MPRA
Munich Personal RePEc Archive

Bad Company: Reconciling Negative Peer Effects in College Achievement

Ryan Brady and Michael Insler and Ahmed Rahman

United States Naval Academy

November 2015

Online at https://mpra.ub.uni-muenchen.de/68354/
MPRA Paper No. 68354, posted 13. December 2015 08:42 UTC

Bad Company: Reconciling Negative Peer Effects in College Achievement[*]

Ryan R. Brady[†]
Michael Insler[‡]
Ahmed S. Rahman[§]

Department of Economics
United States Naval Academy

November 2015

Abstract

Existing peer effects studies produce contradictory findings, including positive, negative, large, and small effects, despite similar contexts. We reconcile these results using U.S. Naval Academy data covering a 22-year history of the random assignment of students to peer groups. Coupled with students' limited discretion over freshman-year courses, our setting affords an opportunity to better understand peer effects in different social networks. We find negative effects at the broader "company" level—students' social and residential group—and positive effects at the narrower course-company level. We suggest that peer spillovers change direction because of differences in the underlying mechanism of peer influence.

Keywords: Peer effects, social network formation, academic achievement, homophily
JEL Codes: D85, I21, I23, I26, J24

[*]Many thanks to James West, Darrell Glaser, and participants of the Alan Stockman Conference at the University of Rochester in October 2014 for their helpful feedback.
[†]Associate Professor of Economics, U.S. Naval Academy. 589 McNair Rd., Mail Stop 10D, Annapolis, MD 21402. Phone: (410) 293-6883. Email: rbrady@usna.edu.
[‡]Assistant Professor of Economics, U.S. Naval Academy. Phone: (410) 293-6881. Email: insler@usna.edu.
[§]Corresponding author. Associate Professor of Economics, U.S. Naval Academy. Phone: (410) 293-6880. Email: rahman@usna.edu.

1 Introduction

Economists have closely studied the role of peer effects at all levels of schooling but, so far, have had difficulty extending these insights into the policy arena. Research has demonstrated the prevalence of peer effects in higher education in particular. Studies of institutional data from Dartmouth College, Williams College, the University of Maryland, Berea College, the United States Military Academy (USMA), and the United States Air Force Academy (USAFA) have revealed peer effects of various sizes on a range of academic outcomes.[1] Some of most notable effects have been observed at USAFA where Carrell *et al.* (2009) estimate that a 100-point increase in the peer-group average SAT verbal score increases freshman students' GPAs, on average, by 0.4 grade points on a 4.0 scale.[2] In a follow-up study, Carrell *et al.* (2013) analyze a direct intervention in which the researchers themselves allocate incoming students into peer groups designed to positively influence academic marks, as predicted by their historical estimates of peer effects. The intervention, however, backfires; the targeted beneficiaries of the experiment experience statistically significant *reductions* in their grades. The implications of these findings are profound, suggesting that policy makers who seek to affect the composition of students, with the intent to improve student performance, are as of yet still peering into a black box.

In this paper, we attempt to unpack the complex process of peer group formation to better understand why researchers have measured so many disparate peer effects, positive and negative, large and small. To do so, we offer both an empirical examination and a conceptual model of peer interaction. For the former, we examine data on college freshmen at the United States Naval Academy (USNA), using a dataset that includes more than 100,000 fall semester grades from over 20,000 freshmen. In employing college level data, we follow in the tradition of studies cited above. Unlike previous studies on college students, however, the size of our dataset, which includes over twenty graduating classes spanning 1991 to 2012, is the largest examined thus far at the college level.[3] In employing service academy data, we also follow in the tradition of Lyle (2007) and Carrell *et al.* (2009), both of which exploit the random assignment of students into residential groups (called "companies") to identify peer effects. Unlike these studies, however, we explore how interaction depends on the size and purpose of the group, which may then lead to either positive or negative peer effects in different situations.

[1]See Sacerdote (2001), Zimmerman (2003), Foster (2006), Stinebrickner and Stinebrickner (2006), Lyle (2007), and Carrell *et al.* (2009).

[2]The researchers suggest that their estimates are larger than previous findings because of the size (approximately 30 students) and critical role of peer groups at USAFA compared to more narrow roommate linkages seen in other studies. But they also find peer group SAT math scores to be statistically insignificant to individual achievement.

[3]Notable datasets on primary and secondary students include those analyzed by Burke and Sass (2013) and Lavy *et al.* (2012). Also, Cornelissen *et al.* (2013) analyze a very large dataset of German retirees.

We provide theoretical motivation with a simple model of peer interaction where a student is motivated both by *homophily*, which pulls her towards individuals of similar attributes, and by academic spillovers from peers, which may pull her towards a different group of individuals. The model illustrates why—in situations with broad social interaction—individuals move towards more similar peers, while situations requiring "narrow" task-based interactions can induce individuals to collaborate with peers who have different attributes. These interactions in turn affect individual performance. The insights of our model are evocative of explanations found in a variety of peer group studies, including the ideas of "invidious comparison" and the explanation of academic outcomes captured by the "boutique" model of peer interaction (see Hoxby and Weingarth (2005) for discussion; we expound on these ideas in section 5.4).[4]

We produce two primary empirical findings. First, we find *negative* peer effects at the full company level (which defines a students' residential and social group). For STEM courses (science, technology, engineering, and mathematics), average peer ability across all freshman companymates, as measured by both verbal and math SAT scores, negatively affects own grades. The negative peer effect measured for both SAT categories is a rare result in the peer effects literature; it stands in contrast, for example, to a positive effect found for verbal SAT scores by both Zimmerman (2003) and Carrell *et al.* (2009). Yet our negative peer effects at the company level are consistent with the findings from Carrell *et al.* (2013)'s hands-on experiment at USAFA. Moreover, the negative effects are consistent with the notion of homophily captured in our conceptual model. Our negative result comes from models with both linear and non-linear specifications, and is robust to numerous sensitivity checks. Second, we find that at the company-course level—students in the same company *and* the same course—average peer ability *positively* affects student performance, but only for relatively small companymate-coursemate peer groups. This latter result is consistent with most of the papers cited thus far. Our contrasting results conform with the idea that students may avoid interaction in broader social settings, yet may benefit from closer interactions in smaller groups that perform specific common tasks.

The unique setting at USNA allows us to better understand the black box of peer interaction by exploiting differences in peer group interaction at both a broad group level (the company), and subgroup levels (specific course-company peer subgroups). Within various peer subgroups, the scope of student interaction can differ and potential spillovers can vary in intensity across peers (see Manresa (2013) for discussion of such a possibility). Companies at USNA are well suited for studying such peer effects, both because dorm selection is randomly assigned and because students spend an inordinate amount of time

[4]Our model might also be considered an alternative to the utility maximization problem found in Cicala *et al.* (2014) where peer effects arise solely from comparative advantage. Our model however embodies the idea that peer effects arise partly due to a spillover from interaction (since utility maximization is a function of peer ability), whereas in Cicala *et al.* (2014) the actual behavior of peers has no effect on an agent's utility.

with their companymates in many aspects of college life. While previous studies have looked at peer effects on college students at the dorm-room level, dorm-floor level, and at the service academy residential group level, ours is the first to examine peer effects within residential groups *and* at the course level. In such settings, there is a natural opportunity for students to collaborate on similar tasks with their neighbors. Importantly, students at USNA have no discretion either over the assignment of their companies or their course enrollment during their first semester of freshman year. Hence, we exploit the variation in pre-treatment peer ability at the company level (*i.e.*, the average SAT score across all freshman companymates), as well as the variation in pre-treatment peer ability at the *company-course* level to analyze how spillovers may differ across various observable subgroups of students. In other words, unique to other peer effects studies we measure peer effects within more narrowly defined subgroups that are clearly designated to engage in common tasks, as opposed to interacting in a broader social group while taking disparate courses.

Our paper helps reconcile the seemingly contradictory findings in the literature, where each study observes a different social network and thus uncovers a different type of social interaction. Our emphasis on subgroup formation within broader peer groups stresses the need to understand the context of social settings and how common tasks can magnify endogenous peer effects. Failure to take these factors into consideration when attempting to manipulate peer groups can otherwise produce disastrous outcomes. Our broader picture of peer interaction has potential implications for policy changes, and we expound on these themes in the remaining sections of the paper. Section 2 provides a brief review of the key insights from the literature of most interest to our paper, and it also offers a theoretical motivation for our empirical findings. In Section 3 we explain our setting and dataset in more detail. The final sections summarize the estimation strategy and results, and provide discussion.

2 Motivation and Conceptual Framework

2.1 Identifying Peer Effects

The measurement of peer effects is complicated by "reflection." Peer effects are the reflection of our own image as it is cast upon our friends, who in turn cast their image back upon us.[5] This reflection is a function of our respective backgrounds, our current interactions with each other, and the environment in which we interact. Manski (1993) categorizes each of these factors in turn as contextual effects, endogenous effects, and correlated effects. Endogenous effects are a function of simultaneity, where peer group members affect each other and the

[5]Manski (1993) deliberately uses the term "reflection" to evoke one looking in a mirror. An alien observer would not be able to tell if the image in the mirror was initiating movement, or if the person in front of the mirror was doing so.

observer cannot tell who is really affecting whom (see Glaser (2009) for succinct discussion).[6] Contextual effects stem from the predetermined attributes (innate ability or training) of the peer group. The term "contextual" comes from the sociology literature whereas economists would perhaps call it "exogenous."

Correlated effects refer to either common factors affecting a peer group or to the tendency of people to consort with those who they perceive as like-minded individuals. Common factors could include the quality of the teacher in a particular class, the air conditioner not working for an entire semester, and so on. The desire to associate with similar people is a selection problem. An individual may gravitate towards those with characteristics they fancy in themselves and hence in others (early risers, tidy, a love of classical music, and so on). This makes it difficult to identify a peer effect since the features of an individual may be correlated with why he or she is a part of that group to begin with (Lyle, 2007).

Consider the following structural form of the reflection issue where the researcher specifies the effect of peers on student i as follows:

$$GPA_{ict} = \beta_0 + \beta_1 \overline{GPA}_{jct} + \beta_2 \overline{PRE}_{jc,t-1} + \beta_3 PRE_{ic,t-1} + u_{ict} \tag{1}$$

GPA is the usual academic measure for grade point average, PRE is a measure of the exogenous "pre-treatment" ability the students bring with them to college (this is the "contextual" effect in Manski's parlance). The subscript indicates person i in peer group c in time period t. Subscript j denotes peer group members, where $j \neq i$. $\overline{GPA}$ and $\overline{PRE}$ are the averages of those peer group members' GPAs and pretreatment characteristics, respectively. The specification can be amended to control for various fixed effects (we discuss this more later with respect to our estimation). The parameter β_1 measures the endogenous peer effect while β_2 captures the contextual peer effect.

The inability to identify peer effects comes in three related forms. First, the unobserved selection process implies the error term will be correlated with the group pretreatment variable. Second, group performance at time t will be correlated with the error term via the endogenous effect (the simultaneity of peer influence). And third, measurement of β_2 may be further biased by common factors affecting all members of the group. Hence, in the structural equation both peer effect parameters β_1 and β_2 are likely to be biased.

The literature approaches the bias in different ways. A recent cadre of papers deals with the selection issue using "natural experiments" where peer groups are randomly assigned. Given the random assignment, the selection issue is mitigated. Relatively recent examples include Sacerdote (2001) and Lyle (2007); both estimate the structural equation (or some close version of it) shown in equation (1). Sacerdote (2001) estimates peer effects

[6] There tends to be some confusion in the literature with respect to Manski's terms. "Reflection" is sometimes listed as a fourth category. However, based on our reading of Manski (1993) and papers such as Glaser (2009) and Durlaf and Ioannides (2010), we believe our description to be consistent with Manski (1993).

with random dorm room assignments of freshmen at Dartmouth. He regresses a freshman's GPA on own level of ability (a "pre-treatment" metric that includes SAT scores and high school class rank), on roommate's pre-treatment ability, and on roommate's GPA. In this case Sacerdote (2001) obtains an unbiased estimate of roommate's contextual background but not roommate's GPA. The key to identification for the former effect is the random assignment which limits any possible correlation between roommate pre-treatment ability and other factors that may influence freshman i's GPA—parental pressure, non-roommate peers, and other components of the error term. As long as those factors are uncorrelated with roommates' background then one has a "clean" estimate of the peer effect from the roommate's pre-treatment ability.

Sacerdote (2001) ultimately finds that roommate pre-treatment ability is not a statistically significant explanatory factor for freshman GPA—so there is no contextual peer effect. However, roommate GPA is a statistically significant predictor of freshman GPA, but the result is obviously plagued by simultaneity bias. Sacerdote however suggests that bias notwithstanding, the statistically significant coefficient provides evidence of a positive peer effect. That is, Sacerdote claims, the biased coefficient represents *some* peer effect, even if the measurement is afflicted by simultaneity.

