

A Test for Classmate Peer Effects in Higher Education

Jessica Hoel, Jeffrey Parker, and Jon Rivenburg*
Reed College

September 2005

Abstract

We use data on ten years of entering students at Reed College to test for peer effects among classmates in a required freshman class. Assignment of students to class sections is effectively random, allowing us to avoid statistical problems that often make testing for peer effects impossible. We find evidence of nonlinear effects of the distribution of classmate abilities on student achievement. Students in sections with high mean and low standard deviation or low mean and high standard deviation of ability do well in their subsequent studies. Students in sections with low mean and low standard deviation or high mean and high standard deviation do poorly.

*The authors are, respectively, research assistant at the Board of Governors of the Federal Reserve System, George Hay Professor of Economics at Reed College, and Director of Institutional Research at Reed College. The opinions expressed are not necessarily those of the Federal Reserve System or of the administration of Reed College. This research was supported by a Bernard Goldhammer Summer Collaborative Research Grant. We would like to thank John Colgrove, Sharon Goodman-Lasher, Paul Marthers, and Ellen Yerkes for assistance with data acquisition, Albyn Jones and Marc Schneiberg for helpful methodological comments, and the summer inhabitants of the Public Policy Workshop, including Greg Agamalian, Michael Anderson, Dan Englander, Yves Savoie, Josh Simon, and Dawn Teele, for their contributions as this research was progressing. Remaining shortcomings are our own.

1. 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. Better students enable a more challenging curriculum and, so the story goes, a better education through interaction with peers inside and outside the classroom. The theoretical analysis of Rothschild and White (1995) supports the use of "merit aid" to attract students who contribute to the education of their peers.

Applicants usually 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.

While the behavior of participants in the higher-education market demonstrates faith in the importance of peer effects, very little research is available to support this belief.¹ This paper examines the effects of a student's classmates in a required first-year course on his or her subsequent academic success. The sample comprises ten years of entering first-year students at Reed College, a selective liberal-arts college. Each first-year student at Reed is assigned (effectively) randomly to a "conference" section of Humanities 110. We explore whether the academic qualifications of a student's classmates in this conference are associated with higher or lower achievement, measured by grade-point average or likelihood of graduation from Reed.

The paper is organized as follows. Section 2 examines the theoretical basis for peer effects in higher education and discusses some relevant literature. The third section describes our basic model of student achievement, which is used to predict a student's grade-point average based on his or her own characteristics. The fourth section presents our peer-effect results; section five interprets these results; section 6 concludes.

¹ In contrast, peer effect interactions on values, lifestyles, and purchasing have long been the focus of marketing research. The notion of "birds of a feather flocking together" is the basis of most geo-demographic marketing. See Weiss (1989).

2. PEER EFFECTS

2.1. Between vs. within-college peer effects

In principle, one can think of peer effects at two levels: between schools and within schools. The strongest peer effects in higher education are probably those between institutions. The general characteristics of a college's student body constrain its curriculum. Schools with bright and motivated students are able to assign more challenging materials and teach courses at a higher level than colleges whose students are academically less able. However, between-institution peer effects are impossible to measure unless one can distinguish between the educational effects of differences in peers and the effects of all the other characteristics that differ across schools.

This study examines peer effects within a single institution—Reed College. Differences among peer groups within the population of Reed students are surely less striking than the differences between groups of Reed students and those at, say, a typical state university. Nonetheless, there are likely some groups of Reed peers whose effects are more conducive to high academic achievement than if one interacted with other groups. We have all participated in classes in which the classroom interaction among students has contributed strongly to learning—and classes where the peer dynamic has been less supportive. We have all known students whose academic careers suffered because of the behavior of classmates, roommates, neighbors, or friends—and others whose performances have improved due to their interaction with these peers.

2.2. Choice of peers and the reflection problem

Estimation of peer effects is made impossible in most higher-education settings because most college students choose their peers. If a student's peers have unobserved characteristics that are systematically related to his or her own, then estimation of peer effects is made impossible by what Charles Manski (1993) calls the "reflection problem." For example, if smart students systematically tend to choose smart peers, then it becomes impossible to distinguish statistically between the effects of the student's own smartness and the effect of peers' smartness.

College students always choose their friends and they often choose roommates, neighbors, and classmates. They are likely to choose peers whose unobservable characteristics are systematically related to their own. Even when the choice of peers is not explicitly voluntary, there may be associations between characteristics of peers. For example, by enrolling in a difficult course (or an advanced or honors section of a course) a good student would surround himself or herself with good peers.

The reflection problem can be overcome by using peer groups where one of the following conditions holds: (a) selection into groups is random, or at least independent of variables related to achievement; or (b) selection into groups is based only on measurable characteristics that can be used as controls in estimation. Neither of these conditions will be fulfilled in situations where students select their peers since at least some of the variables on which such selection is based are impossible to observe. Thus, the estimation of peer effects must be based on externally (and, ideally, randomly) assigned peer groups.

2.3. Studies of peer effects

Students in primary and secondary schools may be assigned into classrooms in a way that is not related to achievement. Thus, studies of classroom peer effects at these levels are more common. Among the studies finding positive classroom peer effects are Coleman et al. (1966), Henderson, Mieszkowski, and Sauvageau (1976), Hoxby and Terry (2000), Zimmer and Toma (2000), Checchi and Zollino (2001), Robertson and Symons (2003), and McEwan (2003). These studies find that having better peers tends to improve a student's own academic performance; many of them also find the effects to be larger for low-ability students than for those with high ability.

Random class assignment is unusual in higher education, which makes testing for classroom peer effects more difficult. García-Diez (2000) estimates positive peer effects among first-year economics students in a Spanish university. Because the primary purpose of the study was assessment of a change in curriculum rather than estimation of peer effects, the author does not address issues of selection bias.

Assignment of first-year roommates, while based to some degree on survey responses, is often approximately random. Relevant selection issues are addressed explicitly in recent studies of residential peer effects at U.S. colleges. Bruce Sacerdote (2001) and

David Zimmerman (2003) examine the effects of the characteristics of first-year roommates at, respectively, Dartmouth College and Williams College on academic achievement.

Winston and Zimmerman (2004) use the Andrew W. Mellon Foundation's College and Beyond survey to estimate roommate peer effects at three unnamed colleges.

