Peer and Leadership Effects in Academic and Athletic Performance

July 13, 2006

Scott E. Carrell*
Dartmouth College

Robert N. Gilchrist*
Adams State College

Richard L. Fullerton*
USAF Academy

James E. West*
USAF Academy

Abstract: Students at the U.S. Air Force Academy are randomly assigned to one of 36 squadrons. In these squadrons, students live in adjacent dorm rooms, dine together, compete in intramural sports together and perform military training together. Due to this unique policy, we are able to identify the *known* peer group with which students spend a majority of their time. Using the squadron as the peer group, we find peer effects of much larger magnitude than those found in the previous literature. For freshmen students, a 1-point increase in peer grade point average (GPA) increases individual GPA by 0.64 to 0.73 grade points. We also find evidence of positive leadership effects from the upper class "supervisors" within the squadron. A 1-point increase in the junior class GPA increases individual freshman GPA by 0.23 grade points. Both the peer and leadership effects continue into the sophomore year after random reassignment to a new squadron, providing evidence of persistence in the effects. Our results demonstrate the importance of properly identifying the relevant peer group when estimating peer effects. As evidence of this, we find that geographic proximity of students in dorm halls alone, as in Foster (forthcoming), does not generate measurable peer effects. We find smaller peer effects at the roommate level, which are nearly identical in magnitude to those found by Sacerdote (2001) at Dartmouth; and, these roommate effects virtually disappear once we control for squadron effects.

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

JEL Classifications: Z13
Key Words: Peer Effects

* Carrell: Department of Economics, Dartmouth College, 301 Rockefeller Hall, Hanover, NH, 03755 (e-mail: scott.carrell@dartmouth.edu); Fullerton and West: Department of Economics and Geography, USAF Academy, 2354 Fairchild Hall, USAFA, CO 80840 (e-mail: richard.fullerton@usafa.af.mil & jim.west@usafa.af.mil); Gilchrist: Department of Chemistry, Computer Science, and Mathematics, Adams State College, 208 Edgemont Blvd. Alamosa, CO 81102. Special thanks goes to USAFA personnel: Col William Carpenter, Rolland Stoneman, Kathleen O'Donnell, Jau Tsau, and Kate Carson for their assistance in obtaining the data and background information required for this project. Thanks also go to Doug Staiger, Bruce Sacerdote, Omari Swinton, and Changhui Kang for helpful comments and suggestions.

I. Introduction

Justification for education policy decisions such as integration, busing, school choice, ability grouping, and affirmative action in admissions decisions are predicated upon the assumption of large positive peer effects in educational outcomes. To date, empirical researchers have only produced modest evidence of positive peer effects in educational outcomes. The most convincing studies, in which students have been randomly assigned to roommates or classrooms have typically found only very small, positive, and nonlinear peer effects (see Sacerdote, 2001; Zimmmerman, 2003; Hoxby 2006). In another recent study, Foster (forthcoming) finds no evidence of peer effects in dormitory hall-wings at the University of Maryland and draws into question the very existence of peer effects in higher education.

Researchers studying peer influence in higher education have typically assumed peer group formation at the roommate, dorm floor, or dorm level (Sacerdote, 2001; Zimmerman, 2004; Kremer & Levy, 2003; Foster, forthcoming). Anecdotal evidence would suggest that college students quickly establish networks of friends and study partners that extend beyond the roommate, dorm floor, or dorm level. To the extent this is true, works in the previous literature have likely underestimated the true magnitude of peer effects, as the influence of peers who reside outside these more narrowly measured groups would be omitted.

Previous works estimating peer effects in higher education typically report reduced form models in which own academic performance is a function of exogenous characteristics of peers (Sacerdote, 2001; Zimmerman, 2004; Kremer & Levy, 2003; Foster, forthcoming). As detailed by Manski (1993), these techniques are useful in testing for the presence of peer effects. But, absent econometric identification of underlying structural parameters, reduced form models do

not decompose endogenous from exogenous peer effects, nor do they accurately estimate the magnitudes.