Lyle (2007) finds similar results to Sacerdote (2001) using random assignment of freshmen to companies at USMA to control for selection in the estimation of the structural equation above. Rather than being assigned to dorms or dorm floors, freshmen at the USMA are sorted into companies, which are peer groups built from all class years (freshman, sophomore, junior, and senior). Lyle (2007) finds the endogenous peer effect is statistically significant, but the contextual effect is statistically insignificant. In his application, however, Lyle (2007) emphasizes that common shocks may explain much of the endogenous effect.[7]

Given the inherent bias in measuring the endogenous effect, both Sacerdote (2001) and Lyle (2007) also estimate reduced form versions of the structural equation shown above. The reduced form can be written,

$$GPA_{ict} = \left(\frac{\beta_0}{1-\beta_1}\right) + \beta_3 PRE_{ic,t-1} + \left(\frac{\beta_2 + \beta_1\beta_3}{1-\beta_1}\right)\overline{PRE}_{jc,t-1} + \varepsilon_{ict} \tag{2}$$

where the parameters map to the structural equation shown in equation (1).[8] Lyle (2007) does not find a statistically significant "reduced form peer effect" (*i.e.,* he cannot reject that the coefficient on $\overline{PRE}$ is zero). Sacerdote (2001), however, finds some evidence in one version of the reduced form model, where having high ability roommates suggests a positive peer effect.

[7]Previously, Sacerdote (2001) had recognized the possibility of common shocks. He imposed dorm-level fixed effects, which made little difference in his results.

[8]See Manski (1993) for more details on the mapping between the structural model and the reduced form in this literature.

More recent studies—such as Guryan *et al.* (2009), Zimmerman (2003), Stinebrickner and Stinebrickner (2006), Foster (2006), and Carrell *et al.* (2009)—eschew the structural form altogether and focus solely on the reduced form. The latter use random assignment of freshmen into "squadrons" at USAFA to search for peer effects in the reduced form setting. Their measure of peer pre-treatment—verbal SAT scores—is statistically significant and positive at the squadron level, yet weak at the roommate level. Peers' average verbal SAT score is statistically significant predictor of academic performance for math and science courses, but strangely not for language courses. Also counter-intuitively, peers' average math SAT score is not statistically significant for math and science courses.

The result in Carrell *et al.* (2009) is consistent with numerous findings that the composition of peer groups is important but also produces some unintuitive results. Zimmerman (2003) cites research on primary school students, whose performance increases with average classroom IQ, but at a diminishing rate. Zimmerman (2003) suggests that mixing students of varying ability (rather than segregating students) should generate higher aggregate learning. Indeed, Carrell *et al.* (2009) find that low ability students benefit from proximity to high ability peers (more than average ability students do).

2.2 Negative Peer Effects and Homophily

With group composition in mind, Carrell *et al.* (2013) implement an "optimal" peer assignment experiment at USAFA to harness the positive spillovers found in Carrell *et al.* (2009). The peer group assignments were meant to improve the performance of the lowest ability students, where the primary treatment group ("bimodal squadrons") includes low ability students alongside a larger fraction of peers with high SAT verbal scores. A second treatment group ("homogeneous squadrons") is comprised of primarily "middle ability" students.

Contrary to most evidence from the empirical literature, Carrell *et al.* (2013) estimate a *negative* treatment effect for the main treatment group of interest—the low ability students perform worse than similar students in the control group. Conversely, students in the homogenous treatment group perform better than their counterparts in the control group.

The authors suggest that homophily effects explain the results. The low-ability students in the treatment group were more likely to study with—and identify as friends—other low-ability students in their squadron (friends were identified via a follow up survey conducted by the authors). The higher ability students in the treatment squadron segregate themselves similarly. The middle ability students, on the other hand, appear to benefit from the homogeneity of their group. In the control group, the tendency for such sorting was less apparent. Another example of a negative estimate is from Foster (2006). For a sample of University of Maryland students, she estimates a statistically significant negative peer effect on male students that stems from their peers' median SAT scores (but not from peers' average SAT scores). Using data on Florida's public school system, Burke and Sass (2013) find that peer effects are negative for low ability students if they are grouped with students at the high end of the distribution.

It thus appears that even with random assignment, the literature has yet to reach a consensus on peer effects. Likely, different social networks which form in different contexts complicate inference (Jackson, 2008). Unfortunately, the nature of this sort of endogenous group formation is a "black box," as Carrell *et al.* (2013) note towards the end of their paper. We can, however, conjecture and empirically examine how the push/pull of homophily might lead to different outcomes in different settings.

2.3 A Simple Framework

To make more concrete some of the ideas mentioned above, let us consider a student who cares about only two things, her grades, defined as G, and her "homophily index," defined as H. Assume each of these depend on the student's own innate characteristics, and *potentially* the characteristics of her peers. Let us further assume that if the student interacts with the peer group, the student's homophily index erodes away as her characteristics become more *different* from the group's. On the other hand, with social interaction, the student's grades will tend to converge to the average of the group. That is, if the student is relatively weak, her grades will improve with more interaction. If the student is relatively strong, her grades will deteriorate with more interaction.

More specifically, assume that the student aims to maximize U, where

$$U = G^{\alpha} H^{1-\alpha} \tag{3}$$

and $0 < \alpha < 1$. Given a certain group of other students, the student chooses with whom to interact. That is, the student chooses both the number and type of people to form her peer group in order to maximize equation (3). Let us consider *ability* as the sole characteristic that matters, and that ability is innate. a_0 is the student's ability, while a_i is peer i's ability. Let us then have following functional forms:

$$G = \left(a_0 + \gamma \sum_{i=1}^{N} (a_i - a_0) \right)^{\psi} \tag{4}$$

$$H = a_0 + N - \phi \sum_{i=1}^{N} |a_i - a_0| \tag{5}$$

where N is the number of the student's *chosen* peers (which conceivably could be zero), $\gamma > 0$ measures the degree of ability spillover to the student's grade performance, $\phi > 0$ measures the degree of distaste for ability differences, and $0 < \psi < 1$. The meaning of each measure is straight-forward. Equation (4) suggests that interacting with peers of stronger ability helps the student achieve better grades, although at a diminishing rate. Interacting with weaker ability students on the other hand deteriorates the student's performance. If the student chooses not to interact at all, then she relies solely on her own ability a_0. On the other hand, equation (5) suggests that while the student is a "social" creature and so derives utility from having more peers in general (hence the inclusion of N),[9] interacting with peers of relative stronger *or* weaker ability erodes one's homophily index, thus lowering overall utility. The student's decision on the number and type of other students to interact with is essentially based on which effect dominates.[10]

2.4 Numerical Examples

Let us demonstrate the simple framework above with some numbers. First consider a student with low relative ability $a_0 = 1$. She has three potential peers, each who range in ability from $2 \geq a_i \geq 2.5$. Assume that $\alpha = \psi = 0.5$, and that $\gamma = \phi = 1$. For a given potential group with given characteristics, the student has eight possible choices to maximize equation (3): befriend no one (and receive a grade of a_0^ψ), befriend one of four students, befriend two students (for which there are three possibilities), or befriend all three.

The top portion of figure 1 demonstrates the effects on the student's grade as the average ability of the peer group increases (we increase ability uniformly across all three of the other students). For the low-ability peer group her grade is highest—she befriends all three because her homophily index is fairly large, and her grade is better because her peers still have higher ability than she does. However, as average peer quality improves, she starts cutting people out of her life because her homophily index starts falling. As her peer group shrinks from three to two to one, her grade suffers, and each drop is larger than the next. Finally with a large enough average group quality, she is left alone to earn a grade of one.

Now consider a relatively high-ability student. In this case suppose the number and average peer group quality is the same, but now the student has an ability of $a_0 = 2.5$. The bottom portion of figure 1 demonstrates the effects on this student's grade as the average ability of the peer group increases across the same range as before. For the low range of peer ability, the high-ability student chooses to be isolated from the others and earns a grade of a_0^ψ. As the average quality of the peer group improves, the high-ability student befriends more and more individuals. But because each new peer is still of lower quality than the

[9] Qualitative results are not sensitive to the inclusion of N in the homophily index.

[10] Alternate approaches to modeling peer effects are of course possible, perhaps most notably in Cicala *et al.* (2014). In their model individuals sort into peer groups based on comparative advantage. We believe our approach is more appropriate for this study, for two reasons. One, Cicala *et al.* (2014) do not model peer spillovers. While an appropriate omission in certain contexts, we believe academic spillovers to be a potentially important effect of peer interaction. Two, our empirical evidence suggests that one's ordinal rank within one's company matters far less for one's academic performance than the *average* ability of the peer group.

student, her grades suffer with each new friend. With quality high enough, she befriends the whole group; after that point any further quality improvements for the group raises her performance.

We thus see that there can exist a range of peer ability such that increases in average peer ability can hurt individual performance for *both* high-ability and low-ability students. For low-ability students, negative peer effects can be seen as arising from an "invidious comparison" effect suggested by Hoxby and Weingarth (2005). For high-ability students, negative peer effects can be viewed as a somewhat novel take on the "bad apple" model, again mentioned in Hoxby and Weingarth (2005). Ironically, we show here that bad apples can pull down performance, even as these apples become *less* bad.

We can also observe that these effects will be sensitive to the degree to which homophily preferences matter. To see this we redo the above exercises, but here we lower the degree of distaste for ability differences, ϕ, to 0.1. Conceivably this may arise with a change in the nature of personal interactions. Results for both low and high-ability students are displayed in figure 2. Here the low-ability student remains friends with all three peers, since grade spillover benefits now outweigh negative homophily effects from peer quality improvements. As such her performance monotonically rises with group quality. On the other hand, the high-ability student moves from one to two to three friends, and her grade performance shows ups and downs with no overall trend.

Thus we see that when dis-utility from group heterogeneity is limited, higher average peer quality can raise own performance. This would be consistent with the "shining light" model of peer effects (Hoxby and Weingarth, 2005).

This simple framework demonstrates that endogenous peer group formation matters even when "formal" groups of potential peers are exogenously created. Perhaps more importantly, the model provides us with a cautionary tale: positive peer effects can be harnessed, but only in situations where the dis-utility from low homophily is not too great. In order to implicitly test this framework, we require different social settings of interactions where ϕ would conceivably vary. Can we find instances where students, thrust together in common purpose, embrace their differences instead of shielding themselves from them? Data derived from features of USNA, described below, provide us such a setting. The following sections test some of the implications presented above with evidence from freshman students enrolled at USNA.

3 Our Setting

3.1 Data

Our data, which were compiled with the aid of USNA's Office of Institutional Research, contain far more observations—more than 100,000 grades from over 20,000 first-semester freshmen-–than comparable studies from other college-level institutions.[11] In addition to every grade assigned to freshmen in the classes of 1991-2012, we observe the following student-specific characteristics: race/ethnicity, whether a recruited athlete, whether previously enlisted in the armed forces, whether attended a one-year preparatory school prior to enrolling at USNA,[12] and math and verbal SAT scores. Table 1 contains summary academic and demographic statistics for freshman students. While a relatively white and male student body, a fair representation of minority and female students exists, especially during the latter years of our sample.

3.2 Company Random Assignment

USNA provides an ideal setting to identify the effects of social interactions on academic achievement. Upon arrival, every freshman is assigned into a company. All students live in one on-campus dormitory, which houses 30 companies of approximately 150 students, each containing an even mix of freshmen, sophomores, juniors, and seniors. A student's company makes up his or her basic group of potential peers. The company assignment procedure, which is administered by the Admissions Office, is designed to produce a demographically diverse but otherwise randomly allocated mix of students in each company. Students are first stratified according to certain predetermined characteristics: race, gender, home state, prior military service, and attendance at a one-year Naval Academy preparatory school. Once administrators ensure balanced representation among these characteristics across all companies, they randomly assign all remaining students to companies.[13] The key features of the procedure are that students have no control over the outcome—USNA does not solicit interests, lifestyle details, or roommate preferences as is typical at other universities—and it produces an allocation that is effectively random. The mechanism prevents students from sorting into residence hall groupings that could offer academic or personal advantages.

If companies are randomly assigned, then we should not be able to predict own SAT scores from freshman companymates' average SAT scores, conditional on the strata mentioned above (gender, race/ethnicity, etc.). We estimate two OLS regressions for each subsample of freshmen from every graduating class 1991-2012: conditional on the predetermined characteristics that affect company assignment, we regress (1) own SAT math score on freshman companymates' average SAT math score, and (2) own SAT verbal score on freshman companymates' average SAT verbal score. Table 2 presents the results. Two out of the 44 regressions show problematic correlations (at $\alpha = .05$), a proportion that is in line with a five percent chance of Type I error. Our findings confirm that the assignment of

[11] For comparison, Carrell *et al.* (2009) utilize a sample of approximately 20,000 grades.

[12] See Kotlikoff et al. (2015) on the effects of these specific programs on educational outcomes.

[13] This procedure is very similar to USAFA's, described by Carrell *et al.* (2009). It differs from the procedure at USMA, described by Lyle (2007), which additionally produces an even mix of academic ability (proxied by incoming SAT score) across companies. Lower variation in academic achievement at USMA across companies yield an environment in which it is much more difficult to estimate peer effects.

freshmen to companies is random.

3.3 Course Random Assignment

In addition to the random assignment of peer groups, there are other features that make USNA an ideal laboratory in which to examine peer effects. All freshmen must pass or validate a set of 11 core courses in a range of subject areas such as calculus, chemistry, political science, and naval history. With little variation, these courses form their entire first year schedule.[14] Importantly for our study, freshmen at USNA cannot select whom they take a course with (either instructor or peers) during the fall semester. As part of the students' "Plebe Summer"—the summer-long indoctrination of the freshman class—fall semester courses and sections are selected for students.[15] Indeed, students only learn of their schedules in mid-August, just prior to the beginning of the semester.