Zimmerman finds some evidence of positive roommate and dormmate peer effects for some categories of students. Using combined SAT scores as his measure of peer quality, he finds that having a roommate with low SAT scores will diminish the academic performance of the student near the middle of the SAT distribution. Students at the low end of the distribution are adversely affected by having a low-SAT-average dorm.

Sacerdote finds that one's own and one's roommate's first-year GPA outcomes are strongly correlated, but because of the reflection problem that does not imply peer causality. Linear effects of roommate admission characteristics are not significant, but he finds evidence of some nonlinear effects: having a roommate whose admission index is in the top 25% raises one's first-year GPA by 0.06 and is statistically significant.

Winston and Zimmerman's study finds some evidence of positive peer effects. The strongest effects seem to be associated with middle-tier students (those between the 15th and 85th percentiles among their peers in SAT scores) and the negative effects of a low-SAT roommate seem stronger than the positive effects of a high-SAT roommate.

2.4 Testing for peer effects in Humanities 110

As noted above, the most vexing econometric challenge to the estimation of peer effects is the reflection problem associated with endogenous selection of peers. Because classmates are allocated by an essentially random mechanism, Reed's Humanities 110 course provides a suitable natural experiment for estimation of classroom peer effects.

Nearly all Reed College first-year students are required to enroll in Humanities 110 for the entire first year.² The course meets for six hours per week. Three hours per week are a lecture in which all 350 or so students meet together; the other three hours are a

² A few first-year students do not enroll in Hum 110. For example, students whose prior training in chemistry allows them to enroll in sophomore-level organic chemistry cannot take Hum 110 due to a time conflict with the humanities lecture. These students would take Hum 110 as sophomores. Any student who did not take Hum 110 in his or her first full-time semester is excluded from our sample, but is included in peer statistics for the sections of Hum 110 in which he or she enrolled.

conference section of approximately 16 students led by a faculty member. Hum 110 is widely viewed as one of the distinctive elements of the Reed curriculum; it introduces Reed students not only to classical scholarship in the humanities, but also to the conference style of teaching and learning that is common at Reed. It also serves as the vehicle for teaching basic writing and critical-reasoning skills to entering students. Therefore, a student's experience in Hum 110 might be expected to have an unusually significant influence on his or her academic success in other courses.

Conference sections are offered at various times of the week. Students sign up for a humanities conference section at their initial meetings with their academic advisors. Instructors' names are not shown in the published class schedule, so the choice of conference section is essentially a choice of time. The required lecture is held at 9am on Mondays, Wednesdays, and Fridays, with most conference sections being at later times on those same days. Thus, a desire to avoid early-morning classes (which might be correlated with late-night partying and low academic motivation) has little effect on the choice of conference time. However, the conference time must not conflict with other courses that the student takes. Introductory courses in biology, chemistry, physics, and psychology have a lecture component that all students in the course must attend, so this imparts some selectivity into the choice of Hum conference times. For example, no student who takes introductory chemistry in his or her first year could enroll in an 11am MWF humanities conference because it would conflict with the intro chemistry lecture.

New students and their advisors meet during orientation to plan a schedule for the first year. Once they have agreed on a preferred conference time, the registrar's office places students into sections. Most conference sections are bunched at popular times.³ When multiple sections are offered at the same time, the registrar fills the sections sequentially in the order in which the students are registered, which depends on the advising appointment times assigned to the students (apparently randomly) by the student services staff. When all sections at a requested time are full, subsequent students requesting that time are assigned to a section at another time that does not conflict with the students' other classes. Once a

³ Nearly 44% of the students in our sample enrolled in conferences meeting at the two most popular times: 10am and 11am on Mondays, Wednesdays, and Fridays compared with only 28% total in the five afternoon time slots.

tentative assignment of all students into sections has been made, the registrar's office re-assigns students as necessary to balance the gender composition of sections.

Thus, the assignment of conference classmates in Humanities 110 can be thought of as nearly random with respect to variables that are likely to affect academic success. As Zimmerman (2003) shows, bias in peer effects can be avoided by including any non-random selection criteria as control variables. Because of the effects of schedule conflicts, time of day is clearly such a criterion. We tested dummy variables for the time of the humanities conference along with other characteristics of the conference (size, instructor) to control for such bias. None of these characteristics had statistically significant effects on student achievement, so they were dropped from the final specification.

3. DATA SET AND SPECIFICATION ISSUES

3.1 Sample and outcome measures

Our sample includes all students who enrolled in Humanities 110 at Reed College as freshmen between fall semester 1988 and fall semester 1997. During that period, the content of Hum 110 changed little and the method of allocating students among sections did not vary. Our sample is not restricted to students who completed their degrees at Reed but includes as well the approximately 30% of the entering students who had not been graduated from Reed as of May 2004.

Cumulative grade-point average is our principal outcome measure. We also look for peer effects on first-year and first-and-second-year GPA under the presumption that first-year Hum 110 peers may have a stronger effect on academic performance in the early years than on performance in the final years. We also looked at students' GPA in humanities-related courses to see if Hum 110 peers had a stronger effect in courses more closely related to the content of Hum 110.

One possible, and adverse, effect of strong peers on grades is that they may raise the "curve" in the student's section in Hum 110, lowering his or her grade in that course, *ceteris paribus*. To avoid this effect, we exclude the grade in Hum 110 from all GPA calculations.

A second category of outcome variable is whether the student was eventually graduated from Reed College. A sizable minority of the students who begin at Reed as

freshmen transfer to other institutions, usually after one or two years at Reed. The reasons for this movement are only partially understood; we test for the possibility that academic characteristics of Hum 110 classmates may play a role in students' decisions about whether or not to depart.

3.2 Control variables

In estimating classmates' effects on a student's academic performance, we must control for his or her own abilities and characteristics. Studies of college roommate peer effects have used scores on the SAT entrance exam, high-school grade-point average, and/or a rating assigned by the admission office as controls for an individual student's ability. We find that for our sample, each of these measures has an independent effect on student performance—none is a sufficient statistic for predicting success.⁴

Moreover, student characteristics other than admission credentials may affect GPA. We include demographic variables (ethnicity, sex, financial-aid status, and U.S. citizenship status), dummy variables for year of entry (to capture any possible changes in grading standards over time), and dummy variables for the student's area of major (to capture differences in grading practices across disciplines).

As noted above, we also tested controls for the time of the Hum 110 conference section and for the identity of the instructor. The estimated coefficients on the meeting-time dummies were statistically indistinguishable from zero, both individually and collectively. The coefficients on instructors were collectively insignificant; while a few instructors had coefficients that were statistically significant, the number of significant results was broadly consistent with the number to be expected under the null hypothesis. Based on these results, we omitted the time and instructor controls from the final model.