The statistical properties of our data set enable us to address the limitations of previous studies. Students at the United States Air Force Academy (USAFA) are randomly assigned to one of 36 squadrons. The students live in adjacent dorm rooms, dine together, compete in intramural sports together and perform military training together. As a result, the squadron to which an individual student belongs, made up of roughly 120 students, comprise the *known* peer group in which a student spends a vast majority of his/her time. As students have no ability to influence the squadron into which they are placed, self-selection is not present. USAFA collects and maintains copious amounts of demographic data and high school (hence exogenous) performance data on all students. This enables us to properly identify structural equations and estimate peer effects using 2 stage least squares.

Members of each squadron perform various leadership roles within the squadron based on their relative seniority (freshmen, sophomore, junior, or senior class). Through the end of March, freshmen students are not allowed to enter the premises of another squadron. Hence, interaction with students from other squadrons is extremely limited for the freshmen.[1] At the start of the sophomore year, each student is then randomly reassigned to a new squadron and remains in that squadron for the remaining three years. This practice originated in response to

---

[1] Students are intermixed during academic classes and can meet with students from other squadrons at the library, gym, church, and what would be considered the student union. Additionally, freshmen students who are on intercollegiate athletic teams or participate in club sports are intermixed with students from other squadrons during practice times and on team trips.

the 1965 USAFA cheating scandal as an attempt to break up peer groups.[2] For our purposes, it enables us to test for the persistence of freshman peer effects on sophomore performance.

Similar to results found by Sacerdote (2001) and Zimmerman (2003), we find only moderate evidence of peer influence at the roommate level which virtually disappears when the broader squadron level peer performance is included as an explanatory variable. We view this result as empirical evidence of the importance of properly identifying the relevant peer group when estimating peer influence. Using the squadron as the peer group, we find peer effects of much larger magnitude than those found in the previous literature. For freshmen students, our models estimate that a 1-point increase in peer grade point average (GPA) increases individual GPA by 0.64 to 0.73 grade points on a scale of 0.0 to 4.0. Additionally, we find evidence of positive leadership effects from the upperclassmen "supervisors" within the squadron. A 1-point increase in the junior class GPA within a squadron increases individual freshman GPA by 0.23 grade points. Both the peer and leadership effects from the freshman year continue into the sophomore year after reassignment to a new squadron, providing evidence of persistence in the effects. We also find similar results in athletic performance.

Our results show that peers play a substantial role in both academic and athletic outcomes in higher education. When roommates are defined as the peer group, we find that the magnitude of roommate peer effects at USAFA almost exactly matches that found by Sacerdote (2001) at Dartmouth.[3] But our results further demonstrate the importance of properly identifying the relevant peer group when estimating peer influence, as the magnitude of squadron peer effects we find is six-fold that of roommate peer effects.

---

[2] See Malmstrom (2006) for further details.
[3] Sacerdote (2001) found a 1-point increase in roommate GPA increased own GPA by 0.120 grade points (Table III, Specification 1). Our model estimates a roommate-only coefficient of 0.1186 (Table 4, Specification 1).

II. Air Force Academy Squadron and Rank Structure

The Air Force Academy is a fully accredited undergraduate institution of higher education with an approximate enrollment of 4,200 students. There are 32 majors offered including the humanities, social sciences, basic sciences, and engineering. The average SAT for the 2005 entering class was 1309 with an average high school GPA of 3.60 (Princeton Review, 2006). Applicants are selected for admission on the basis of academic, athletic, and leadership potential. In addition, applicants must receive a nomination from a legal nominating authority including Members of Congress, the Vice President, or President of the United States, and other related sources. All students attending the Air Force Academy receive 100% scholarship to cover their tuition, room, and board. Additionally, each student receives a monthly stipend of $845 to cover books, uniforms, computer, and other living expenses. All students are required to graduate within four years[4] and serve a five-year commitment as a commissioned officer in the United States Air Force following graduation.

All students are assigned to one of 36 squadrons, with each squadron comprised of approximately 120 students evenly split between freshmen, sophomores, juniors, and seniors. Upperclassmen within the squadron act as the military training instructors, called cadre, during "basic cadet training"[5] and serve in various leadership roles throughout the academic year. The seniors are the "leaders." Their primary role is to "develop" the juniors, "shape" the sophomores, and "inspire" the freshmen. The juniors are the "workers" within the squadron. Their primary role is to "develop" the sophomores and "train" the freshmen. In practice, the

---

[4] Special exceptions are given for religious missions, medical "set-backs", and other instances beyond the control of the individual.
[5] All incoming freshmen are required to attend a five-week basic training prior to the start of the first academic year.

juniors supervise the freshmen within the squadron. The sophomores are the "role models" within the squadron and act as mentors and "coach" the freshmen. Finally, the freshmen are the "followers" and "learn and live loyalty" and "lead by example" (ODS, 2004).[6]

In theory, any member of the squadron could potentially help a freshman student with his/her coursework. As freshman students are junior, probationary members of a squadron, we would expect the primary peer group of freshman students to be that of other freshman students within the same squadron. It is also plausible that more senior members of a squadron could provide academic assistance as well as being mentors and leaders to the freshmen.