The course schedules are determined unilaterally by the Registrar's office. Once results from validation or placement exams administered during the summer are considered, the registrar generates freshman schedules for the fall semester.[16] As part of that process, some course selection is simply a function of whichever company to which a student is assigned. For example, in the fall semester all freshmen in 1^{st} through 15^{th} Companies will take American Government, and the rest will take American Naval History. Hence, attempting to select a course to be with one's friends, or for a particular instructor that an upperclassman recommends, is simply not possible at USNA.[17] Later, in Section 5.3, we provide some statistical support for the randomness of course assignment.

Lastly with respect to our dataset, USNA has relatively low grade inflation. Based on our sample, average GPAs have risen over the years, from an average of 2.7 in 1991 to 3.0 in 2009. Rojstaczer and Healy (2010) show that average GPAs from a large sample of American four-year colleges have increased approximately linearly from 2.9 in 1991 to 3.1 in 2006. Thus the trend at USNA is steeper but averages remain lower, compared to other schools. Given that GPAs are far more frequently bounded from above than below, this suggests both that there is higher grade variation at USNA and that grades produce higher signal-to-noise ratios than at other institutions.

[14] In fact, freshmen are restricted to only 11 courses in the first year. Only with direct approval from the Academic Dean can a freshman take a twelfth course (and a student at most can take six courses during the fall semester).

[15] "Plebe Summer" begins in late June and runs up to the beginning of the fall semester in late August. It is USNA's version of boot camp for the incoming freshman class. Freshmen are commonly referred to as "Plebes" at USNA.

[16] The validation or placement exams are the only form of indirect student input in the process. For example a subset of students will validate Calculus I via AP scores or a placement exam and instead start with Calculus II during the fall semester. Nevertheless, regardless at whichever level the student places, they cannot choose the particular section of the course.

[17] The information described in this section is from USNA's "Plebe Academic Handbook," provided to freshmen during Plebe Summer, and USNA's "Plebe Advising Handbook," which is provided to faculty members serving as academic advisers to the students during their freshman year.

4 Econometric Model

4.1 Baseline Model

We envision a freshman's first semester academic grades following a "production process" with inputs: (1) own high school (*i.e.,* pre-USNA) characteristics; (2) peers' high school characteristics; and (3) variation that is specific to course and academic year. Given the similarities between USNA and USAFA, we adopt a specification similar to that of Carrell *et al.* (2009). We use the following linear model:

$$G_{igct} = \alpha + \beta Z_{igt} + \gamma \frac{\sum_{k \neq i} Z_{kgt}}{n_{gt} - 1} + \delta X_{igt} + \theta Y_{ct} + \eta_t + \varepsilon_{igct}. \tag{6}$$

In this specification, G_{igct} is grade (on a standard four point scale) of student i in peer group g (*i.e.,* company g) for course c in academic year t.[18] We only use grades from the fall semester of freshman year to avoid issues related to potential self-selection into courses in subsequent semesters. Z_{igt} includes student i's pre-USNA characteristics that may directly affect academic achievement, for which we proxy with SAT math and SAT verbal scores. Z_{kgt} represents SAT math and verbal scores for student k, who is one of i's companymates within company g. γ captures the influence of i's "average" companymate, as we average over these two characteristics for all $k \neq i$. X_{igt} is the set of pre-USNA controls for student i: race/ethnicity, gender, whether a recruited athlete, whether attended a feeder school, and whether he or she possessed prior military experience. Y_{ct} includes information that is specific to course-years (which can be accounted for via a fixed effect) and η_t is a set of academic year dummies. ε_{igct} represents all omitted factors. By construction, ε_{igct} is uncorrelated with Z_{kgt} because peer assignment is random and all peer characteristics in Z_{kgt} were established before arriving at USNA (*i.e.*, prior to treatment).

4.2 Incorporating Coursemate Subgroups into the Model

For the model in equation (6), there remains an interpretation issue that, if not properly understood, can lead to unanticipated effects from policy interventions as seen in Carrell *et al.* (2013). We cannot distinguish what is driving the value of γ. Recall from Section 2 the reduced form parameter is an amalgam of the structural parameters that identify the endogenous effect β_1, the contextual effect β_2, and the own-pre-treatment effect β_3 (see equation (2)). That is, the reduced form peer effect that we can estimate is:

$$\gamma = \left(\frac{\beta_2 + \beta_1 \beta_3}{1 - \beta_1} \right). \tag{7}$$

[18] Academic year t does not represent the time period in the usual context of panel data, because we use each student's grades only from his or her initial semester. In other words, there is only one time period t for each student i, but we include the subscript to indicate possible academic year effects (*e.g.,* grade inflation).

Our conceptual framework in Section 2 indicates that negative peer effects may arise when students choose to group themselves by ability within a broader group, and this effect becomes more pronounced as the preference for homophily rises. In that setting, it is not clear if the negative peer effect arises due to the simultaneous peer interaction (the endogenous effect) or from the contextual effects.

An important distinction in understanding the possible mechanism driving positive or negative peer effects is the social multiplier discussed by Manski (1993), Glaeser *et al.* (2003), and Durlaf and Ioannides (2010) where positive peer influence occurs simultaneously. Also, Sacerdote (2001) notes, "positive student behavior leads to more positive behavior." Glaser (2009) explains that there is no feedback with respect to contextual features since the factors are predetermined. Instead, any contextual peer effects must occur due to variation in the levels of pretreatment ability among the group. Is there a circumstance where the social multiplier—the endogenous effect β_1 in equation (1)—might have a negative sign?

Consider the endogenous effect captured by β_1. For β_1 to be negative it must be that the high performance of the peer group somehow leads to individual students' performance, on average, being diminished by interacting in time t with the high-performers (controlling for the contextual differences). At first glance, one might imagine a scenario where a member of the group is hazed by other members of the group, so much so their performance suffers. Or more benignly, perhaps there exists a "rebel" scenario where one rebel's performance declines as high-performers' behavior defines the group. We think it is reasonable to suggest that such scenarios are unlikely to dominate, on average. Instead, it seems likely that any idiosyncratic negative effects would be dominated by other, positive contemporaneous effects. It is also reasonable to assume that $\beta_1 < 1$, since peers' contemporaneous achievement should not have a multiplied impact on own achievement. Therefore, we assume:

Assumption 1. $0 \leq \beta_1 < 1$

Second, consider the coefficient on own ability, β_3. It is very reasonable to assume that a student's own predetermined ability is positively correlated with his or her grades:

Assumption 2. $\beta_3 > 0$

In equation (7), under Assumptions 1 and 2, a negative "total peer effect" γ can only arise from a negative contextual effect β_2. On the other hand, a positive estimate of γ could be produced by either a positive β_2, or a negative β_2 that is drowned out by relatively strong β_1 and β_3 parameters.

The key insight here is that our conceptual model (Section 2) and the discussion of structural estimation (above) are compatible. Both suggest that any negative estimate of γ must stem from a negative contextual effect, such as sorting on predetermined characteristics. On the other hand, any positive estimate of γ could stem from a positive contextual and/or a positive endogenous effect. This observation has important policy implications, which we discuss in Section 5.4.

Given the natural experiment at USNA, we can consistently estimate γ but cannot identify its three components separately. However, we harness USNA's joint random assignment into companies and course sections, which allows us to test whether higher-quality peers may either positively or negatively influence academic achievement, depending on peer group setting. Or in other words, we can test that the contextual effect can differ in sign, as the conceptual model of Section 2 suggests. In particular, we focus on one possible channel through which peer effects may operate: collaboration between students in the same company *and* the same course. That is, the peer effect now stems from companymate-coursemates, rather than just companymates. We modify equation (6) as follows:

$$G_{igct,s} = \alpha_s + \beta_s Z_{igt,s} + \gamma_s \frac{\sum_{k \neq i} Z_{kgct,s}}{n_{gct,s} - 1} + \delta_s X_{igt,s} + \theta_s Y_{ct,s} + \eta_{t,s} + \varepsilon_{igct,s} \tag{8}$$

We estimate equation (8) on subsamples of the data that are stratified by the sizes of companymate-coursemate peer groups. To reflect that the size and direction of peer effects may differ depending on the composition of one's peer group, equation (8) introduces a subscript for peer group sizes, s.[19] Smaller companymate-coursemate groups suggest a special more narrow relationship affecting performance. As the companymate-coursemate group size increases, this relationship dilutes. In the most extreme case, the companymate-coursemate group is the same as the companymate group (*i.e.,* everyone in the company takes the same course). Here, we are simply back to estimating companymate effects, as in equation (6). Figure 3 shows the distributions of these peer groups' sizes for humanities/social science and math/science courses, separately. Given the distributions, it is reasonable to use peer group stratifications of sizes 2-10, 11-30, and over 30, for s.

Based on the conceptual model in Section 2, our prior is that γ from equation (6) is less than γ_s from equation (8) for more narrowly-defined peer groups s. By isolating smaller peer groups whose students are engaged in course-specific tasks, peer ability would conceivably matter more for positive grade performance due to greater peer interaction. In the conceptual framework, this would be consistent with a lower distaste for peer differences, ϕ; students of disparate backgrounds would be more likely to engage in narrow and focused interactions and thereby benefit from higher-ability peers.

[19] Also note that the peer effect variable $\frac{\sum_{k \neq i} Z_{kgct,s}}{n_{gct,s}-1}$ now includes subscripts for shared course c.

5 Results

5.1 Baseline Model

Table 3 contains OLS estimates of equation (6). We cluster standard errors by academic year-company groups and estimate separate regressions for math or science (henceforth math/sci) course grades, and humanities or social science (henceforth hum/ss) course grades. Each provides us roughly 50,000 grade observations.

Estimates of all control variables have reasonable signs and magnitudes consistent with prior research on academic performance. Women perform better than men in humanities and social sciences, but worse in math and science courses. Minorities, recruited athletes, and preparatory school attendees tend to under-perform in the classroom, likely due to selection. Freshmen who were previously enlisted in the armed forces earn math and science grades that are higher, on average, by 0.16 grade points.

Turning to our measures of background academic ability, we unsurprisingly find that own verbal and math SAT scores have strong statistically significant effects on performance for both course types. Specifically, 100 additional points in one's verbal SAT score is associated with an increase in one's own grade in a hum/ss class by a third of a letter grade (*e.g.,* from a B to a B+), with smaller results for math SAT scores. On the other hand, a 100 point increase in one's math SAT score corresponds to a higher grade in a math/sci course by half a letter grade, with smaller results for verbal SAT scores.

Results become more startling when we look at the average SAT scores for one's peer group, as measured by fellow freshman in the same company. Estimated coefficients on these measures are all negative. Effects from average peer verbal SAT scores are negative and statistically significant for both hum/ss and math/sci courses. Effects from average peer math SAT scores are negative and statistically significant for only math/sci courses. For example, a 100 point increase in average peer math SAT scores produces a fifth of a letter grade *lower* own performance in math/sci courses.

Columns 2 and 5 include the standard deviation of peer's verbal and math SAT scores. Lyle (2007) suggests that peer heterogeneity may affect individual performance. Here we see that estimated coefficients on peer SAT standard deviation are all statistically insignificant. Further, their inclusion does not meaningfully alter our estimates of peer effects from average SAT scores.

Columns 3 and 6 take a non-linear in means approach, as used in Carrell *et al.* (2013). Here the peer effect variable is the fraction of a student's potential peers in the top and bottom SAT score quartiles of her graduating class. Findings echo our basic results on the negative effects from higher peer ability. Degrees of statistical significance vary, but in all cases, a larger fraction of high-ability peers harms academic performance, while a larger fraction of low-ability peers benefits it.

Though modest in size, our estimates of negative peer effects at the company level seem unshakable, while the positive effects discussed in other studies appear fragile to alternative specifications. Many other studies fail to estimate statistically significant effects, perhaps because they use far fewer grade observations. Aside from Carrell *et al.* (2009)'s result for peers' verbal SAT scores, they find that peers' math SAT scores are statistically insignificant with respect to a freshman's GPA. Lyle (2007) also finds no effect from peers' math SAT scores (he does not report results for verbal scores). Nor does Zimmerman (2003) find a statistically significant effect for peers' math SAT in his primary estimation, though the estimated effect does have a negative sign.

On the other hand, our reduced form estimates are evocative of the negative outcomes of the peer group experiment at USAFA implemented and discussed by Carrel *et al.* (2013). As described earlier, Carrell *et al.* (2013) suggest negative effects from grouping lower ability students with a large number of high-ability students stem from homophily considerations (in the peer group measured at the squadron level). Our company level results are consistent with this idea. Students may self-select in particular ways within companies and this behavior manifests in negative spillovers. Here we can argue that in a context where students are assigned into some broad peer group and may interact in many different academic and non-academic settings, any positive peer effects that might occur via studying are overwhelmed by the tendency towards homophily. We discuss further the possible mechanisms at work in Section 5.4 below. First, however, we consider a setting for peer effects where students are more likely to interact, performing specific and common tasks.