While we had records of SAT scores, high-school grade-point average, high-school class rank, and admission office reader rating for the majority of students, one or more of

⁴ Among the five admission "quality" measures we use—math SAT, verbal SAT, high-school GPA, a dummy variable for being in the top five percent of the high-school class, and the admissions office's reader rating—the highest pairwise correlation is 0.53. A linear regression of reader rating on the other four measures yields an *R*-square of just 0.41. Presumably the residual in this regression reflects information from admission essays, interviews, and other information that is not reflected in the raw numbers.

these indicators was missing for over one-third of the sample. Using a traditional linear regression treating these as continuous variables would restrict our sample to about 1900 of the 3155 potential observations.

To avoid this reduction in sample size, and to allow for possible non-linearities in the effects of these variables on GPA, we specified verbal and math SAT, high-school GPA, and admission reader rating as categorical variables, dividing the observed sample into four categories as closely as possible to the 85th, 50th, and 15th percentiles.⁵ In addition to the high, medium-high, medium-low, and low categories for each measure, we coded another dummy variable in each set that was set equal to one for observations with missing values.⁶ This allowed all observations to be retained in the sample. The information in the high-school rank variable was captured effectively with a two-level dummy-variable specification indicating whether or not the student was in the top five percent of his or her high-school class, with a “missing” dummy variable if no class rank was reported. Of those reporting class rank, 31% were in the top five percent.

Several adjustments to the raw data should be noted. First, the College Board changed the scaling of SAT scores during our sample period to compensate for accumulated drift in the distribution of scores over time. We use “recentered” scores for the earlier period, so all reported SAT scores are comparable and measured according to the new scale. Second, we observed high-school GPA values in excess of 4.0 for a large number of observations, with one as large as 9.5. Consultation with staff in the Reed admissions office suggests that many of the GPAs that are greater than 4.0 result from a policy of giving higher grade values for advanced work such as Advanced Placement courses. Thus, to some extent, GPAs greater than 4.0 may be informative. Without detailed knowledge of the grading systems of the individual high schools involved, no perfect solution to this problem is possible. Based on the advice of the Reed admissions staff, we excluded from our sample

⁵ Zimmerman and Winston and Zimmerman both use 85th and 15th percentiles in dividing their samples. We chose these levels for consistency with these earlier studies.

⁶ There is only a single “missing” dummy variable for the two SAT scores because the math and verbal scores were missing for the same 110 students. 726 students had missing high-school GPAs and 922 were missing class rank. Many students are missing multiple variables, so the total number with one or more of these variables missing is 1217.

the six observations with GPA greater than 5.0 under the assumption that these reflect non-standard grading systems, but retained the 280 observations between 4.0 and 5.0.

3.3 Individual characteristics and GPA

We examine the effects of individual student characteristics on cumulative GPA for two reasons. First, this establishes a baseline model of student achievement to which we add variables measuring the characteristics of classroom peers. Second, we achieve a parsimonious measure of peer quality by using the predicted GPA values from this baseline model, which incorporate in a single variable the multidimensional information in the student's admission file.⁷

The baseline regression using the categories for the admission variables is shown in Table 1.⁸ The "low" category is omitted for each set of dummy variables. For comparison, the right-hand columns of Table 1 show a continuous specification of SAT scores and high-school GPA for the sub-sample for which all variables were observed.

The pattern of effects described by the non-linear continuous specification (quadratic for verbal and cubic for math SAT scores) matches the pattern of the categorical specification quite closely. In particular, while students with very high math SAT scores do significantly better at Reed, there is essentially no difference between the GPA performance of students with medium and low math scores. In contrast, students with low verbal SAT scores perform significantly worse than those with medium or high scores, but there is little difference (and, in fact, a slightly inverse relationship) among the GPAs of those in the higher categories. This difference reflects the structure of course requirements at Reed. A student in a non-quantitative major can navigate the Reed curriculum without having to take any highly quantitative course, thus weak quantitative skills need not lower ones grades. However, the Reed curriculum requires extensive writing from every student, not least in

⁷ The use of the predicted GPA from this equation provides another reason for adopting the category specification that maximizes the sample size. In attempting to characterize the quality of classroom peers, it is desirable to include all of one's classmates. With the categorical specification, the only classmates for whom a predicted GPA cannot be calculated are a handful of special students such as visiting students and local high-school students taking the course by special arrangement.

⁸ Of the 3155 students in our full sample, thirteen did not complete any courses at Reed, and thus had no grade-point average. Therefore, there are only 3142 observations in these regressions.

Hum 110 and the required senior thesis. Thus the grades of students with weak verbal skills are likely to suffer.

Table 1. Basic regressions for undergraduate GPA

Basic regressions for cumulative undergraduate GPA				
Explanatory Variable	Category specification	Coefficient (standard error)	Continuous specification	Coefficient (standard error)
Verbal SAT	High	0.074* (0.04)	SAT verbal/100	0.466* (0.239)
	Medium high	0.084** (0.03)	SAT verbal/100 squared	-0.033* (0.018)
	Medium low	0.094*** (0.03)		
Math SAT	High	0.142*** (0.04)	SAT math/100	3.547* (1.907)
	Medium high	0.03 (0.03)	SAT math/100 squared	-0.622** (0.303)
	Medium low	0.007 (0.03)	SAT math/100 cubed	0.036** (0.016)
	SAT Missing	0.066 (0.07)		
High-school GPA	High	0.268*** (0.05)	High-school GPA	0.295*** (0.044)
	Medium high	0.225*** (0.04)		
	Medium low	0.113*** (0.04)		
	Missing	0.066*** (0.07)		
Reader rating	High	0.466*** (0.04)	Reader rating	0.259*** (0.034)
	Medium high	0.292*** (0.03)		
	Medium low	0.143*** (0.03)		
	Missing	0.365*** (0.13)		
High-school rank	Top five percent	0.116*** (0.03)	Top five percent	0.102*** (0.032)
	Missing	0.035 (0.03)		
Demographic	Asian	-0.028 (0.05)	Asian	-0.114* (0.060)
	Black	-0.243** (0.10)	Black	-0.288** (0.131)
	Caucasian	0.051 (0.03)	Caucasian	0.027 (0.043)
	Financial-aid eligible	-0.014 (0.02)	Financial aid eligible	-0.044* (0.026)
	Hispanic	-0.027 (0.06)	Hispanic	-0.101 (0.073)
	Male	-0.086*** (0.02)	Male	-0.100*** (0.026)
	Native American	-0.069 (0.13)	Native American	-0.014 (0.143)
	U.S. citizen	-0.111** (0.04)	U.S. citizen	-0.093 (0.063)
	Observations	3142	Observations	1937
	R-squared	0.20	R-squared	0.24

significant at 10%; ** significant at 5%; *** significant at 1%
Equation also includes a constant term and dummies for year of entry and division of major.