III. Background

This paper is similar in nature to the works of Foster (2006), Sacerdote (2001), Kremer and Levy (2003) and Zimmerman (2003) in that it measures peer effects in academic outcomes for college students. It goes further in measuring peer effects in physical education outcomes and testing for leadership effects that persist through a student's undergraduate education at USAFA.

As in Carrell, Malmstrom, and West (2005), we consider three types of peer influence outlined by Manski (1993): 1) exogenous effects, 2) endogenous effects, and 3) correlated effects. Exogenous effects occur when individual behavior varies with the pre-treatment group characteristics (Manski, 1993). Endogenous effects occur when individual behavior varies with the during treatment behavior of the group. Finally, correlated effects are those driven by the self-selection of individuals into groups (Sacerdote, 2001).

Actual measurement of peer effects can prove difficult at times. Two main challenges exist. First, it is difficult to separate out the individual and group influence on one another

---

[6] Information obtained from, "Officers Development System: Building Leaders of Character", United States Air Force Academy, February 2004.

(Vidgor & Nechyba, 2004). This problem is often referred to as the endogeneity problem (Moffitt, 2001; Sacerdote, 2001) or the reflection problem (Manski, 1993). The second issue in measuring peer influence occurs because individuals tend to self-select into peer groups. In the presence of self-selection, it is difficult to distinguish the peer effects from the selection effects (Sacerdote, 2001).

The endogeneity problem is typically handled by finding suitable instruments for peer behavior that are exogenous with respect to the stochastic error component of the dependent variable. A more recent strategy in the education peer effects literature has used lagged peer achievement as an instrument for current achievement (Betts & Zau, 2004; Burke & Sass, 2004; Hanushek, et al., 2003; Vidgor & Nechyba, 2004). This is in accordance with sound econometric theory, since past achievement is historic with respect to current achievement, and therefore is plausibly exogenous. We use a similar methodology in our specifications, particularly when measuring sophomore achievement and leadership effects.

The selection problem has typically been resolved using situations in which a "natural experiment" occurs where individuals are randomly assigned to peer groups (Boozer & Cacciola, 2001; Foster, forthcoming; Sacerdote, 2001; Zimmerman, 2003). Another strategy, widely used in the primary education peer effects literature, is to exploit the variation across classrooms or cohorts within a school (see Hoxby, 2006; Vidgor & Nechyba, 2004; Betts & Zau, 2004; Burke & Sass, 2004; Hanushek, et al., 2003). This has typically been accomplished using large administrative panel data sets while employing a series of fixed effects models. Foster (forthcoming), Sacerdote (2001), and Zimmerman (2003) go through great lengths to show that freshman roommate assignments at their respective institutions are effectively random (or conditionally random); however, self-selection is less of an issue at USAFA since

freshman roommate and squadron assignments are accomplished without *any* input from freshman students. In attempting to develop an ability to work with peers of all abilities and backgrounds, USAFA does not ask any questions of incoming students as to their likes, dislikes, or roommate preferences. One might argue that the effect the institution is trying to achieve in bypassing student preferences (and, fortunately for us, self-selection bias) is a behavioral model similar to the Rainbow model outlined in Hoxby (2006). When reassigning students into new squadrons for their sophomore year, USAFA does systematically distribute students with high military and/or academic performance across squadrons to prevent clustering of high achievers. As a result, we employ techniques similar to Sacerdote (2001) and Zimmerman (2003) to correct for this negative selection bias.[7]