5.2 Coursemate Subgroup Model

To explore further the intuition of peer group formation, we estimate possible peer effects at the companymate-coursemate level.[20] We stratify by both course type (humanities/social science, math/science) and peer group size (2-10, 11-30, 31 or more), where freshman i's peer group—and there is potentially a different peer group for each of i's distinct grades—is defined by the set of freshmen companymates contemporaneously taking the same course. Table 4 contains OLS estimates of equation (8). Standard errors are clustered by academic year-company groups.

Across all models, estimates of the control variables' coefficients are consistent with previous findings. Striking differences appear, however, when comparing the companymate-coursemate-specific peer effects across different peer group sizes. For small peer group sizes (2-10), the results of which are shown in columns 1 and 4, we observe positive peer effects in

[20]We can also try to estimate peer effects simply at the coursemate level. This however produces difficulties in identification, mainly due to the large size of these groups. Average SAT scores for these large groups tend not to vary enough across courses (recall that all freshmen select essentially the same schedule of courses). These results, although they are uninteresting, are available upon request.

hum/ss courses when those peers have stronger verbal SAT scores, and we observe positive peer effects in math/sci courses when those peers have stronger math SAT scores. These estimates are significant at the 0.1 percent level but appear smaller than estimates in previous literature: A one standard deviation (11.7 point) increase in coursemate-companymates' average SAT verbal score yields, on average, a 0.013 point higher hum/ss course grade, and a one standard deviation (10.0 point) increase in coursemate-companymates' average SAT math score yields, on average, a 0.0074 point higher math/sci course grades. Carrell *et al.* (2009), for example, find that a one standard deviation increase in peers' average verbal SAT results in 0.05 additional grade points. Here it appears that the ability of peers who both take the same course and live in the same company positively influence academic outcomes.

As the cohort of companymates increases in size within a course, however, the positive peer effects disappear. There is modest evidence of a negative peer effect through math SAT for groups of size 11-30 in math/sci courses. For peer groups greater than thirty, the group compositions approach the full company of freshman students. This simply captures those who are all taking the basic core courses that most freshmen are required to take. Results are consistent with our framework discussed in Section 2. Smaller peer groups working on a narrow range of course-specific tasks are far more likely to interact with each other and influence one another's performance (in the conceptual framework this would be represented by a smaller ϕ). As the scope of social interaction becomes more narrow and task-oriented, differences in innate characteristics play a smaller role in causing negative homophily effects.

A possible criticism is that, since we argue that both types of peer groups—broad company and narrower course-company—influence grades, the course-companymate subgroup model might suffer omitted variable bias since it excludes company-wide peer effect controls. If we include both peer effect channels and observe that signs and magnitudes of coefficients do not change, then there is no such bias. Therefore it is instructive to estimate a "horse-race" model that contains peer effect variables at *both* full company and course-company levels, but still stratified by coursemate-companymate peer group size. Estimation of this model essentially produces the results of table 4, now adding the peer effect controls from columns 1 and 4 of table 3. We do not report these results here (available upon request), but they affirm the findings of both tables 3 and 4. Horse-race models preserve negative company-wide peer effects and positive coursemate-companymate effects in the small subgroups (both types of peer effects coefficients are statistically significant in these models).

5.3 Robustness Checks

We further explore these positive peer effects from coursemate-companymate groups by performing a number of alternative empirical exercises.

5.3.1 Non-linear Channels of Peer Influence

Here we examine if our findings differ when we use a non-linear specification for peer influence—the fraction of top and bottom quartile SAT performers in the same course-company group. Table 5 reports results. Our thesis on positive peer effects for small course-company groups is supported here. In column 1, a greater share of "high-verbal" peers, as proxied by a greater share of high verbal SAT performers, helps individual performance in hum/ss courses, while a greater share of "low-verbal" peers hurts individual performance. And we see that "low-math" peers on net further pull down grades for hum/ss classes.

For math/sci courses, we see that low-math ability peers in particular hurt individual performance. The estimated coefficient on high-math ability peers is positive but not statistically significant.

We also see that these results reverse as peer group size increases. For larger peer groups, our findings revert back to those in table 2 (columns 3 and 6), suggesting negative peer effects at the broader company level.

Another non-linear approach is to interact average peer quality with the quantile ranking of students' own SAT scores. This would conceivably allow us to observe potential differences of peer effects for relatively poor students versus relatively strong students (Burke and Sass, 2013; Lavy *et al.,* 2012). These results however provide us no additional insights; students at the low end of the distribution appear similarly affected by peer spillovers as those at the high end (specific estimations are not reported here but are available upon request). Thus consistent with our discussion from Section 2, peer effects appear similar across student ability levels.

5.3.2 Are Results Driven Merely by Subgroup Size?

Here check if this positive effect is simply an artifact of picking smaller subgroups within the larger company. We redo the exercise of table 4, now using a placebo peer group: for each grade G_{igct}, we define freshman i's peer group as his or her company g mates who are *not* taking course c. We stratify grades, as always, by humanities/social science courses and math/science courses. Table 6 displays results. For proper comparison, we also reproduce the same peer subgroup stratifications as in the main results of table 4. For example, columns 2 and 6 use observations of G_{igct} only when student i had between two and ten companymates in course c, even though the peer effect itself is calculated from companymates *not* in course c.

Here we observe that all peer effects are either negative or not statistically different from zero. These falsification tests suggest that academic interaction is a key factor in generating positive peer spillovers, not merely peer group size.

5.3.3 Is There Something Special About Small Course-Company Subgroups?

To answer this question table 7 provides more summary statistics over grade observations, broken down into the three standard subgroup sizes (2-10, 11-30, 31 or more) and the two standard course types (hum/ss, math/sci). These different groups look quite similar in terms of both demographics and academic backgrounds. This is further echoed in figure 4. We can see that the distributions of ability, as proxied by SAT scores, are very similar across different peer groups.

Perhaps these course-company group sizes correlate with average grades in the course. A potential concern is that structural differences in courses produces the differences in peer effects across group sizes seen in table 4. For example, grades assigned when coursemate-companymate peer groups are only of size 2-10 may more frequently come from courses with lower enrollment. If it were the case that, only in such courses, instructors tend to assign higher grades overall—specifically when there are more high ability students present—then our positive estimates of γ_s would be biased upward by course size-related grade inflation. Due to a lack of variation within small companymate-coursemate subgroups, including a full set of company-course fixed effects (Y_{ct}) is not possible in the model shown in equation (8).

Table 8 addresses this by controlling for the average grade given in that course the previous year (a proxy for Y_{ct}).[21] Conceptually this may be an important control. Estimates show, however, that our positive peer effects for small peer group sizes hold up with the additional control. Freshman-year courses associated with smaller peer groups do not appear to distribute higher or lower grades in any systematic way. We also see a fair amount of grade persistence over time—the average grade of a course in prior years is a strong predictor of one's own grade in that course. But the addition of this control keeps our overall findings on peer effects at the company-course level intact.

5.3.4 Is There Selection Into Course-Company Subgroups?

As mentioned previously, freshman have little to no discretion over their first semester course content. There is a standard mix of classes nearly all freshman must complete. But in a college where there is such focus on standard course content for freshmen, how do we observe so many cases of small coursemate-companymate groups, and what determines their selection?

There are indeed many cases where incoming freshmen take courses for which there are only a few other freshmen in the same company who are also enrolled. There are essentially four possibilities: (1) they are deemed to have insufficient background knowledge

[21] Technically, the variable "Avg. course grade (previous year)" is the residual from a regression of the one-year lag of the course's average grade on the peer group average SAT scores of students enrolled in the course. We include this measure to capture those factors other than student ability that may influence faculty to give higher or lower course grades from the school average.

and must take remedial courses; (2) they have validated core courses and so are allowed to take advanced courses; (3) they are taking specialized courses based on background or interest (such as a "critical" language such as Chinese or Arabic); (4) they are taking a standard core course that, for some random reason, many others in their company will take in a later semester. In any of these circumstances, students still has no control over who they are in the course with (companymates or otherwise) or their instructor (see Section 3.3 for discussion of this point).

Still, one might question whether selection on common attributes might be driving positive peer effects in some way. That is, perhaps selection into small courses is based on some individual-level characteristics (unobservable to us) that also influence grades. To test this idea we re-estimate the coursemate-companymate specification, but now we only include instances of the fourth case mentioned above. That is, we entirely exclude non-core and potentially specialized courses from the analysis.

Results are presented in table 9. Despite the dramatically lower number of observations, we still observe a positive and statistically significant peer effect within small groups via high-verbal ability peers in hum/ss courses. We also see a positive coefficient for small groups with high math-ability peers for math/sci courses, although it is not statistically significant (recall that in table 4, the SAT math-related positive peer effect was smaller, and here the number of observations are almost one fourth of those used in table 4). On the other hand, all estimated coefficients for peer groups larger than ten have negative signs.

We consider an additional robustness check for potentially problematic course selection. We perform regression tests for course random assignment that are similar to those presented at the company level in table 2. That is, for each course in each academic year that contained at least 60 freshmen,[22] we regress own SAT math score on coursemate-companymates' average SAT math score, conditional on pre-USNA characteristics (and we repeat for SAT verbal scores). There are 305 such courses in our sample, so we perform a total of 610 distinct regressions, collecting the p-values of the peer effect variables' coefficients from each. Uniformly distributed p-values would provide evidence against problematic course selection. Figure 5 shows distributions of p-values from SAT math and SAT verbal regressions separately. Using the Kolmogorov–Smirnov test, we cannot reject uniformity of p-values from the SAT verbal regressions ($p = 0.382$) but we reject it for p-values from the SAT math regressions ($p = 0.045$).

Since the course-by-course regressions may indicate problematic selection on SAT math scores, we carry out one more robustness check. In table 10, we re-estimate the main coursemate-companymate peer effects model—equation (8)—now including observations only from the courses that generated p-values over 0.10 for *both* the SAT math and verbal regressions above (*i.e.,* courses for which there is no strong evidence of selection).

[22]The analysis is not sensitive to this cutoff; a cutoff of 60 provides an average of at least two freshmen from each company in each course.

While sample sizes are now lower, the previously-seen pattern of peer effects persists. There is a positive and statistically significant peer effect within small groups via high-verbal ability peers in hum/ss courses. There is also a positive coefficient for small groups with high math-ability peers for math/sci courses. All coefficient estimates for peer groups larger than ten have negative signs or are statistically significant.

These results suggest that our positive peer effects are not merely driven by selection on similar characteristics into small courses. Again, relationships based on narrowly defined tasks can generate positive peer influence. In these situations, homophily may simply not matter as much.

5.4 Discussion

The simple model in Section 2 posits a mechanism to explain the opposing findings from company and company-course levels. It must be the case that at some point the costs of mingling with the high-ability persons simply outweighs the benefits and this calculus is different depending on the peer group's scope. With respect to various explanations of peer effects, our results and our simple model evoke the idea of "invidious comparison" discussed in Hoxby and Weingarth (2005).[23] In such a scenario, the low achieving student is put off by the presence of high ability individuals, since it lowers her own status in the peer group and affects the student's self-esteem (and hence her performance). The idea of invidious comparison is captured in the model in Section 2 since the presence of high ability types affects the decision on how many peers to interact with—driven by the trade-off between group identification versus academic achievement—which ultimately affects performance.

The model in Section 2 also helps explain the results from Carrell *et al.* (2013) where the bimodal treatment group leads to a sort of self-segregation. The low-ability students cluster together while the high-ability students do the same. A post-experiment survey indicates that low-ability students are more likely to study with low-ability peer group members (more so than low ability students in the control peer group). The authors speculate that there may be a threshold, where once the fraction of low-ability members of the group reaches some point, then the endogenous sorting with like-ability individuals becomes more prevalent.

Moreover, one could imagine the idea of invidious comparison to be particularly strong for college freshmen, relative to their high school experiences. The typical high school senior who might attend USNA is a "big fish in a small pond," confident in her standing in the high school's distribution of ability. Once at college, however, that same

[23]This idea also hints at the tension between popularity and academic achievement among high school students highlighted in Fryer and Torelli (2010). In the Fryer and Torelli (2010) explanation there is a disincentive to accumulate human capital out of uncertainly of labor market outcomes, which may make more sense from the perspective of a high school student. That is probably less applicable for students already in college, whom arguably have made the decision to enroll in college out of some understanding of labor market payoffs relative to not attending college.

student is the proverbial small fish in a big pond and the tendency towards homophily over achievement becomes more natural. Now at college, the freshman, perhaps still quite immature, is overwhelmed by the caliber of her peers. Under this scenario, the penchant for homophily may dominate, as the overwhelmed student seeks refuge in students with similar ability. Yet, at the company-course level, such mechanisms are less likely.

At the company-course level, the peer group is fundamentally different, both in size and purpose. Here, the mechanisms driving the positive effects are possibly explained by the "boutique" model of peer effects (as reviewed by Hoxby and Weingarth (2005)). In this case homogeneity of ability benefits performance perhaps due to the opportunity for the teacher or school to tailor instruction to the common ability level.[24] In the USNA setting, the freshman courses include common curricula and exams, and the section sizes are small. Engaged together on common tasks in small groups, the environment for invidious comparison as inspiration for homophily is mitigated. The fact that positive effects disappear as the group size grows (and increasingly resembles the company) supports this story.