4. TESTING FOR CLASSROOM PEER EFFECTS

4.1 Classmates and academic performance

There are many ways in which freshman humanities classmates might affect one's academic performance. The behavior of classmates may instruct, inspire, or build confidence in a student, leading to greater academic success. Alternatively, it may fail to do these things and may even lower achievement by demoralizing or intimidating the student. We list below a few of the channels through which classmates may affect success and describe how these might be captured in our data.

Keep in mind that our main academic performance measures are grades in class *other than Hum 110*. Thus, we are testing for classmate effects that would show up in grades in other classes. Given the foundational nature of the Hum 110 course in the Reed curriculum, one might expect a student's academic experience in that class to have significant effects on other classes both during and after the first year. These effects might be stronger early in the student's Reed career and in courses more closely related to the humanities.

Better classmates make for better learning: Having a stronger cohort of humanities classmates should make for more informed and dynamic discussion. This may help all students in the conference to learn the material better and to become more adept at participating in classroom conferences in general. Conversely, a class in which all students are weak may not engage in productive discussion and is unlikely to elicit the subtleties of the texts being analyzed. If this effect is important, we would expect a higher mean ability of classmates to improve a student's performance.

Intimidation or demoralization effects: A student with classmates who seem more intelligent or academically able than his or her own perceived abilities may become discouraged. This could lead the student to perform below his or her true abilities. Thus weaker students with higher-mean classmates may suffer reduced performance.

Conference dynamics may be related to the variation in abilities: A conference section with extremely high variance of abilities may not function effectively because students will absorb and understand the material at very different levels. Alternatively, the presence of at least some variation in abilities may provide beneficial opportunities for peer learning.

Some “abilities” may be more relevant in peers than others: A conference composed mainly of students with outstanding abilities in mathematics and sciences may have less appreciation for and aptitude to understand humanities material than one with students with stronger verbal skills. Thus, although our main indicator of peer quality is predicted GPA, we also consider whether peer characteristics such as verbal and math SAT scores have potentially different effects.

4.2 Effects of mean peer ability

The first set of tests we report examine linear effects of the mean classmate predicted GPA on the individual student’s grades. While classmate means naturally vary much less than individual predicted GPAs, our sample shows considerable variation. The classmate-mean variable has a mean of 2.99 with a standard deviation of 0.071 and a range of (2.77, 3.24). The lowest classmate mean value is near the 20th percentile of the overall distribution of predicted GPAs and the highest mean is near the 80th percentile. Table 2 shows the coefficient on the average classmate predicted GPA in regressions with five own GPA measures as the dependent variable.⁹ All regressions include the full battery of control variables shown in Table 1.¹⁰

⁹ The narrow hum-related GPA includes grades in courses in humanities (other than 110), history, literature, English, philosophy, upper-level modern foreign languages, political philosophy, afrr history, religion, classics, Latin, and Greek. The broad measure includes all courses other than mathematics and the natural sciences. Senior thesis and independent study courses are excluded from both.

¹⁰ The dummy variables for division of major are unnecessary in the regressions for Hum-related GPA because there is no need to control for interdisciplinary differences in grading practices when only Hum-related courses are included. Excluding these variables had a negligible effect on the peer effect coefficients.

Table 2. Linear effects of classmate mean predicted GPA

Sample	Effect of Hum 110 classmates' mean predicted GPA on student's own GPA (various measures of own GPA all exclude Hum 110)				
	Cumulative	First-and-second year	First-year	Hum-related (narrow)	Hum-related (broad)
All	0.078 (0.162)	0.011 (0.165)	0.110 (0.178)	-0.073 (0.190)	0.037 (0.169)
Males	0.117 (0.250)	-0.021 (0.259)	-0.004 (0.271)	0.085 (0.299)	0.160 (0.262)
Females	0.039 (0.211)	0.050 (0.210)	0.221 (0.233)	-0.192 (0.241)	-0.069 (0.218)
Top 15%	-0.504 (0.314)	-0.521 (0.355)	-0.509 (0.392)	-0.479 (0.329)	-0.548*
Middle 70%	0.308 (0.197)	0.256 (0.200)	0.335 (0.215)	0.093 (0.235)	0.311 (0.207)
Bottom 15%	-0.652 (0.495)	-0.756 (0.488)	-0.429 (0.524)	-0.689 (0.593)	-0.986*

Standard errors in parentheses below coefficients
 *Statistically significantly different than zero at the 10% level. No coefficient was significant at 5%.

Of the 30 estimated coefficients reported in Table 2, only two in the lower right-hand corner are statistically significant at the 10% level; none is significant at 5%. One expects 3 out of 30 estimated coefficients to show 10% significance under the null hypothesis that all coefficients are zero, so the results in Table 2 are fully consistent with the absence of any linear effects of average peer ability on one's own grades.

However, there is a consistent pattern in Table 2 that is suggestive. The bottom three rows indicate consistent and meaningfully large point estimates that differ markedly across the student's own quality categories. Good peers seem to be beneficial for the middle 70% of students but detrimental for those at the top and the bottom of the distribution. Thus, Table 2 suggests that peer effects may be different for sub-samples of students stratified by predicted GPA.

The most straightforward specification is linear in mean peer quality, but evidence from prior studies of roommate peer effects suggests that nonlinear and even non-monotonic specifications may be appropriate. However, adding quadratic and cubic terms to the regressions in Table 2 fails to reveal any pattern of significant coefficients. Nor do significant classmate effects emerge when classmate-mean peer quality is broken into quantiles and entered into the equation as a set of dummy variables.

