0143)	0.0745*** (0.0142)
(SAT verbal / 100) ^ 2			-0.0183 (0.0138)		
High-school GPA	0.314*** (0.0567)	0.446*** (0.0730)	0.319*** (0.0422)	18.62*** (4.261)	19.64*** (4.412)
(High-school GPA) ^ 2				-5.361*** (1.247)	-5.952*** (1.324)
(High-school GPA) ^ 3				0.524*** (0.122)	0.611*** (0.131)
High-school percentile / 100	-0.762* (0.433)	-0.341 (0.490)	0.464*** (0.123)	0.00607*** (0.00123)	0.00535*** (0.00128)
(High-school pcntile / 100) ^2	0.877*** (0.289)	0.357 (0.342)			
Admission rating residual	0.0198** (0.00767)	0.0346*** (0.0108)	0.271*** (0.0213)		
Affinity rating				0.0427*** (0.00961)	
Demographic, aid, high-school type					
Female	0.171*** (0.0200)	0.116*** (0.0359)	0.0765*** (0.0176)	0.107*** (0.0187)	0.0883*** (0.0187)
Foreign student	0.038 (0.102)	0.00121 (0.0993)	0.0979** (0.0412)	0.0839 (0.0789)	0.199** (0.0793)
Black	-0.367*** (0.0768)	-0.217* (0.111)	-0.296** (0.110)	-0.0496 (0.0784)	-0.138* (0.0770)
Asian/Pacific Islander	-0.141*** (0.0517)	-0.217*** (0.0780)	-0.0759* (0.0417)	-0.125*** (0.0337)	-0.0321 (0.0337)
Hispanic	-0.143** (0.0647)	-0.182** (0.0796)	-0.0914* (0.0531)	-0.264*** (0.0520)	0.0114 (0.0564)
Native American	-0.0574 (0.0769)	-0.531** (0.228)	-0.102 (0.0860)	-0.226** (0.0980)	-0.0655 (0.0854)
Regression statistics					
Number of observations	2634	1138	4452	1800	1788
R-square (average over 10 imputations)	0.377		0.237	0.394	
Year dummies (p value)	0.026		0.010	0.013	
Faculty dummies (p value)	0.295		0.003	0.384	

Regressions also include financial-aid and high-school type dummies, and dummies for core-course instructor, entry year, and area of major. Regressions of sub-samples were done jointly constraining dummies to be the same for both sub-samples. Standard errors are shown in parenthesis below coefficients. Statistical significance at the 10% (5%, 1%) level is shown by * (**, ***).

Table 5. Analysis of variance: Outcomes by core section

Outcome (excluding core grades):	Lewis & Clark	Reed	Whitman
	p value of ANOVA *F* test		
Cumulative GPA	0.231	0.144	0.001
First-year GPA	0.057	0.013	0.016
1st & 2nd year GPA	0.106	0.133	0.002
Narrow core-related GPA	0.458	0.227	0.003
Broad core-related GPA	0.485	0.131	0.003

Table 6. Probability values on *F* test of core-section dummy variables

GPA measure (excluding core):	Lewis & Clark	Reed	Whitman
	p-value on *F* test for section dummies		
All courses	0.56	0.44	0.51
First-year courses	0.21	0.38	0.30
First- & second-year courses	0.35	0.43	0.50
Narrow core-related courses	0.08	0.35	0.62
Broad core-related courses	0.57	0.49	0.50

Table 7. Peer-effects regressions: Estimated effects of peer averages of SAT scores and admission ratings

GPA measure (excl. core):	Lewis & Clark	Reed		Whitman
	Adm. rtg	SAT	Adm. rtg.	SAT
All courses	0.012 (0.016)	0.00114 (0.0326)	0.0286 (0.0773)	-0.00597 (0.0247)
1st year	0.017 (0.020)	0.013 (0.0363)	0.0428 (0.0862)	-0.0127 (0.0283)
1st & 2nd year	0.015 (0.017)	-0.00456 (0.0332)	-0.00872 (0.0781)	-0.00182 (0.0256)
Narrow core-related	0.021 (0.024)	-0.0121 (0.0345)	0.0584 (0.0776)	0.00659 (0.0277)
Broad core-related	0.008 (0.016)	0.0196 (0.0334)	0.0967 (0.0755)	-0.00471 (0.0261)

Standard errors are shown in parentheses. No coefficient was statistically significant at the 0.10 level