Another explanation is that at the company-course level the positive endogenous effects from interaction—studying and working together—outweigh any possible negative contextual effect. Indeed, this discussion is speculative but is supported by both our conceptual model and our reduced form results as they relate to the structural model. As noted in Hoxby and Weingarth (2005) peer effects are likely non-linear, so the typical linear-in-means estimation restricts its measurement. So our linear results may ultimately mask the true effects, as is the case with all attempts using linear-in-means models. However, the contrasting results for our different peer group levels, and the different effects found for each, provide insight into what may be occurring within the black box of peer interaction.

6 Conclusion

This paper attempts to reconcile the seemingly inconsistent findings of positive peer effect spillovers in some contexts and seemingly negative spillovers in other contexts. In short, our paper highlights the idea that context matters. In large social settings or living arrangements, more favorable average peer attributes can perversely lower individual performance as individuals increasingly group with those of like-traits. In other settings where individuals are engaged in common work on tasks, such grouping patterns can be overcome. Our results can guide policy interventions meant to harness peer effects. Cognizance about potential complications of peer group assignment is critical. Increasing peer quality can foster beneficial collaboration in common tasks in close-knit, small peer groups, but the specific

[24]Related to the boutique model is the "focus" model which takes boutique idea a bit further. In the focus model the positive effects from a relatively homogenous peer group may even spill over to the few students somewhat differentiated from the typical member of the group. And as the number of atypical students increases and the ability level becomes bimodal or multimodal the alleged benefits of the homogeneity will decrease.

design of "optimal peer groups" for positive educational outcomes is indeed complicated and remains an avenue for future research.

References

[1] Burke, Mary A., and Sass, Tim R. (2013). Classroom Peer Effects and Student Achievement. *Journal of Labor Economics*, 31, (1), 51-82.

[2] Carrell, S., Fullerton, R., and West, J. (2009). Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics,* 27(3), 439-464.

[3] Carrell, S., Sacerdote, B., and West, J. (2013). From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica,* 81(3), 855-882.

[4] Cicala, S., Fryer, R. G., and Spenkuch, J. L. (2014). A Comparative Advantage Theory of Social Interactions. Working paper.

[5] Cornelissen, T., Dustmann, C., and Schoenberg, U. (2013). Peer Effects in the Workplace. IZA Discussion paper no. 7617.

[6] Durlaf, S., and Ioannides, Y. (2010). Social Interactions. *The Annual Review of Economics,* 2, 541-578.

[7] Foster, G. (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, 1455-1475.

[8] Fryer, R. G., and Torelli, P. (2010). An Empirical Analysis of 'Acting White'. *Journal of Public Economics,* 94(5), 380-396.

[9] Glaser, D. (2009). Teenage Dropouts and Drug Use: Does the Specification of Peer Group Structure Matter? *Economics of Education Review,* 28, 497-504.

[10] Guryan, J., Kroft, K., and Notowidigdo, M. (2009). Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments. *American Economic Journal: Applied Economics,* 1(4), 34-68.

[11] Hoxby, C. M., & Weingarth, G. (2005). Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects. Working paper.

[12] Jackson, M. O. (2008). Social and economic networks. Princeton University Press.

[13] Kotlikoff, P., Rahman, A. S., and Smith, K. A. (2015). What Can a One-Year College Preparatory Program Do For College Performance? Working paper.

[14] Lavy, V., Silva, O., and Weinhardt, F. (2012) The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools. *Journal of Labor Economics,* 30(2), 4367-414.

[15] Lyle, D. (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.

[16] Lyle, D. (2009) The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point. *American Economic Journal: Applied Economics*, 1(4), 69-84.

[17] Manresa, E. (2013). Estimating the Structure of Social Interactions Using Panel Data. Working paper.

[18] Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies,* 60, 531-542.

[19] Rojstaczer, S. and Healy, C. (2010). Grading in American Colleges and Universities. Teachers College Record. March 4, 2010.

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

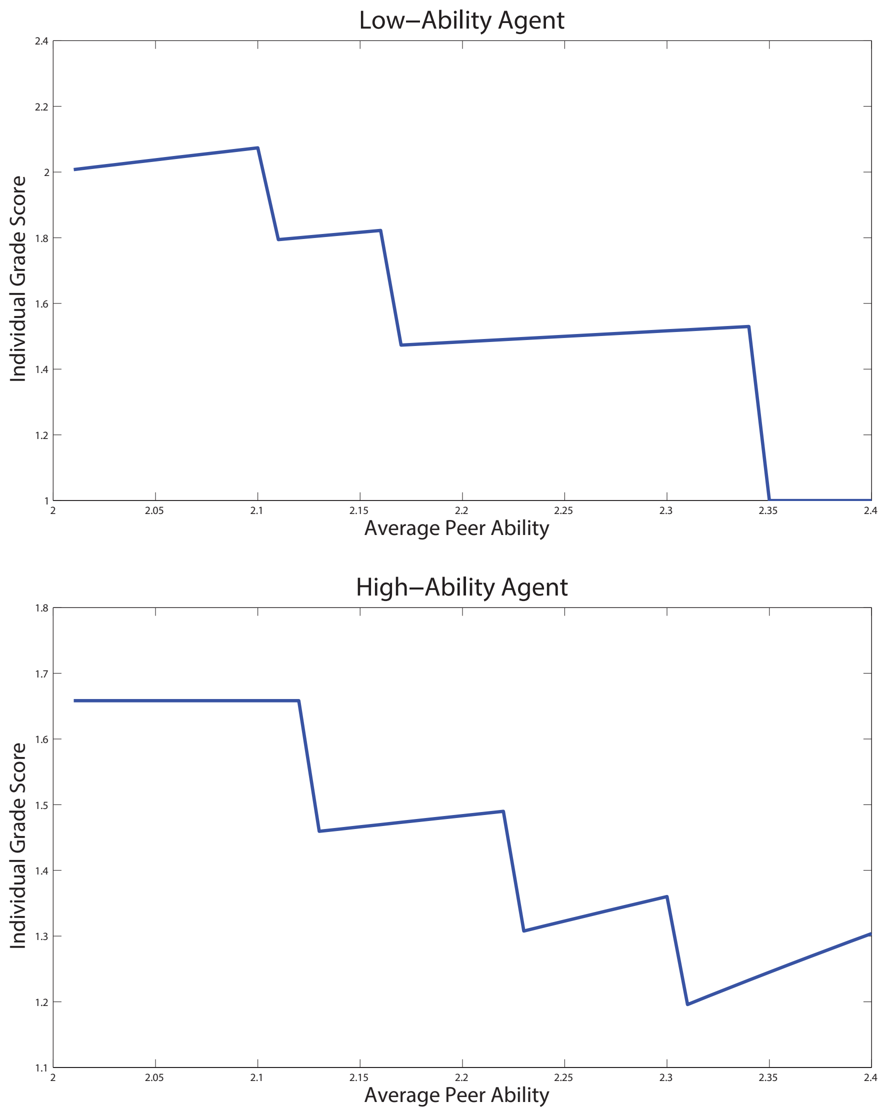
[21] Stinebrickner, R. and Stinebrickner, T. (2006). What can be Learned about Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds. *Journal of Public Economics,* 90, 1435-1454.

[22] United States Naval Academy Academic Dean Instruction 1531.56S, "Plebe Advising Handbook."

[23] United States Naval Academy Academic Dean Instruction 1531.111, "Plebe Academic Handbook."

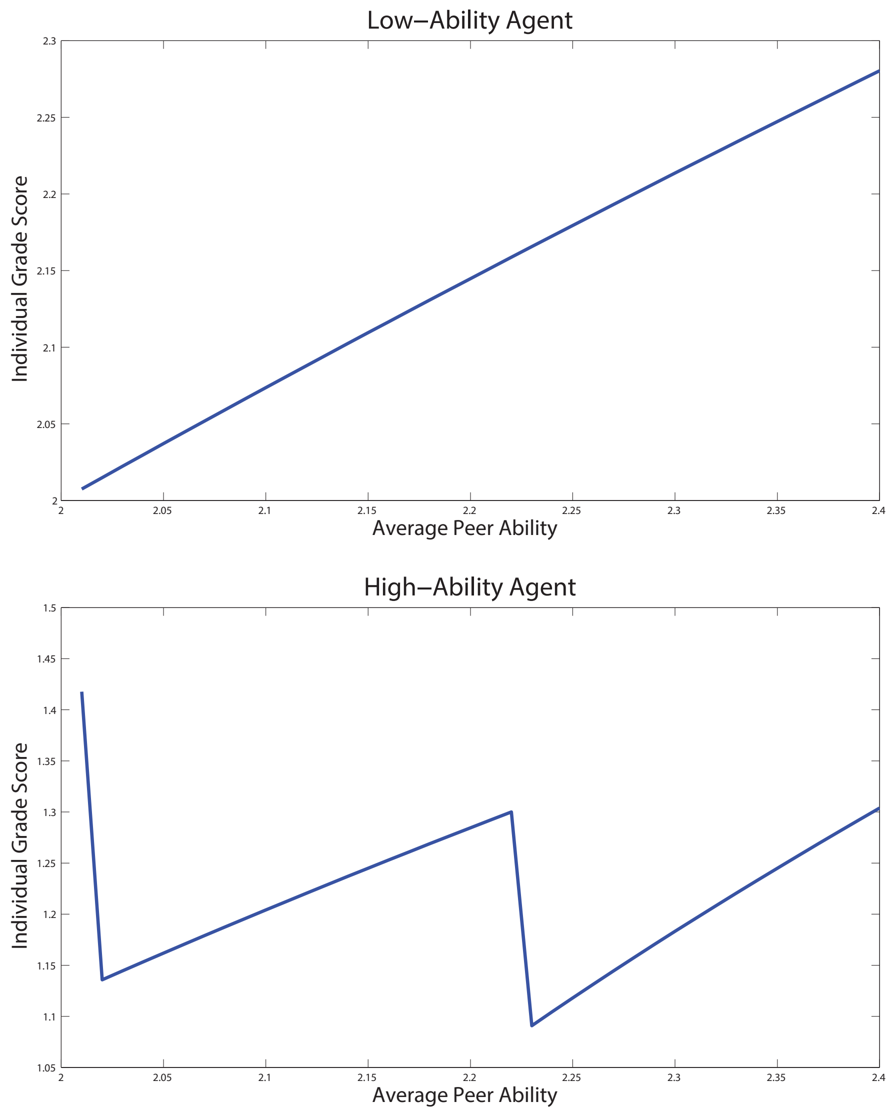
[24] Zimmerman, D. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics,* 85(1), 9-23.

Figure 1: Grade Effects from Changes in Average Peer Ability with High "Distaste" for Peer Differences



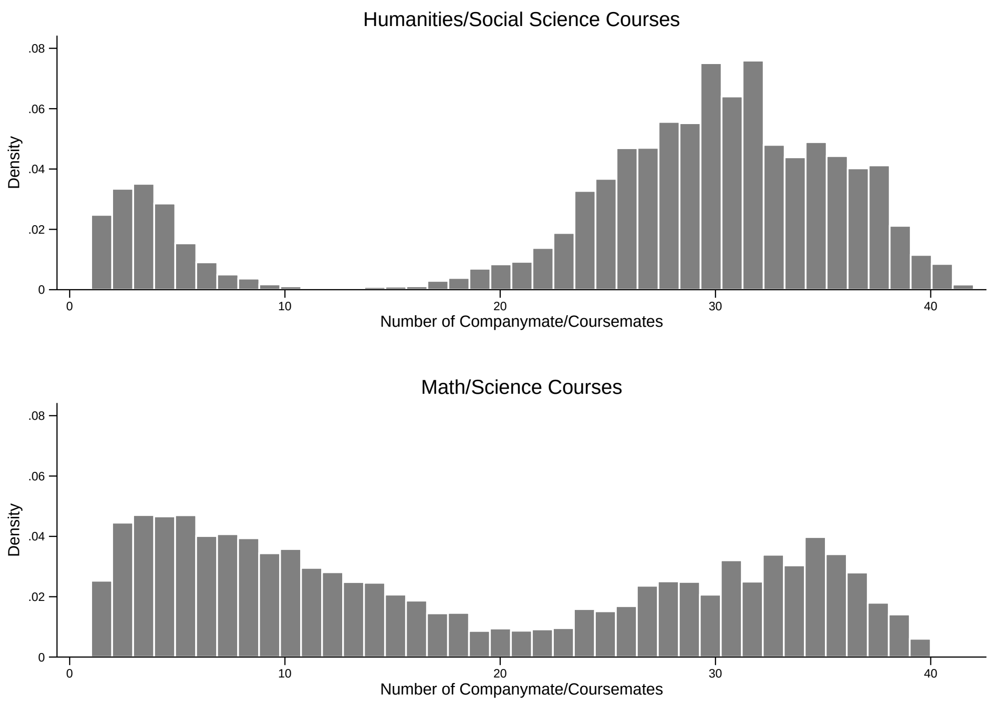
Note: Blue solid lines depict the individual's grade performance as the average ability of the potential peer group rises uniformly.