4.3 Effects of peer standard deviation

As noted above, high variance of ability within a section could affect the educational effectiveness of the conference. We test for this initially by including the standard deviation of predicted GPA among members of the Hum 110 section as an explanatory variable in the achievement regression. The classmate-standard-deviation variable has a mean of 0.245, a standard deviation of 0.045, and a range of (0.119, 0.385). The estimated coefficients on classmate standard deviation are shown in Table 3. As with the classmate means, most of the coefficients on classmate standard deviation are small and statistically insignificant. However, a strong and significant negative effect of variation in abilities shows up for the weakest students in the sample. This implies that the weakest students may be negatively impacted by participating in a conference with a high degree of variation in abilities. Coupled with the (statistically weaker) evidence of negative mean effects in Table 2, this raises the possibility that the weakest students may be negatively impacted by being in the same conference with very strong students. Perhaps the weakest students receive less constructive attention from faculty when there is a high mean and a wide variation of skills. Alternatively, they may be intimidated into low participation, and thus reduced learning, by the presence of a wide range of abilities.

Table 3. Effects of standard deviation of classmates' predicted GPA

Sample	Effect of standard deviation of Hum 110 classmates' predicted GPA on student's own GPA (various measures of own GPA all exclude Hum 110)				
	Cumulative	First-and-second year	First-year	Hum-related (narrow)	Hum-related (broad)
All	0.111 (0.239)	0.122 (0.246)	-0.165 (0.264)	0.148 (0.277)	0.053 (0.247)
Males	0.163 (0.366)	0.185 (0.382)	-0.188 (0.399)	-0.025 (0.434)	-0.047 (0.380)
Females	0.206 (0.316)	0.199 (0.316)	0.024 (0.351)	0.406 (0.356)	0.294 (0.324)
Top 15%	0.295 (0.532)	0.090 (0.603)	-0.044 (0.666)	-0.180 (0.569)	0.073 (0.539)
Middle 70%	0.346 (0.290)	0.151 (0.295)	-0.093 (0.317)	0.285 (0.337)	0.209 (0.300)
Bottom 15%	-1.762** (0.715)	-0.676 (0.718)	-1.272* (0.769)	-0.822 (0.866)	-1.287* (0.748)
Standard errors in parentheses below coefficients * (**) Statistically significantly different than zero at the 10% (5%) level.					

4.4 Effects of position in classroom distribution

The possibility that students may be intimidated where they are weak relative to their classmates or not stimulated in a class where they are stronger than their classmates suggests that the student's relative position in the classroom ability distribution may matter. We measure this position as the difference between the student's own predicted GPA and the mean of his or her peers, divided by the class standard deviation—analogue to a standardized z -score. Thus, a student who is one standard deviation above the mean of her peers would have a value of +1.0; one who is exactly at the mean would have 0.0.

Table 4 shows the estimated coefficients on the positional measure in GPA regressions. Such effects are very small and with two exceptions statistically insignificant. The Middle 70% sample has significant negative positional effects for a couple of GPA measures. For these students, the results suggest that they do better in their subsequent classes if they are lower in the distribution of their humanities conference. Because a given student will be relatively low in the distribution of a class with a high classmate mean, this supports the positive (but statistically insignificant) effect of the class mean shown in Table 2. It is worth noting that the students in the Top 15% are nearly always at the top of the classroom distribution (positional variable ranges from 0.4 to 2.7) and those in the Bottom 15% are nearly always at the bottom (ranging from -0.5 to -2.9), so the greatest variation in the positional variable occurs in the middle group, which may help to explain why the standard errors of the coefficients for this group are much smaller than those of the other groups.

Table 4. Effects of position in classroom distribution

Sample	Effect of position in distribution of Hum 110 classmates' predicted GPA (own minus mean, divided by standard deviation) on student's own GPA (various measures of own GPA all exclude Hum 110)				
	Cumulative	First-and-second year	First-year	Hum-related (narrow)	Hum-related (broad)
All	-0.051 (0.035)	-0.026 (0.036)	-0.041 (0.039)	-0.025 (0.041)	-0.049 (0.036)
Males	-0.053 (0.054)	-0.020 (0.057)	-0.029 (0.059)	-0.076 (0.065)	-0.068 (0.056)
Females	-0.052 (0.045)	-0.042 (0.045)	-0.065 (0.050)	0.008 (0.052)	-0.040 (0.046)
Top 15%	0.071 (0.066)	0.090 (0.075)	0.099 (0.082)	0.109 (0.070)	0.091 (0.066)
Middle 70%	-0.067 (0.044)	-0.068 (0.044)	-0.079* (0.048)	-0.047 (0.051)	-0.090** (0.045)
Bottom 15%	-0.126 (0.122)	0.056 (0.121)	-0.096 (0.129)	-0.000 (0.147)	0.008 (0.128)
Standard errors in parentheses below coefficients * (**). Statistically significantly different than zero at the 10% (5%) level.					

4.5 Interactions of peer mean and standard deviation

The results presented in Table 2, Table 3, and Table 4 do not reveal a simple relationship between a student's humanities classmates and his or her subsequent academic success. To determine whether there are important interactions between the various characteristics of the classroom distribution of abilities, we examine two additional models. The first includes both the classmate mean and standard deviation along with their product, to test for the possibilities of interactions between class mean and class variance. The results are shown in Table 5.