Table 8. Peer-effects regressions: Estimated effects of peer-mean predicted GPA

Dependent variable = GPA measure (excluding core):	Coefficient (standard error) on peer-mean predicted GPA		
	Lewis & Clark	Reed	Whitman
All courses	0.114 (0.160)	0.00328 (0.152)	0.0163 (0.0135)
First-year courses	0.050 (0.194)	-0.00751 (0.169)	0.00959 (0.0155)
First- & second-year courses	0.103 (0.164)	-0.0907 (0.153)	0.00900 (0.0140)
Narrow core-related courses	0.224 (0.211)	0.0639 (0.155)	0.00119 (0.0151)
Broad core-related courses	0.146 (0.161)	0.131 (0.150)	0.00626 (0.0142)

Table 9. Peer-effects regressions: Effects of peer mean/standard deviation/cross-product of predicted GPA

GPA measure (excl core)	Peer measure (based on predicted GPA)	Lewis & Clark	Reed	Whitman
All courses	Peer mean	-0.849 (1.124)	1.349* (0.681)	-0.177 (0.170)
	Class standard dev	-8.947 (10.91)	13.11* (6.680)	-0.0747 (0.0686)
	Mean * st dev	3.527 (4.172)	-5.392* (2.731)	0.0125 (0.0147)
First-year	Peer mean	-2.155* (1.256)	1.313* (0.755)	-0.205 (0.197)
	Class standard dev	-21.17* (12.04)	12.97* (7.362)	-0.0913 (0.0795)
	Mean * st dev	8.270* (4.611)	-5.275* (3.009)	0.0222 (0.0171)
Narrow core-related	Peer mean	-1.539 (1.452)	0.891 (0.740)	-0.0712 (0.191)
	Class standard dev	-17.18 (14.22)	7.996 (7.242)	-0.0269 (0.0767)
	Mean * st dev	6.535 (5.437)	-3.316 (2.961)	0.00370 (0.0164)

Table 10. Effects of shares of students in each quartile of overall distribution

College	GPA measure: All courses excluding core			
		Effect of:		
	Sample:	Top quartile share	Med-high quartile	Med-low quartile
Lewis & Clark	All	0.025 (0.132)	0.040 (0.125)	–0.046 (0.124)
	Top quartile	0.010 (0.182)	0.074 (0.183)	0.199 (0.178)
	Med-high quartile	0.045 (0.259)	0.114 (0.316)	–0.187 (0.278)
	Med-low quartile	–0.256 (0.338)	0.110 (0.300)	0.110 (0.300)
	Bottom quartile	0.259 (0.354)	0.243 (0.343)	–0.058 (0.336)
Reed	All	0.120 (0.118)	0.0606 (0.123)	0.141 (0.122)
	Top quartile	-0.0335 (0.200)	0.0798 (0.206)	-0.180 (0.237)
	Med-high quartile	0.116 (0.246)	-0.0116 (0.242)	0.266 (0.273)
	Med-low quartile	0.140 (0.259)	0.108 (0.283)	0.334 (0.319)
	Bottom quartile	0.190 (0.342)	0.232 (0.347)	0.302 (0.339)
Whitman	All	-0.0456 (0.116)	-0.151 (0.112)	-0.0199 (0.105)
	Top quartile	0.137 (0.161)	0.033 (0.161)	0.277* (0.157)
	Med-high quartile	0.0241 (0.259)	-0.0679 (0.239)	0.0829 (0.283)
	Med-low quartile	-0.185 (0.264)	-0.108 (0.235)	-0.225 (0.319)
	Bottom quartile	-0.155 (0.341)	-0.570* (0.335)	-0.241 (0.293)

References Cited

Ammermueller, Andreas, and Jörn-Steffen Pischke. 2006. Peer Effects in European Primary Schools: Evidence from PIRLS. Cambridge, Mass.: National Bureau of Economic Research. Working Paper #12180.

Angrist, Joshua D., and Kevin Lang. 2004. How Important are Classroom Peer Effects? Evidence from Boston's Metco Program. Cambridge, Mass.: National Bureau of Economic Research. Working Paper #9263.

Boozer, Michael A., and Stephen E. Cacciola. 2001. Inside the 'Black Box' of Project Star: Estimation of Peer Effects Using Experimental Data. New Haven, Conn.: Yale University Economic Growth Center. Discussion Paper #832.

Brownstone, David, and Robert Valletta. 2001. The Bootstrap and Multiple Imputations: Harnessing Increased Computing Power for Improved Statistical Tests. *Journal of Economic Perspectives* 15 (4):129-141.

Burke, Mary A., and Tim R. Sass. 2006. Classroom Peer Effects and Student Achievement. Tallahassee, Fla.: Florida State University Department of Economics. Working Paper.

De Long, J. Bradford, and Kevin Lang. 1992. Are All Economic Hypotheses False? *Journal of Political Economy* 100 (6):1257-72.

Dills, Angela K. 2005. Does Cream-Skimming Curdle the Milk? A Study of Peer Effects. *Economics of Education Review* 24 (1):19-28.