Figure 2: Grade Effects from Changes in Average Peer Ability with Low "Distaste" for Peer Differences



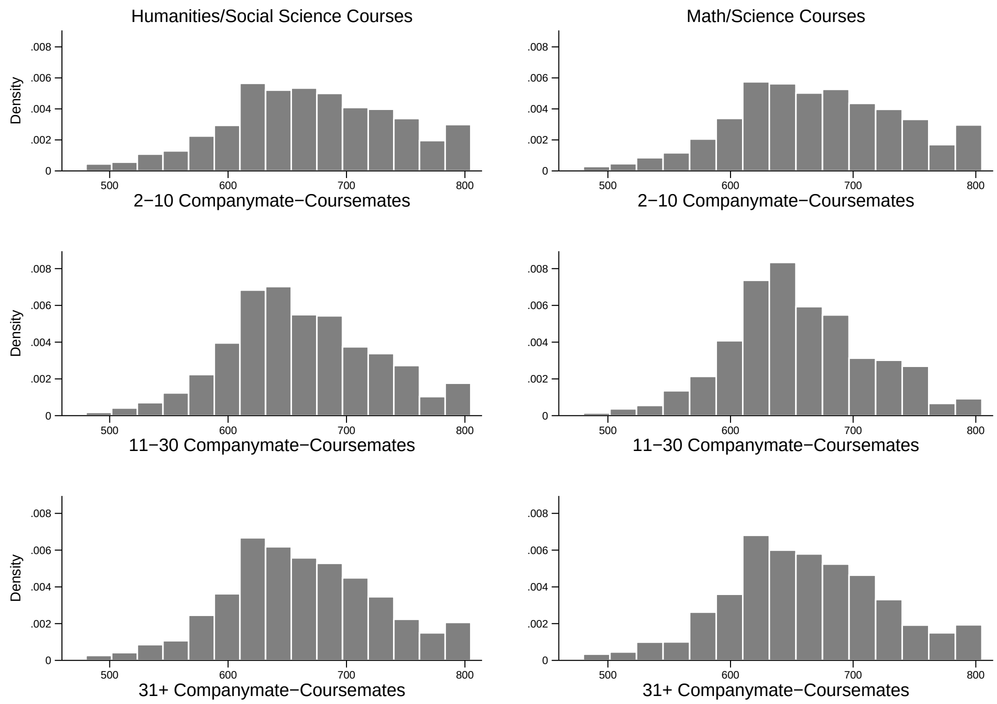
Note: Blue solid lines depict the individual's grade performance as the average ability of the potential peer group rises uniformly.

Figure 3: Distribution of Companymate-Coursemate Peer Groups



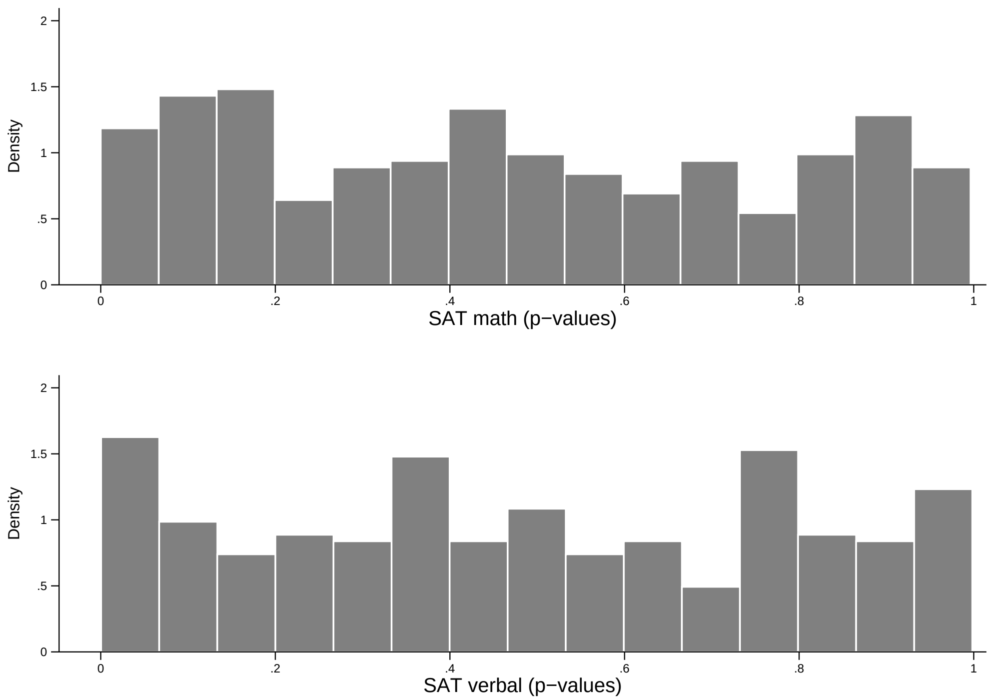
Note: Shows the distribution of companymate-coursemate peer group sizes, divided into humanities/social science courses and math/science courses. Distributions naturally yield the peer group size divisions used in the regression analysis: 2-10, 11-30, 31 or more.

Figure 4: Distribution of Math SAT Score for Companymate-Coursemate Peer Groups



Note: Shows the distribution of math SAT score for each bin of companymate-coursemate peer group size, divided into humanities/social science courses and math/science courses (we do not present similar distributions for verbal SAT score).

Figure 5: Distributions of p-values from Course-by-course Regressions (checking random assignment)



Note: For each course containing 60 or more freshman, we estimate two distinct regressions: (1) own SAT math regressed on coursemate-companymates' average SAT math; (2) own SAT verbal regressed on coursemate-companymates' average SAT verbal. All regressions also control for gender, race/ethnicity, prior enlisted status, and feeder source. Histograms show the distributions of p-values for each coefficient of the peer effect variable (*i.e.,* coursemate-companymates' average SAT), across all eligible courses.

Table 1: Summary Statistics

	Mean*	Std.Dev.
Female	0.15	0.36
Race/ethnicity:		
Black	0.06	0.24
Asian	0.04	0.20
Hispanic	0.08	0.27
White	0.79	0.41
Other	0.02	0.15
Recruited athlete	0.27	0.44
Prior enlisted	0.09	0.29
Feeder source:		
NAPS	0.17	0.38
Foundation school	0.07	0.25
None	0.75	0.43
Other	0.02	0.13
Own SAT math	662	64.3
Own SAT verbal	639	68.8
Companymates' SAT math	662	9.9
Companymates' SAT verbal	638	12.0
Course-companymates' SAT math	661	31.3
Course-companymates' SAT verbal	638	32.0
Observations	100,593	

Note: *Column shows sample means for SAT scores and sample proportions for all other variables.

Table 2: Regression Checks for Company Random Assignment

Dep. Var.:	Own SAT Math		Own SAT Verbal		
Indep. Var.:	Peers' SAT Math (average)		Peers' SAT Verbal (average)		
Ac. Year:	Coefficient	Standard Error	Coefficient	Standard Error	Obs.
1988	0.0546	(0.186)	0.144	(0.164)	1176
1989	-0.306	(0.329)	-0.140	(0.232)	1198
1990	-0.369	(0.246)	-0.215	(0.297)	1234
1991	-0.820**	(0.280)	0.0597	(0.185)	1101
1992	-0.667	(0.404)	-0.232	(0.260)	1030
1993	-0.114	(0.233)	-0.270	(0.246)	1121
1994	-0.614	(0.345)	-0.215	(0.282)	1064
1995	-0.275	(0.253)	0.0524	(0.211)	989
1996	-0.300	(0.417)	-0.328	(0.225)	1015
1997	-0.744	(0.455)	-0.765	(0.395)	1031
1998	-0.233	(0.254)	-0.148	(0.241)	1040
1999	-0.830*	(0.381)	-0.489	(0.602)	1113
2000	-0.0493	(0.294)	-0.548	(0.372)	1105
2001	-0.0471	(0.292)	-1.169	(0.615)	1117
2002	-0.222	(0.293)	-0.399	(0.392)	1146
2003	-0.259	(0.257)	-0.0614	(0.242)	1127
2004	-0.770	(0.453)	-0.113	(0.312)	1155
2005	-0.874	(0.453)	0.251	(0.187)	1165
2006	-0.569	(0.395)	-0.329	(0.332)	1152
2007	-0.0159	(0.179)	0.0197	(0.242)	1167
2008	-0.240	(0.277)	0.102	(0.179)	1128
2009	-0.248	(0.209)	-0.00832	(0.240)	1188

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. This table presents point estimates and standard errors of one particular coefficient in 44 distinct regressions. OLS estimations are carried out separately for subsamples of freshmen from each graduating class. The dependent variable is either own SAT math or own SAT verbal and the independent variable of interest is either companymates' average SAT math or companymates' average SAT verbal. Each specification also controls for gender, race/ethnicity, prior enlisted status, and feeder source. Standard errors are clustered by company.

Table 3: Regressions - All Freshman Companymates as Peers

Dependent Variable: Grade	Humanities and Social Science Courses			Math and Science Courses		
Female	0.0430***	0.0430***	0.0433***	-0.0896***	-0.0897***	-0.0893***
	(0.0113)	(0.0113)	(0.0113)	(0.0144)	(0.0144)	(0.0144)
Race/ethnicity (ref.: white and other)						
Black	-0.277***	-0.277***	-0.278***	-0.267***	-0.267***	-0.267***
	(0.0176)	(0.0176)	(0.0176)	(0.0213)	(0.0213)	(0.0213)
Asian	-0.0896***	-0.0896***	-0.0895***	-0.0802***	-0.0801***	-0.0792***
	(0.0196)	(0.0196)	(0.0196)	(0.0237)	(0.0237)	(0.0236)
Hispanic	-0.142***	-0.142***	-0.142***	-0.200***	-0.200***	-0.200***
	(0.0155)	(0.0155)	(0.0155)	(0.0191)	(0.0191)	(0.0191)
Recruited athlete	-0.0822***	-0.0821***	-0.0822***	-0.132***	-0.132***	-0.132***
	(0.00932)	(0.00933)	(0.00933)	(0.0118)	(0.0118)	(0.0118)
Prior enlisted	0.00877	0.00908	0.00887	0.158***	0.159***	0.157***
	(0.0172)	(0.0172)	(0.0172)	(0.0209)	(0.0209)	(0.0209)
Feeder source (ref.: none)						
NAPS	-0.161***	-0.162***	-0.162***	0.141***	0.141***	0.141***
	(0.0128)	(0.0129)	(0.0128)	(0.0159)	(0.0159)	(0.0159)
Foundation school	-0.0594***	-0.0590***	-0.0590***	-0.0821***	-0.0814***	-0.0814***
	(0.0157)	(0.0157)	(0.0157)	(0.0208)	(0.0209)	(0.0208)
Other	-0.0271	-0.0273	-0.0269	0.248***	0.247***	0.248***
	(0.0393)	(0.0393)	(0.0394)	(0.0416)	(0.0416)	(0.0416)
Own SAT math	0.000550***	0.000547***	0.000549***	0.00513***	0.00513***	0.00513***
	(0.0000770)	(0.0000771)	(0.0000770)	(0.000101)	(0.000101)	(0.000101)
Own SAT verbal	0.00297***	0.00297***	0.00297***	0.000307***	0.000308***	0.000307***
	(0.0000726)	(0.0000726)	(0.0000727)	(0.0000834)	(0.0000835)	(0.0000836)
Peers' SAT math (avg.)	-0.000945	-0.000978		-0.00198**	-0.00199**	
	(0.000547)	(0.000547)		(0.000710)	(0.000706)	
Peers' SAT math (std. dev.)		-0.000709			-0.00141	
		(0.000590)			(0.000870)	
Fraction of peers in top SAT math quartile			-0.143*			-0.228**
			(0.0676)			(0.0879)
Fraction of peers in bottom SAT math quartile			0.0882			0.152
			(0.0612)			(0.0831)
Peers' SAT verbal (avg.)	-0.000785	-0.000823*		-0.00161**	-0.00165**	
	(0.000408)	(0.000406)		(0.000562)	(0.000562)	
Peers' SAT verbal (std. dev.)		-0.000290			0.000320	
		(0.000498)			(0.000656)	
Fraction of peers in top SAT verbal quartile			-0.0729			-0.0667
			(0.0561)			(0.0778)
Fraction of peers in bottom SAT verbal quartile			0.0147			0.220**
			(0.0592)			(0.0743)
Constant	1.619***	1.726***	0.514***	1.150*	1.245*	-1.235***
	(0.405)	(0.406)	(0.0662)	(0.500)	(0.503)	(0.0827)
Observations	47748	47748	47748	52845	52845	52845
R^2	0.169	0.169	0.169	0.146	0.146	0.146

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. OLS estimations are carried out separately for humanities and social science course grades (columns 1-3) and math and science course grades (column 4-6). Models differ in the specifications of peer effect variables. Coefficients for academic year dummy variables are included in the estimations but are not shown. Standard errors are clustered by company-academic year groups.