Table 5. Interactions between peer mean and standard deviation

Sample	Classmate Variable	Effect of classmate variable on GPA (various measures of own GPA all exclude Hum 110)				
		Cumulative	First-and-second year	First-year	Hum-related (narrow)	Hum-related (broad)
All	Mean	1.394* (0.824)	1.730** (0.845)	1.626* (0.908)	1.511 (0.954)	1.727** (0.851)
	St. Dev.	16.817* (10.192)	21.955** (10.448)	18.639* (11.220)	20.057* (11.776)	21.123** (10.537)
	Mean * St. Dev.	-5.600 (3.412)	-7.311** (3.498)	-6.307* (3.756)	-6.663* (3.943)	-7.060** (3.528)
Males	Mean	1.389 (1.264)	2.292* (1.317)	2.571* (1.379)	1.875 (1.485)	1.828 (1.316)
	St. Dev.	16.398 (15.637)	29.668* (16.299)	31.977* (17.032)	21.651 (18.386)	20.137 (16.328)
	Mean * St. Dev.	-5.436 (5.227)	-9.855* (5.449)	-10.757* (5.693)	-7.261 (6.146)	-6.767 (5.458)
Females	Mean	1.286 (1.089)	1.048 (1.087)	0.577 (1.204)	1.187 (1.236)	1.523 (1.109)
	St. Dev.	16.167 (13.458)	13.015 (13.442)	4.429 (14.876)	18.453 (15.220)	20.760 (13.708)
	Mean * St. Dev.	-5.355 (4.512)	-4.301 (4.507)	-1.494 (4.987)	-6.036 (5.105)	-6.857 (4.596)
Top 15%	Mean	2.187 (1.741)	1.899 (1.975)	-0.921 (2.187)	2.428 (1.793)	2.394 (1.749)
	St. Dev.	32.143 (20.445)	28.873 (23.193)	-4.940 (25.683)	35.001* (21.108)	35.052* (20.550)
	Mean * St. Dev.	-10.670 (6.825)	-9.650 (7.742)	1.598 (8.574)	-11.803* (7.051)	-11.722* (6.861)
Middle 70%	Mean	1.367 (1.000)	1.703* (1.016)	1.759 (1.094)	1.042 (1.175)	1.737* (1.037)
	St. Dev.	14.376 (12.675)	18.922 (12.872)	17.940 (13.849)	12.427 (14.851)	18.300 (13.158)
	Mean * St. Dev.	-4.721 (4.243)	-6.306 (4.309)	-6.068 (4.636)	-4.074 (4.972)	-6.084 (4.405)
Bottom 15%	Mean	-1.876 (2.882)	-2.260 (2.877)	0.894 (3.068)	1.297 (3.431)	-2.076 (3.000)
	St. Dev.	-21.389 (33.152)	-18.777 (33.156)	10.341 (35.277)	21.307 (39.498)	-17.111 (34.648)
	Mean * St. Dev.	6.625 (11.123)	6.213 (11.126)	-3.878 (11.834)	-7.312 (13.253)	5.455 (11.625)

Because of the model's nonlinearity, the results in Table 5 are difficult to interpret directly. Figure 1 shows the estimated response of a student's first-and-second-year GPA to the mean and standard deviation of the classmate distribution, using the estimates for the entire sample. The effects are both large and statistically significant. The model predicts a

positive effect for conferences with high mean and low standard deviation and for those with low mean and high standard deviation. Humanities conferences in which both mean and standard deviation are either high or low seem to be associated with poorer student performance in other classes. Class mean and standard deviation have a correlation coefficient of 0.11, so there is no strong tendency for the conferences in our sample to cluster along either diagonal.

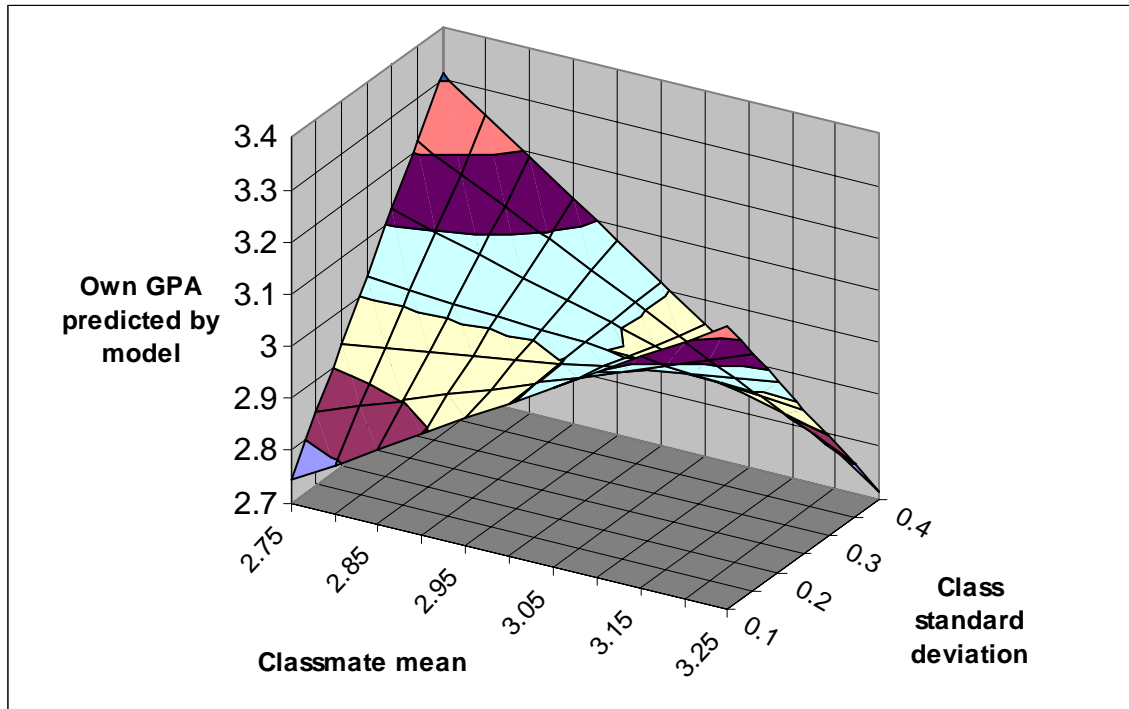


Figure 1. Estimated response of GPA to classmate mean and standard deviation

The partial effects of mean and standard deviation shown in Figure 1 show clearly why the simple linear specification fails to yield consistently significant effects. Higher mean classmate ability raises a student’s performance in conferences where students have uniform abilities, but lowers it in high-dispersion conferences. Similarly, higher standard deviation of conference peers raises a student’s performance when the peer-ability mean is low, but reduces it when peers have high average ability.

The nature of these effects is broadly confirmed by regressions using categorical dummy variables for bivariate terciles of the joint mean/standard deviation distribution. The highest performance (in order) is associated with the three cells: (high mean, low standard

deviation), (medium mean, low standard deviation), and (medium mean, medium standard deviation). The lowest performance is from students in conferences with (low mean, low standard deviation), (high mean, medium standard deviation), and (medium mean, high standard deviation).¹¹

Low-mean conferences with high standard deviation most likely have more strong students than low-mean conferences with low standard deviation, because it is the strong students who vary most from a low mean. High-mean conferences with low standard deviation contain fewer weak students than high-mean conferences with high standard deviation. Since it is students in the high-mean-low-standard-deviation and low-mean-high-standard-deviation conferences who seem to outperform their predicted GPA, this suggests that perhaps having at least a critical number of strong students (or at most a critical number of weak ones) might be important. However, this result was not supported by tests incorporating the number of top-20%, top-30%, or top-50% students in the conference.