We estimate peer effects using two separate approaches; reduced form equations, and two-stage least squares. In the first approach, we regress individual outcomes on pre-treatment variables to avoid simultaneous equation bias or the reflection problem (Sacerdote, 2001; Zimmerman, 2003). We use a variety of own, roommate, peer (other freshmen in squadron), and upperclassmen pre-treatment variables. Freshman GPA is presumed to be exogenous with respect to such variables as SAT scores (both math and verbal), academic composite (to include high school GPA, class rank, quality of school, size of school), fitness scores, and leadership composite scores required for entry to USAFA. We do not attempt to determine any non-linear effects by stratifying our exogenous explanatory variables as in Hoxby (2006), Sacerdote (2001), and Zimmerman (2003). Our specification uses the linear-in-means model common to the peer effects literature. While we recognize the limitations of linear-in-means models (Hoxby, 2006), we nevertheless use it for ease of comparison to the existing literature. In our

---

[7] A full discussion of our data and potential selection bias is conducted in the data section of the study.

8

second set of specifications, we identify the endogenous peer effect by specifying the freshman GPA as a function of roommate, peer (other freshmen), current upperclassmen freshman GPA, and own pre-treatment variables. We estimate these equations using two-stage least squares (2SLS) as in Foster (forthcoming) and Hoxby (2006) with all roommate, squadron level peer, and upper class average pre-treatment and demographic characteristics as first stage regressors.

In general, we find strong, robust peer effects of much larger magnitude than those found in previous studies, especially Foster (forthcoming), who defines her peer group as a hall-wing of a dormitory floor at the University of Maryland. We credit this to the copious amounts of data that USAFA keeps on all students and the nature of the squadron structure, which allows us to cleanly identify the group of possible peers for freshman students. Within squadrons, significant amounts of social, academic, athletic, and leadership interactions take place early and often, which forms a solid foundation for the "total peer effect" (Sacerdote, 2001).

Despite the more involved admissions process of seeking an appointment, we believe students at the Air Force Academy are largely drawn from the same pool as other selective academic institutions throughout the United States.[8] The appointment process ensures broad geographic diversity in the student population. Informal discussions with current and former students reveals many applied to a broad variety of universities, selecting the school they perceived gave them the "best deal." As admission to USAFA is based on academic, leadership, and athletic potential, applicants are not uniformly gifted in any one area. Some candidates may be relatively weaker with respect to their peers in certain areas, but stronger in others. Accordingly, there is substantial variation in pre-treatment variables despite the

---

[8] A possible exception would be an overrepresentation of students from military families. In the classes of 2000-2007, 16.5% of students had one or both parents in the military.

selectivity of the institution (selectivity comparable to Dartmouth and Williams) (U.S. News, 2006).[9] This allows us to test for the presence of social interactions between individuals and their roommates, classmates, and leaders within the squadron.

IV. Data & Peer Group Assignment

Data

Data for our study consists of individual level admissions and performance data during the freshmen and sophomore years for the U.S. Air Force Academy (USAFA) graduating classes of 2000 – 2007. For each individual, our data set contains high school (pre-treatment) measures of academic, athletic, and leadership aptitude as well as the individual's race/ethnicity, gender, whether they were a recruited athlete, and whether they attended a military preparatory school.

Our measure of pre-treatment academic aptitude is an *academic composite* computed by the USAFA admissions office, which is a weighted average of an individual's SAT or ACT score, high school GPA, class rank, and the quality of the high school attended. The sample mean academic composite is 3,251 with a standard deviation of 295. Our measure of pre-treatment athletic aptitude consists of a score on a fitness test (*fitness score*), required by all applicants prior to entrance.[10] The sample mean fitness score is 460 with a standard deviation of 97. Our measure of pre-treatment leadership aptitude is a *leadership composite* computed by the USAFA admissions office, which is a weighted average of high school and community activities (e.g., student council offices, Eagle Scout, captain of sports team, etc.) The sample

---

[9] All three schools (USAFA, Williams and Dartmouth) are designated "most selective" by the U.S. News and World Report's "Americas Best Colleges 2006". This rating is based on a formula that "accounts for enrollees' test scores and class standing and the school's acceptance rate (the percentage of applicants who are accepted)."

[10] The fitness score measures timed scores in pull-ups, sit-ups, push-ups and a 600-yard shuttle run, in addition to a standing long jump and a basketball throw.

mean leadership composite is 1,724 with a standard deviation of 183. Eighteen percent of the student population is female, 5-percent is black, 6-percent is Hispanic and 5-percent is Asian. Twenty-seven percent are recruited athletes and 2-percent attended a military preparatory school. Additionally, 7-percent of students at USAFA have a parent who graduated from a service academy and 17-percent have a parent who served in the military.