Table 4: Regressions - Freshman Coursemate-Companymates as Peers

Dependent Variable: Grade	Humanities and Social Science Courses			Math and Science Courses		
Peer group size:	2-10	11-30	31 or more	2-10	11-30	31 or more
Female	0.139***	0.106***	-0.0512**	-0.0238	-0.144***	-0.0993***
	(0.0238)	(0.0164)	(0.0175)	(0.0192)	(0.0230)	(0.0205)
Race/ethnicity (ref.: white and other)						
Black	-0.225***	-0.253***	-0.326***	-0.217***	-0.334***	-0.268***
	(0.0367)	(0.0261)	(0.0252)	(0.0313)	(0.0292)	(0.0329)
Asian	-0.0483	-0.0848**	-0.118***	-0.0185	-0.128***	-0.102**
	(0.0462)	(0.0269)	(0.0293)	(0.0340)	(0.0346)	(0.0383)
Hispanic	-0.0728*	-0.120***	-0.182***	-0.173***	-0.214***	-0.211***
	(0.0352)	(0.0214)	(0.0223)	(0.0285)	(0.0277)	(0.0283)
Recruited athlete	-0.108***	-0.0655***	-0.0920***	-0.0760***	-0.154***	-0.185***
	(0.0204)	(0.0125)	(0.0135)	(0.0180)	(0.0169)	(0.0183)
Prior enlisted	0.0446	0.000958	0.00419	0.143***	0.180***	0.120**
	(0.0379)	(0.0220)	(0.0276)	(0.0313)	(0.0274)	(0.0377)
Feeder source (ref.: none)						
NAPS	-0.119***	-0.158***	-0.191***	0.0404	0.135***	0.271***
	(0.0302)	(0.0174)	(0.0194)	(0.0256)	(0.0227)	(0.0257)
Foundation school	-0.0785*	-0.0546*	-0.0509*	-0.126***	-0.101***	0.0246
	(0.0392)	(0.0211)	(0.0222)	(0.0298)	(0.0296)	(0.0314)
Other	-0.0352	-0.0148	-0.0221	0.272***	0.127*	0.395***
	(0.0899)	(0.0565)	(0.0520)	(0.0646)	(0.0638)	(0.0601)
Own SAT math	0.000699***	0.000387***	0.000599***	0.00399***	0.00508***	0.00616***
	(0.000181)	(0.000108)	(0.000110)	(0.000149)	(0.000169)	(0.000146)
Own SAT verbal	0.00252***	0.00286***	0.00285***	-0.000325**	0.000389**	0.000952***
	(0.000166)	(0.000108)	(0.000103)	(0.000118)	(0.000129)	(0.000126)
Coursemate-peers' SAT math (avg.)	0.0000572	-0.00118	-0.000179	0.000738***	-0.00246***	-0.000542
	(0.000213)	(0.000624)	(0.000768)	(0.000183)	(0.000563)	(0.000915)
Coursemate-peers' SAT verbal (avg.)	0.00109***	-0.000646	-0.000795	0.000147	-0.000894	-0.00179*
	(0.000180)	(0.000503)	(0.000629)	(0.000194)	(0.000612)	(0.000797)
Constant	-0.0483	1.821***	1.197*	-0.665***	1.017**	-0.922
	(0.153)	(0.430)	(0.560)	(0.131)	(0.375)	(0.692)
Observations	6342	19896	21510	19842	19197	13806
R^2	0.292	0.138	0.166	0.148	0.144	0.215

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. OLS estimations are carried out separately for humanities and social science course grades (columns 1-3) and math and science course grades (columns 4-6). Estimations are further stratified by the size of the coursemate-peer group associated with each grade observation (sizes 2-10, 11-30, and over 30). Coefficients for academic year dummy variables are included in the estimations but are not shown. Standard errors are clustered by company-academic year groups.

Table 5: Robustness - Freshman Coursemate-Companymates as Peers (Alternate Peer Effect Specification)

Dependent Variable: Grade	Humanities and Social Science Courses			Math and Science Courses		
Peer group size:	2-10	11-30	31 or more	2-10	11-30	31 or more
Female	0.138***	0.107***	-0.0509**	-0.0241	-0.143***	-0.0993***
	(0.0237)	(0.0163)	(0.0175)	(0.0191)	(0.0230)	(0.0205)
Race/ethnicity (ref.: white and other)						
Black	-0.222***	-0.253***	-0.326***	-0.213***	-0.335***	-0.268***
	(0.0367)	(0.0260)	(0.0252)	(0.0314)	(0.0292)	(0.0329)
Asian	-0.0483	-0.0855**	-0.118***	-0.0179	-0.128***	-0.102**
	(0.0461)	(0.0269)	(0.0293)	(0.0341)	(0.0347)	(0.0383)
Hispanic	-0.0708*	-0.120***	-0.181***	-0.173***	-0.214***	-0.212***
	(0.0352)	(0.0214)	(0.0222)	(0.0285)	(0.0276)	(0.0283)
Recruited athlete	-0.108***	-0.0653***	-0.0921***	-0.0748***	-0.155***	-0.185***
	(0.0204)	(0.0125)	(0.0135)	(0.0180)	(0.0169)	(0.0183)
Prior enlisted	0.0508	0.000858	0.00421	0.145***	0.178***	0.119**
	(0.0377)	(0.0219)	(0.0276)	(0.0314)	(0.0274)	(0.0377)
Feeder source (ref.: none)						
NAPS	-0.117***	-0.159***	-0.191***	0.0409	0.135***	0.270***
	(0.0301)	(0.0174)	(0.0195)	(0.0256)	(0.0227)	(0.0257)
Foundation school	-0.0773*	-0.0543*	-0.0502*	-0.126***	-0.0998***	0.0239
	(0.0390)	(0.0211)	(0.0223)	(0.0298)	(0.0297)	(0.0315)
Other	-0.0323	-0.0144	-0.0225	0.268***	0.127*	0.396***
	(0.0906)	(0.0564)	(0.0520)	(0.0646)	(0.0635)	(0.0600)
Own SAT math	0.000717***	0.000385***	0.000596***	0.00399***	0.00508***	0.00616***
	(0.000180)	(0.000108)	(0.000110)	(0.000150)	(0.000169)	(0.000145)
Own SAT verbal	0.00250***	0.00286***	0.00285***	-0.000320**	0.000389**	0.000950***
	(0.000169)	(0.000108)	(0.000103)	(0.000118)	(0.000129)	(0.000126)
Fraction of peers in top SAT math quartile	-0.0667*	-0.169*	-0.117	0.00361	-0.316***	0.0855
	(0.0294)	(0.0851)	(0.0932)	(0.0279)	(0.0912)	(0.154)
Fraction of peers in bottom SAT math quartile	-0.111***	0.0908	0.0194	-0.119***	0.172*	0.0994
	(0.0316)	(0.0792)	(0.0901)	(0.0335)	(0.0776)	(0.131)
Fraction of peers in top SAT verbal quartile	0.120***	-0.196**	-0.0341	0.0102	-0.0455	-0.296*
	(0.0298)	(0.0693)	(0.0923)	(0.0289)	(0.0815)	(0.134)
Fraction of peers in bottom SAT verbal quartile	-0.102**	-0.0738	0.0255	-0.0461	0.0833	0.0410
	(0.0360)	(0.0675)	(0.0858)	(0.0326)	(0.0863)	(0.124)
Constant	0.732***	0.719***	0.598***	-0.0412	-1.165***	-2.411***
	(0.142)	(0.0979)	(0.102)	(0.115)	(0.137)	(0.166)
Observations	6342	19896	21510	19842	19197	13806
R^2	0.294	0.139	0.166	0.148	0.145	0.215

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. OLS estimations are carried out separately for humanities and social science course grades (columns 1-3) and math and science course grades (columns 4-6). Estimations are further stratified by the size of the coursemate-peer group associated with each grade observation (sizes 2-10, 11-30, and over 30). Coefficients for academic year dummy variables are included in the estimations but are not shown. Standard errors are clustered by company-academic year groups.

Table 6: Robustness - Freshman Companymates Not in Same Course as Peers

Dependent Variable: Grade	Humanities and Social Science Courses				Math and Science Courses			
Peer group size:	All	2-10	11-30	31 or more	All	2-10	11-30	31 or more
Female	0.0601***	-0.0866***	0.153***	0.124***	-0.0383*	-0.0234	-0.144***	-0.0908***
	(0.0116)	(0.0144)	(0.0239)	(0.0168)	(0.0185)	(0.0192)	(0.0232)	(0.0207)
Race/ethnicity (ref.: white and other)								
Black	-0.271***	-0.270***	-0.223***	-0.253***	-0.316***	-0.214***	-0.341***	-0.271***
	(0.0185)	(0.0215)	(0.0369)	(0.0272)	(0.0274)	(0.0313)	(0.0294)	(0.0336)
Asian	-0.0811***	-0.0765**	-0.0448	-0.0903***	-0.0971**	-0.0174	-0.119***	-0.102**
	(0.0202)	(0.0234)	(0.0462)	(0.0270)	(0.0319)	(0.0341)	(0.0337)	(0.0393)
Hispanic	-0.139***	-0.202***	-0.0761*	-0.125***	-0.171***	-0.170***	-0.212***	-0.222***
	(0.0158)	(0.0192)	(0.0356)	(0.0218)	(0.0233)	(0.0286)	(0.0278)	(0.0290)
Recruited athlete	-0.0810***	-0.130***	-0.113***	-0.0698***	-0.0856***	-0.0779***	-0.152***	-0.184***
	(0.00956)	(0.0119)	(0.0204)	(0.0127)	(0.0144)	(0.0181)	(0.0168)	(0.0188)
Prior enlisted	0.0108	0.160***	0.0517	0.00310	0.00692	0.145***	0.180***	0.120**
	(0.0176)	(0.0209)	(0.0377)	(0.0224)	(0.0297)	(0.0314)	(0.0281)	(0.0374)
Feeder source (ref.: none)								
NAPS	-0.165***	0.133***	-0.120***	-0.160***	-0.201***	0.0500	0.129***	0.267***
	(0.0131)	(0.0161)	(0.0303)	(0.0177)	(0.0202)	(0.0256)	(0.0231)	(0.0264)
Foundation school	-0.0538**	-0.0889***	-0.0815*	-0.0609**	-0.0293	-0.124***	-0.109***	0.0193
	(0.0163)	(0.0209)	(0.0397)	(0.0218)	(0.0240)	(0.0298)	(0.0298)	(0.0319)
Other	-0.0196	0.251***	-0.0606	0.00685	-0.0258	0.276***	0.124	0.411***
	(0.0412)	(0.0420)	(0.0901)	(0.0552)	(0.0584)	(0.0646)	(0.0650)	(0.0617)
Own SAT math	0.000553***	0.00498***	0.000707***	0.000395***	0.000586***	0.00417***	0.00487***	0.00613***
	(0.0000793)	(0.000102)	(0.000182)	(0.000108)	(0.000118)	(0.000141)	(0.000170)	(0.000147)
Own SAT verbal	0.00290***	0.000255**	0.00299***	0.00275***	0.00276***	-0.000305*	0.000350**	0.000908***
	(0.0000739)	(0.0000843)	(0.000158)	(0.000109)	(0.000109)	(0.000119)	(0.000130)	(0.000128)
Non-coursemate-peers' SAT math (avg.)	0.0000947	-0.00183***	-0.00198*	0.000320	0.0000107	-0.00246**	-0.000256	0.0000367
	(0.000128)	(0.000258)	(0.000942)	(0.000223)	(0.000132)	(0.000756)	(0.000407)	(0.000213)
Non-coursemate-peers' SAT verbal (avg.)	-0.000328***	-0.000528*	-0.00205*	-0.000526***	-0.000169	-0.00111	-0.000258	-0.0000286
	(0.0000889)	(0.000248)	(0.000888)	(0.000146)	(0.0000982)	(0.000689)	(0.000313)	(0.000214)
Constant	0.692***	0.473**	3.043***	0.842***	0.750***	2.136***	-0.688*	-2.411***
	(0.0969)	(0.177)	(0.739)	(0.162)	(0.125)	(0.554)	(0.285)	(0.203)
Observations	43338	51932	6342	18524	18472	19842	18911	13179
R^2	0.167	0.149	0.286	0.135	0.162	0.148	0.140	0.213

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. OLS estimations are carried out separately for all humanities and social science course grades and all math and science course grades, as well as across coursemate-companymate peer subgroups. Coefficients for academic year dummy variables are included in the estimations but are not shown. Standard errors are clustered by company-academic year groups.