4.6 Effects of gender composition of conference

Given the low significance levels of most measures of classroom peer quality, a striking result is the significant effect that gender composition has on GPA. This result is somewhat surprising given that Reed's policy of attempting to balance gender composition across sections limits the variation in this independent variable. The male share (measured as a fraction between zero and one) and has a mean of 0.49 and standard deviation of 0.08, reflecting the fact that most humanities conferences are nearly balanced.

Table 6 shows the effects of the male share on the various GPA measures.¹² Effects are small but significantly negative for the female population for all GPA measures and that significance carries over into the aggregate sample for four of the five measures. A one-standard-deviation increase in the share of females is estimated to raise a female student's cumulative non-Hum-110 GPA by about 0.05. Effects on male students are weak, inconsistent, and insignificant.

¹¹ The exact ranking of categories differs somewhat across the various dependent variables. The ranking described here is for first-and-second-year GPA to facilitate comparison with Figure 1.

¹² The estimates shown include standard control variables, but no peer quality measures. Estimates including peer mean, conference standard deviation, and the interaction between them are similar.

Table 6. Effects of gender composition of Hum 110 conference

Sample	Effect of share of males in Hum 110 conference on student's own GPA (various measures of own GPA all exclude Hum 110)				
	Cumulative	First-and-second year	First-year	Hum-related (narrow)	Hum-related (broad)
All	-0.386*** (0.147)	-0.279* (0.151)	-0.091 (0.162)	-0.522*** (0.172)	-0.337** (0.152)
Males	-0.108 (0.232)	0.114 (0.242)	0.289 (0.253)	-0.244 (0.279)	-0.012 (0.242)
Females	-0.676*** (0.190)	-0.674*** (0.189)	-0.466** (0.210)	-0.833*** (0.215)	-0.646*** (0.193)
Top 15%	0.122 (0.318)	0.188 (0.360)	0.468 (0.397)	0.151 (0.343)	0.047 (0.323)
Middle 70%	-0.473*** (0.180)	-0.393** (0.182)	-0.197 (0.196)	-0.600*** (0.211)	-0.427** (0.186)
Bottom 15%	-0.754* (0.419)	-0.402 (0.417)	-0.458 (0.443)	-0.882* (0.516)	-0.511 (0.440)
Standard errors in parentheses below coefficients * (**, ***) Statistically significantly different than zero at the 10% (5%, 1%) level.					

The negative effect of male classmates seems to be stronger for students with weaker admission credentials. Having more male classmates has a small positive effect on the GPA outcomes of students in the top 15% of predicted GPA, though this effect is not statistically significant. (This positive, but insignificant, effect also holds for females in the top 15%.) However, strong and significant negative effects emerge for the middle 70%. Moreover, although the much-smaller sample leads to higher standard errors (and therefore to lower significance levels), the estimated effect on every GPA measure is even larger for the lowest 15%.

4.7 Effects of Hum 110 peers on persistence to graduation

The effects of the mean and standard deviation of humanities classmate quality on GPA shown in Table 5 and Figure 1 show up in a similar manner for the probability of graduation from Reed. Over the ten-year sample, 69.1% of the students who took humanities as first-year college students eventually received a Reed bachelor's degree. A probit regression of graduation on a similar set of control variables yields positive coefficients with *t* statistics of 2.98 and 3.37, respectively, on peer mean predicted GPA and

conference standard deviation of predicted GPA.¹³ The interaction term has a negative coefficient and a t statistic of -3.36 .

The predicted probability of graduation is shown as a function of mean classmate quality and class standard deviation in Figure 2. Students in high-mean, low-standard-deviation conferences and low-mean, high-standard-deviation conferences have higher probabilities of graduation than those in high-mean, high-standard-deviation and low-mean, low-standard-deviation conferences.¹⁴

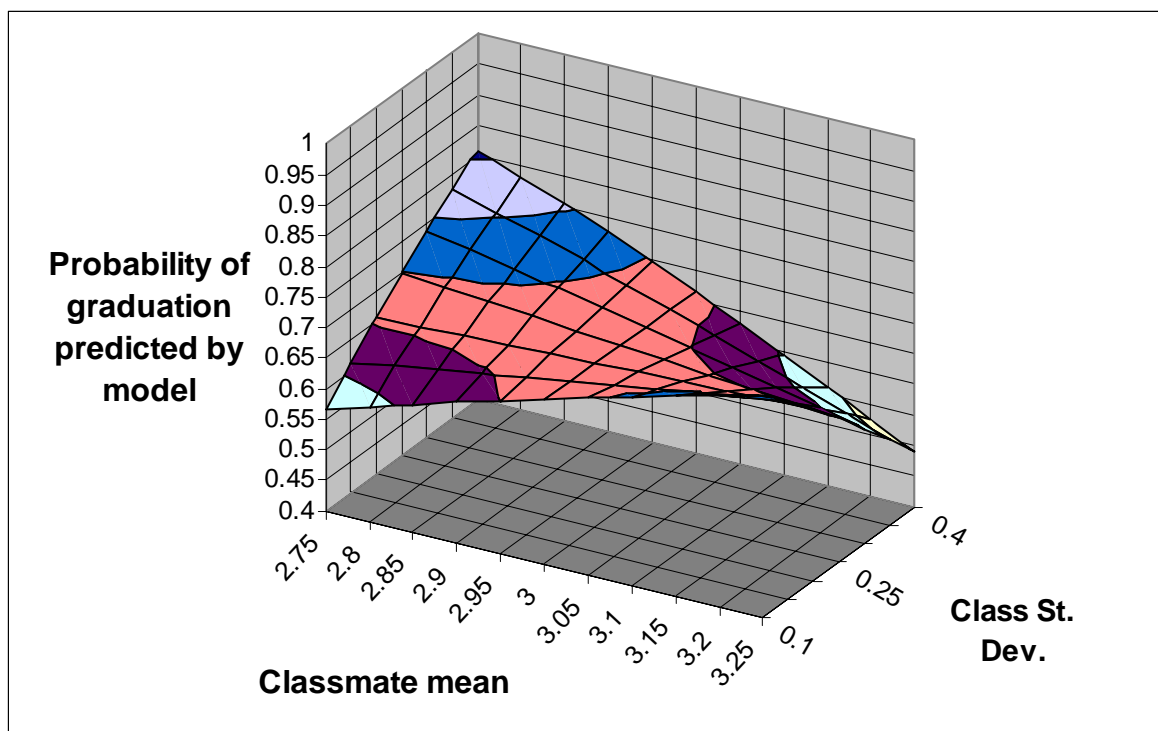


Figure 2. Effects of mean and standard deviation on probability of graduation

The results of probit regressions using subsets of the sample show similar qualitative responses, though the magnitude of the coefficients and their statistical significance vary. For all sub-groups, the coefficients of mean and standard deviation are positive with the interaction term having a negative coefficient. The effects of peers on probability of

¹³ The controls for major were excluded.