Our data contains each individual's freshmen and sophomore (during treatment) academic and athletic performance measured by a grade point average (GPA) and a physical education average (PEA).[11] Both the GPA and PEA are computed on a zero to 4.0 scale. The GPA comprises traditional academic coursework, while the PEA consists of scores on a physical fitness test (pull-ups, long jump, sit-ups, push-ups, and a 600-yard run), time on an aerobic fitness test (1.5 mile run), and grades in physical education courses.

Nearly all students at USAFA spend their entire freshmen year taking required core courses and have virtually no ability to select their own coursework. The USAFA Registrar generates the fall semester academic schedules for the freshmen without any input from the affected students (the one exception is the choice of the foreign language requirement). Nor can students choose their professors. Core courses are taught in small sections with all faculty using the same syllabus and giving the same exam during a common testing period. This institutional characteristic virtually assures there is no self-selection of students into courses or towards certain professors and, consequently, GPA is a consistent measure of academic performance across all students in our sample. Data on pre-treatment and during treatment performance was provided by USAFA Institutional Research and Assessment and de-identified

---

[11] Students also earn a military performance average (MPA); however, we do not use this measure because military performance is primarily determined within the squadron through peer and leadership evaluations (i.e., room inspections, squadron scores in marching, etc.).

by the USAFA Institutional Review Board. A complete list of summary statistics is provided in Table 1.[12]

Random Squadron and Freshmen Roommate Assignment

Each student is randomly assigned to one of 36 squadrons at the start of their freshmen year.[13] Freshmen fall semester roommates are then assigned within each squadron without input from the affected students. We were not able to find any official USAFA records for freshmen roommate assignment; however, using a log of issuing and returning dorm room keys, we were able to successfully match approximately 2/3 of freshman students as roommates. We considered individuals as roommates if students were issued a key to the same room for a minimum of 2 overlapping months.

Students are re-assigned to a new squadron at the start of their sophomore year and remain in that squadron for the next three years. In addition to the sorting mechanisms used to assign students to their freshmen squadron to ensure diversity (sibling, same last name, gender, race, athlete, military preparatory school), students with a 3.5 or greater cumulative freshmen GPA (approximately 16 percent of all students) or a cumulative freshmen military performance average (MPA) of 3.15 or greater (approximately 17 percent of all students) are blocked and randomly assigned to a sophomore squadron first. This mechanism ensures a relatively even spread of the top performers across all 36 squadrons. Of prime importance to our study is that students are not placed into squadrons or with (freshmen) roommates based on pre-treatment

---

[12] As fully discussed in the next section, due to concerns with potential non-random placement of students into squadrons prior to the class of 2005, the summary statistics provided only include the graduating classes of 2005-2007.

[13] The algorithm that assigns students into squadrons prevents siblings as well as students within the same graduating class, with the same last name (i.e., Smith), from being placed in the same squadron. Additionally, females, minorities, athletes, and students who attended a military preparatory school are randomly sorted into squadrons first, to ensure diversity across squadrons.

characteristics. To test for randomness in our squadron and roommate assignments, Table 2 shows results, by graduating class, when regressing individual pre-treatment characteristics on roommate and peer pre-treatment characteristics (*academic composite, fitness score,* and *leadership composite*).

Freshmen squadron placements were unavailable for the graduating classes of 2000, 2001, and 2003; therefore, results for these classes only include sophomore squadron assignments. The negative and significant coefficients on the freshman squadron peer academic and peer athletic composite variables for the classes of 2002 and 2004 indicates a negative selection effect on freshmen squadron placements during these years. That is, it appears USAFA personnel may have sorted students into squadrons based on pre-treatment characteristics with the intention of balancing each squadron's overall academic and/or athletic ability.[14] Sophomore squadron placements appear to have this same negative selection for the class of 2003.

However, there appear to be no squadron selection effects in the data for the classes of 2005 through 2007, as indicated by the statistically insignificant coefficients in all regressions for these years, but one. At the roommate level, the one exception is a positive and significant coefficient on the roommate fitness score for the class of 2007, indicating a potential positive selection of roommates on athletic ability. However, this positive coefficient diminishes and is statistically insignificant when including a squadron fixed-effect, indicating that within squadrons, where roommates are assigned, there appears to be no positive selection.