Table 7: Summary Statistics - By Coursemate-Companymate Subgroup Sizes

	Humanities and Social Science Courses					
Peer group size:	2-10		11-30		31 or more	
	Mean*	Std. Dev.	Mean*	Std. Dev.	Mean*	Std. Dev.
Female	0.17	0.38	0.14	0.35	0.16	0.37
Race/ethnicity:						
Black	0.07	0.26	0.06	0.24	0.06	0.24
Asian	0.04	0.20	0.04	0.19	0.04	0.19
Hispanic	0.08	0.27	0.08	0.27	0.08	0.28
White	0.76	0.43	0.80	0.40	0.79	0.41
Other	0.03	0.17	0.02	0.14	0.03	0.16
Recruited athlete	0.26	0.44	0.27	0.44	0.28	0.45
Prior enlisted	0.09	0.28	0.10	0.30	0.09	0.28
Feeder source:						
NAPS	0.20	0.40	0.16	0.37	0.17	0.38
Foundation school	0.04	0.21	0.07	0.26	0.07	0.25
None	0.74	0.44	0.75	0.43	0.74	0.44
Other	0.01	0.11	0.02	0.12	0.02	0.13
Own SAT math	667	72.1	660	61.9	662	64.9
Own SAT verbal	650	89.5	637	61.6	637	66.6
Observations	6,342		19,896		21,510	

	Math and Science Courses					
Peer group size:	2-10		11-30		31 or more	
	Mean*	Std. Dev.	Mean*	Std. Dev.	Mean*	Std. Dev.
Female	0.14	0.35	0.14	0.34	0.17	0.38
Race/ethnicity:						
Black	0.06	0.24	0.07	0.25	0.06	0.24
Asian	0.04	0.20	0.04	0.20	0.04	0.19
Hispanic	0.07	0.26	0.08	0.27	0.09	0.29
White	0.80	0.40	0.79	0.41	0.77	0.42
Other	0.02	0.14	0.02	0.13	0.03	0.18
Recruited athlete	0.26	0.44	0.27	0.44	0.28	0.45
Prior enlisted	0.09	0.29	0.11	0.31	0.08	0.27
Feeder source:						
NAPS	0.16	0.37	0.17	0.38	0.17	0.38
Foundation school	0.06	0.25	0.07	0.26	0.06	0.24
None	0.76	0.43	0.74	0.44	0.75	0.44
Other	0.01	0.12	0.02	0.13	0.02	0.14
Own SAT math	669	68.7	655	56.6	659	65.2
Own SAT verbal	643	70.8	636	66.5	636	70.4
Observations	19,842		19,197		13,806	

Note: *Column shows sample means for SAT scores and sample proportions for all other variables.

Table 8: Robustness - Freshman Coursemate-Companymates as Peers (with Average Course Grade Control)

Dependent Variable: Grade	Humanities and Social Science Courses			Math and Science Courses		
Peer group size:	2-10	11-30	31 or more	2-10	11-30	31 or more
Female	0.108***	0.115***	-0.0480**	-0.0202	-0.147***	-0.100***
	(0.0240)	(0.0163)	(0.0177)	(0.0194)	(0.0239)	(0.0205)
Race/ethnicity (ref.: white and other)						
Black	-0.211***	-0.255***	-0.331***	-0.207***	-0.334***	-0.269***
	(0.0375)	(0.0265)	(0.0259)	(0.0322)	(0.0301)	(0.0330)
Asian	-0.0548	-0.0837**	-0.106***	-0.0209	-0.136***	-0.100**
	(0.0465)	(0.0268)	(0.0298)	(0.0339)	(0.0339)	(0.0383)
Hispanic	-0.0574	-0.123***	-0.176***	-0.168***	-0.226***	-0.212***
	(0.0341)	(0.0218)	(0.0222)	(0.0293)	(0.0287)	(0.0283)
Recruited athlete	-0.0881***	-0.0679***	-0.0927***	-0.0569**	-0.153***	-0.187***
	(0.0203)	(0.0125)	(0.0138)	(0.0181)	(0.0171)	(0.0183)
Prior enlisted	0.0441	-0.00321	-0.00167	0.137***	0.176***	0.116**
	(0.0376)	(0.0221)	(0.0285)	(0.0327)	(0.0281)	(0.0376)
Feeder source (ref.: none)						
NAPS	-0.102***	-0.156***	-0.199***	0.0482	0.120***	0.267***
	(0.0302)	(0.0177)	(0.0197)	(0.0265)	(0.0231)	(0.0257)
Foundation school	-0.0814*	-0.0573**	-0.0517*	-0.120***	-0.0930**	0.0224
	(0.0386)	(0.0214)	(0.0228)	(0.0305)	(0.0302)	(0.0314)
Other	0.00276	0.0115	-0.00861	0.233***	0.134*	0.401***
	(0.0932)	(0.0548)	(0.0542)	(0.0698)	(0.0667)	(0.0600)
Own SAT math	0.000694***	0.000394***	0.000602***	0.00339***	0.00486***	0.00609***
	(0.000179)	(0.000109)	(0.000113)	(0.000160)	(0.000169)	(0.000145)
Own SAT verbal	0.00203***	0.00284***	0.00280***	-0.000322**	0.000389**	0.000937***
	(0.000174)	(0.000107)	(0.000104)	(0.000122)	(0.000134)	(0.000125)
Coursemate-peers' SAT math (avg.)	0.000137	-0.00132*	-0.0000744	0.00128***	-0.00222***	-0.000569
	(0.000208)	(0.000598)	(0.000749)	(0.000189)	(0.000553)	(0.000912)
Coursemate-peers' SAT verbal (avg.)	0.00150***	-0.000697	-0.000910	0.000116	-0.00101	-0.00181*
	(0.000178)	(0.000497)	(0.000610)	(0.000199)	(0.000593)	(0.000796)
Avg. course grade (previous year)	0.390***	0.563***	0.479***	0.533***	0.767***	1.686***
	(0.0407)	(0.0796)	(0.0785)	(0.0273)	(0.0364)	(0.170)
Constant	-0.0426	1.958***	1.232*	-0.612***	1.075**	-0.840
	(0.152)	(0.417)	(0.548)	(0.133)	(0.378)	(0.689)
Observations	6032	19434	20863	18389	17857	13775
R^2	0.312	0.139	0.166	0.169	0.160	0.218

Note: $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$. OLS estimations are carried out separately for humanities and social science course grades (columns 1-3) and math and science course grades (columns 4-6). Estimations are further stratified by the size of the coursemate-peer group associated with each grade observation (sizes 2-10, 11-30, and over 30). Coefficients for academic year dummy variables are included in the estimations but are not shown. Standard errors are clustered by company-academic year groups. Average course grade from the previous year was dropped in the far-right column due to high collinearity with academic year dummies in that subsample.

Table 9: Robustness - Freshman Coursemate-Companymates as Peers (only Core Course Grades)

Dependent Variable: Grade	Humanities and Social Science Core Courses			Math and Science Core Courses		
Peer group size:	2-10	11-30	31 or more	2-10	11-30	31 or more
Female	0.153***	0.111***	-0.0528**	-0.0433	-0.140***	-0.0992***
	(0.0358)	(0.0163)	(0.0181)	(0.0340)	(0.0240)	(0.0209)
Race/ethnicity (ref.: white and other)						
Black	-0.181**	-0.252***	-0.333***	-0.144**	-0.314***	-0.259***
	(0.0579)	(0.0265)	(0.0265)	(0.0468)	(0.0318)	(0.0342)
Asian	-0.0555	-0.0886**	-0.115***	-0.0881	-0.130***	-0.0959*
	(0.0665)	(0.0273)	(0.0305)	(0.0700)	(0.0382)	(0.0403)
Hispanic	-0.110*	-0.117***	-0.195***	-0.141**	-0.214***	-0.217***
	(0.0548)	(0.0215)	(0.0230)	(0.0456)	(0.0301)	(0.0288)
Recruited athlete	-0.126***	-0.0639***	-0.0914***	-0.101**	-0.177***	-0.184***
	(0.0326)	(0.0126)	(0.0141)	(0.0307)	(0.0187)	(0.0187)
Prior enlisted	0.0687	0.00460	0.00441	0.263***	0.198***	0.138***
	(0.0603)	(0.0225)	(0.0291)	(0.0513)	(0.0308)	(0.0396)
Feeder source (ref.: none)						
NAPS	-0.128**	-0.156***	-0.187***	0.112**	0.133***	0.278***
	(0.0454)	(0.0177)	(0.0203)	(0.0387)	(0.0241)	(0.0264)
Foundation school	-0.0588	-0.0457*	-0.0489*	-0.0108	-0.120***	0.0336
	(0.0505)	(0.0213)	(0.0230)	(0.0475)	(0.0315)	(0.0329)
Other	0.134	-0.0140	-0.0142	0.380***	0.196**	0.375***
	(0.124)	(0.0562)	(0.0537)	(0.0944)	(0.0627)	(0.0630)
Own SAT math	0.000752**	0.000420***	0.000616***	0.00491***	0.00500***	0.00614***
	(0.000253)	(0.000107)	(0.000115)	(0.000307)	(0.000175)	(0.000150)
Own SAT verbal	0.00271***	0.00282***	0.00283***	-0.000665**	0.000461**	0.000926***
	(0.000244)	(0.000109)	(0.000107)	(0.000240)	(0.000142)	(0.000127)
Coursemate-peers' SAT math (avg.)	0.0000340	-0.00120	-0.000350	0.000667	-0.00217***	-0.000404
	(0.000310)	(0.000632)	(0.000808)	(0.000567)	(0.000637)	(0.000928)
Coursemate-peers' SAT verbal (avg.)	0.000843**	-0.000534	-0.000592	0.000315	-0.000370	-0.00187*
	(0.000273)	(0.000518)	(0.000648)	(0.000482)	(0.000600)	(0.000817)
Constant	0.166	1.766***	1.190*	-1.384***	0.603	-0.929
	(0.243)	(0.437)	(0.585)	(0.389)	(0.413)	(0.681)
Observations	2864	19130	19933	5502	14357	12847
R^2	0.226	0.137	0.166	0.114	0.172	0.214

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. OLS estimations are carried out separately for humanities and social science course grades (columns 1-3) and math and science course grades (columns 4-6). Estimations are further stratified by the size of the coursemate-peer group associated with each grade observation (sizes 2-10, 11-30, and over 30). Coefficients for academic year dummy variables are included in the estimations but are not shown. Standard errors are clustered by company-academic year groups.

Table 10: Robustness - Freshman Coursemate-Companymates as Peers (only "verified-random" courses)

Dependent Variable: Grade	Humanities and Social Science "Verified" Courses			Math and Science "Verified" Courses		
Peer group size:	2-10	11-30	31 or more	2-10	11-30	31 or more
Female	0.166***	0.110***	-0.0465*	-0.0296	-0.141***	-0.104***
	(0.0289)	(0.0195)	(0.0211)	(0.0237)	(0.0266)	(0.0273)
Race/ethnicity (ref.: white and other)						
Black	-0.185***	-0.267***	-0.321***	-0.237***	-0.334***	-0.316***
	(0.0468)	(0.0301)	(0.0304)	(0.0368)	(0.0341)	(0.0417)
Asian	-0.0665	-0.0649*	-0.104**	-0.0291	-0.139***	-0.0892
	(0.0618)	(0.0308)	(0.0349)	(0.0391)	(0.0372)	(0.0474)
Hispanic	-0.0884*	-0.123***	-0.176***	-0.185***	-0.214***	-0.253***
	(0.0415)	(0.0262)	(0.0270)	(0.0331)	(0.0313)	(0.0392)
Recruited athlete	-0.0809**	-0.0803***	-0.0863***	-0.0737***	-0.150***	-0.210***
	(0.0262)	(0.0146)	(0.0161)	(0.0201)	(0.0188)	(0.0244)
Prior enlisted	-0.000576	0.00268	-0.0202	0.134***	0.158***	0.0839
	(0.0491)	(0.0249)	(0.0324)	(0.0360)	(0.0301)	(0.0479)
Feeder source (ref.: none)						
NAPS	-0.0917*	-0.157***	-0.183***	0.0650*	0.124***	0.266***
	(0.0383)	(0.0203)	(0.0237)	(0.0288)	(0.0256)	(0.0347)
Foundation school	-0.0387	-0.0516*	-0.0564*	-0.118***	-0.107**	0.00957
	(0.0465)	(0.0251)	(0.0256)	(0.0355)	(0.0334)	(0.0409)
Other	0.0801	-0.0310	-0.0258	0.310***	0.108	0.354***
	(0.100)	(0.0650)	(0.0605)	(0.0719)	(0.0668)	(0.0782)
Own SAT math	0.000789***	0.000382**	0.000664***	0.00393***	0.00492***	0.00630***
	(0.000207)	(0.000131)	(0.000132)	(0.000177)	(0.000182)	(0.000200)
Own SAT verbal	0.00247***	0.00294***	0.00293***	-0.000360**	0.000337*	0.000821***
	(0.000194)	(0.000123)	(0.000118)	(0.000138)	(0.000149)	(0.000170)
Coursemate-peers' SAT math (avg.)	-0.000344	-0.00157*	0.000416	0.000774**	-0.00121*	0.000397
	(0.000271)	(0.000736)	(0.000887)	(0.000246)	(0.000577)	(0.00129)
Coursemate-peers' SAT verbal (avg.)	0.00163***	-0.000467	-0.000540	0.000349	-0.00113	-0.000720
	(0.000238)	(0.000596)	(0.000775)	(0.000255)	(0.000641)	(0.00110)
Constant	-0.180	1.920***	0.541	-0.773***	0.389	-2.281*
	(0.191)	(0.509)	(0.666)	(0.162)	(0.395)	(1.041)
Observations	3862	14304	14873	14092	14763	8022
R^2	0.303	0.137	0.171	0.137	0.134	0.209

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. OLS estimations are carried out separately for humanities and social science course grades (columns 1-3) and math and science course grades (columns 4-6). Estimations are further stratified by the size of the coursemate-peer group associated with each grade observation (sizes 2-10, 11-30, and over 30). Coefficients for academic year dummy variables are included in the estimations but are not shown. Standard errors are clustered by company-academic year groups.