¹⁴ The nonlinearity of partial effects associated with the cumulative normal distribution function is present, but not visually apparent, in the results shown in Figure 2.

graduation are stronger for female students than for males and are stronger for the students in the middle of the predicted-GPA distribution than for those at either end.

5. INTERPRETATION OF ECONOMETRIC RESULTS

The measured effects of Humanities 110 peers on Reed freshmen's academic achievement do not yield a simple interpretation. Classmates who are more academically able do not seem to lead directly to higher achievement unless they are set in a class with relatively low variance of abilities. A low average classmate ability has a negative unless there is wide variation in abilities within the class. This pattern holds whether achievement is measure by grade-point average or by persistence to graduation.

It is not surprising that students in a low-mean, low-variance conference have lower-than-predicted achievement in their other studies. The low-mean, low-variance conference is likely to have very few above-average students and may not provide positive peer effects through the intellectual stimulation of excellent classmates. Adding even a few stronger students to raise variance and mean is likely to improve the educational environment.

Similarly, the high-mean, low-variance conference is a section with mostly high-quality students. It is not surprising that such an environment stimulates its members to above-expected achievement. Combined with the previous result, this suggests that a conference with relatively homogeneous abilities will tend to push the achievement of its members even further in the same direction than would have been expected from an average humanities experience. Good students together do better than expected; weak students together do worse than expected.

The result that students in a high-variance conference do better (relative to expectations) if the mean is low is more difficult to explain. If we think of an example with some very strong and some very weak students present in the same conference, this result suggests that the average student's subsequent performance improves if the weaker students are in the majority and deteriorates if the majority is academically strong.

The significant effect of more female presence in the conference on achievement of female classmates is striking. Educators have long recognized that female students may be overshadowed in class participation by their more aggressive male counterparts. Given that the purpose of the Humanities 110 conference is to enable students to discuss the reading

material, this appears to be a setting in which such effects might be especially strong. The most academically able female students seem to be capable of “holding their own with the boys,” as suggested by the positive (but insignificant) coefficient for the top 15% of female students. It is the female students “in the middle” in terms of ability who seem to thrive on having more female classmates in their humanities conference. This evidence seems strongly consistent with the idea that some female students’ participation and learning may be curtailed by male classmates.

6. CONCLUSIONS AND FURTHER RESEARCH

We have presented results that suggest that the academic abilities of humanities classmates may have significant effects on the academic performance of Reed students in other classes. These effects depend in a nonlinear way on the mean and standard deviation of peers’ abilities.

Although these nonlinear effects of classmate mean and standard deviation shown above are quite robust across outcome measures and sub-sample, one must draw conclusions carefully. These results emerged after extensive econometric testing with alternative functional forms; none of the specifications with mean or standard deviation alone (without the interaction term) yielded consistent evidence of classmate peer effects.

A positive interpretation of these results, which we cautiously adopt here, is that the recognition of the interaction between classmate mean and standard deviation is crucial to the identification of their effects on academic performance. However, one must also consider the possibility that the statistical significance of these results is merely the result of having continued searching specifications until happening onto one that “worked.” Clearly, more research is needed to examine the robustness of these results. Indeed, given that we are looking at the effects of peers in Humanities 110 on measured academic achievement *in other courses*, one might expect peer effects to be difficult to detect.

We have obtained data for a similar gateway course at another liberal-arts college, in which students are allocated randomly into sections. Analysis of these data may help clarify the effects of the classmate ability distribution on student achievement. We hope to proceed with this analysis in the near future.

Beyond this, an ideal setting for examining classroom peer effects would allow one to examine learning *within the course* as a function of the distribution of peer abilities. In order to conduct such a test, one would need a course in which allocation to sections was random (or based on observable characteristics) but in which grades in the course were assigned according to a standard that did not vary across sections. It would be crucial as well to control for the instructor in such a setting. We are exploring a possible example of such a course at Reed College.

Peer effects are important to the economics of higher education for reasons demonstrated clearly by Rothschild and White (1995). While roommate studies may hint at the importance of peer quality, classroom effects are more central to the mission of higher education. Few parents are likely to send their students to a particular college because they are likely to have a smart roommate. Further study of classmate effects is important to elucidate further the empirical relevance of peer effects.

REFERENCES

- Checchi, Daniele, and Francesco Zollino. 2001. Sistema scolastico e selezione sociale in Italia. *Rivista di Politica Economica* 91(7-8), 43-84. (with English summary)
- Coleman, James S., et al. 1966. *Equality of Educational Opportunity*. Washington, D.C.: U.S. Department of Health, Education, and Welfare, Office of Education.
- García-Diez, Mercedes. 2000. The effects of curriculum reform on economics education in a Spanish college. *Education Economics* 8(1), 5–15.
- Henderson, Vernon, Peter Mieszkowski, and Yvon Sauvegeau. 1976. Peer group effects and educational production functions. Ottawa: Economic Council of Canada.
- Hoxby, Caroline M., and Bridget Terry. 1999. Explaining rising income and wage inequality among the college educated. Cambridge, Mass.: National Bureau of Economic Research Working Paper 6873.
- McEwan, Patrick J. 2003. Peer effects on student achievement: Evidence from Chile. *Economics of Education Review* 22(2), 131-41.
- Manski, Charles F. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economics and Statistics* 60, 531–542.
- Robertson, Donald, and James Symons. 2003. Do peer groups matter? Peer group versus schooling effects on academic attainment. *Economica* 70(277), 31–53.
- Rothschild, Michael, and Lawrence J. White. 1995. The Analytics of the Pricing of Higher Education and Other Services in which the Customers Are Inputs. *Journal of Political Economy* 103(3), 573–586.
- Sacerdote, Bruce. 2001. Peer effects with random assignment: Results with Dartmouth roommates. *Quarterly Journal of Economics* 116(2), 681–704.
- Weiss, Michael J. 1989. *The Clustering of America*. New York: Harper & Row.
- 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 Caroline H. Hoxby. Chicago: National Bureau of Economic Research and University of Chicago Press.
- 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.