Ding, Weili, and Steven F. Lehrer. 2006. Do Peers Affect Student Achievement in China's Secondary Schools? Cambridge, Mass.: National Bureau of Economic Research. Working Paper #12305.

Epple, Dennis, and Richard E. Romano. 1998. Competition between Private and Public Schools, Vouchers, and Peer-Group Effects. *American Economic Review* 88 (1):33-62.

Foster, Gigi. 2006. It's Not Your Peers, and It's Not Your Friends: Some Progress toward Understanding the Educational Peer Effect Mechanism. *Journal of Public Economics* 90 (8-9):1455-75.

García-Diez, Mercedes. 2000. The Effects of Curriculum Reform on Economics Education in a Spanish College. *Education Economics* 8 (1):5-15.

Gaviria, Alejandro, and Steven Raphael. 2001. School-Based Peer Effects and Juvenile Behavior. *Review of Economics and Statistics* 83 (2):257-68.

Gwati, Ranganai. 2007. The Prediction versus Estimation Paradox in the Presence of Missing Data. Undergraduate Thesis, Program in Mathematics and Economics, Reed College, Portland, Ore.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2003. Does Peer Ability Affect Student Achievement? *Journal of Applied Econometrics* 18 (5):527-544.

Henderson, Vernon, Peter Miseszkowski, and Yvon Sauvegeau. 1976. Peer Group Effects and Educational Production Functions. Ottawa: Economic Council of Canada.

Hoel, Jessica, Jeffrey Parker, and Jon Rivenburg. 2006. A Test for Classmate Peer Effects in Higher Education. Portland, Ore.: Reed College.

Hoxby, Caroline. 2000. Peer Effects in the Classroom: Learning from Gender and Race Variation. Cambridge, Mass.: National Bureau of Economic Research. Working Paper #7867.

Jones, Michael P. 1996. Indicator and Stratification Methods for Missing Explanatory Variables in Multiple Linear Regression. *Journal of the American Statistical Association* 91 (433):222-230.

Kang, Changhui. 2007. Classroom Peer Effects and Academic Achievement: Quasi-Randomization Evidence from South Korea. *Journal of Urban Economics* 61 (3):458-495.

Kremer, Michael, and Dan Levy. 2008. Peer Effects and Alcohol Use among College Students. *Journal of Economic Perspectives* 22 (3):189-206.

Lefgren, Lars. 2004. Educational Peer Effects and the Chicago Public Schools. *Journal of Urban Economics* 56 (2):169-191.

Little, Roderick J. A. 1992. Regression with Missing X's: A Review. *Journal of the American Statistical Association* 87 (420):1227-37.

Lyle, David S. 2007. Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point. *Review of Economics and Statistics* 89 (2):289-299.

Manski, Charles F. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies* 60 (3):531-542.

McEwan, Patrick J. 2003. Peer Effects on Student Achievement: Evidence from Chile. *Economics of Education Review* 22 (2):131-41.

Rangvid, Beatrice Schindler. 2003. Educational Peer Effects: Regression Evidence from Denmark with PISA2000 Data. Copenhagen: AKF, Institute for Local Government Studies.

Robertson, Donald, and James Symons. 2003. Do Peer Groups Matter? Peer Group versus Schooling Effects on Academic Attainment. *Economica* 70 (277):31-53.

Royston, Patrick. 2005. Multiple Imputation of Missing Values: Update. *Stata Journal* 5:188-201.

Rubin, Donald B. 1996. Multiple Imputation after 18+ Years. *Journal of the American Statistical Association* 91 (434):473-489.

Sacerdote, Bruce. 2001. Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics* 116 (2):681-704.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2006. What Can Be Learned about Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds. *Journal of Public Economics* 90 (8-9):1435-54.

———. 2008. The Causal Effect of Studying on Academic Performance. *B.E. Journal of Economic Analysis & Policy* 8 (1: Frontiers):Article 14.

Vandenberghe, V. 2002. Evaluating the Magnitude and the Stakes of Peer Effects Analysing Science and Math Achievement across OECD. *Applied Economics* 34 (10):1283-90.

Winston, Gordon C., and David J. Zimmerman. 2004. Peer Effects in Higher Education. In *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*, edited by C. Hoxby. Chicago: National Bureau of Economic Research and University of Chicago Press.

Zimmer, Ron. 2003. A New Twist in the Educational Tracking Debate. *Economics of Education Review* 22 (3):307-315.
Zimmer, Ron W., and Eugenia F. Toma. 2000. Peer Effects in Private and Public Schools across Countries. *Journal of Policy Analysis and Management* 19 (1):75-92.
Zimmerman, David J. 2003. Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics* 85 (1):9-23.