---

[14] We currently do not know if the algorithm used to place students into squadrons was different during those years compared to the current algorithm provided to us.

Based on these findings we chose to estimate all our empirical models using the outcomes for the classes 2005 through 2007 only.

IV. Methods and Results

Reduced Form Estimation of Peer Effects

We begin by analyzing the peer and leadership effects using the traditional reduced form linear-in-means model where we regress individual outcomes on roommate and peer pre-treatment characteristics. Specifically, we estimate the following equation for academic performance:

$$(1) \quad GPA_{isc} = \phi_0 + \phi_1 X_{isc}^r + \phi_2 \frac{\sum_{k \neq i} X_{ksc}}{n_{sc} - 1} + \beta X_{isc} + \varepsilon_{isc},$$

where GPA_{isc} is the freshmen fall semester GPA for individual i in squadron s, and graduating class c. X_{isc}^r are the pre-treatment characteristics of individual i's roommate[15] and $\frac{\sum_{k \neq i} X_{ksc}}{n_{sc} - 1}$ are the average pre-treatment characteristics of all other classmates in squadron s except individual i. X_{isc} is a vector of individual i's specific (pre-treatment) characteristics, including academic composite, fitness score, leadership composite, race/ethnicity, gender, recruited athlete, and whether they attended a military preparatory school. ε_{isc} is the error term. We include graduating class fixed effects to control for unobserved mean differences across years in GPA. Given the potential for error correlation across individuals within a given squadron and class, we correct all standard errors to reflect clustering at the squadron by class level.

---

[15] Average GPA is used for individual with two roommates.

Reduced Form Results

We estimate various specifications of equation (1) using ordinary least squares (OLS) for freshmen academic performance, with results shown in Table 3.[16] To provide a basis of comparison to other existing studies, Specification 1 estimates the peer influence at the roommate level only using the SAT verbal and SAT math scores as the primary explanatory variables.[17] Results for this specification are generally uninteresting, with small, negative, and insignificant coefficients on both the Roommate SAT verbal and math scores. Own SAT verbal and SAT math scores are positive and highly significant and provide only a moderate-level of explanatory power ($R^2 = 0.20$).

Specification 2 repeats this analysis, but uses the full array of pre-treatment measures by adding the roommate and own academic composite, fitness score, and leadership composite as explanatory variables to the model. Results for this specification show insignificant coefficients for the roommate academic composite and fitness score variables; however, the coefficient on the roommate leadership composite is positive and significant (0.0126). The effect is relatively small; the model predicts a one-standard deviation increase in the roommate leadership composite results in an increased freshmen fall semester GPA of 0.02 grade points. Additionally, the inclusion of the full array of pre-treatment measures provides a better fit to the data ($R^2 = 0.34$). Both the own academic composite (0.1090) and fitness score (0.0449) are positive and highly significant. Own SAT math and the leadership composite are positive and statistically insignificant and own SAT verbal is negative and significant (-0.0498). This negative result implies that the formula used to compute the academic composite likely places

---

[16] SAT scores, academic composite, leadership composite and fitness scores have all been divided by 100 prior to estimating the regressions.

[17] For student who only have a reported ACT score, we converted the ACT scores to SAT scores using conversions from the College Board (Dorans, 1999).

too much weight on the SAT verbal score. We chose to enter the SAT verbal and math scores in addition to the academic composite as they provide additional explanatory power to the model.[18]

Specification 3 adds to the model the average pre-treatment characteristics of individual i's peers (other freshmen) in squadron s. Of the five additional peer variables included in the specification, only peer SAT verbal is statistically significant. The coefficient (0.4218) is large, positive, and statistically significant at the 0.05-level. Compared to previous studies the magnitude of this effect is quite large, providing empirical evidence that this broader peer group plays a more important role than just roommates. Similar to Zimmerman (2003), the reduced form academic peer effect appears to be driven through SAT verbal scores versus other academic pre-treatment measures. The model predicts a 1-standard deviation increase in the peer SAT verbal score results in an increased own GPA of 0.05 grade points.

Finally, Specification 4 adds to the model the average pre-treatment characteristics of all the upper three classes in squadron s to measure the leadership effects from the upperclassmen within the squadron. In total, we estimate 25 different effects with 5 each for roommate(s), peers, sophomores, juniors, and seniors within the squadron. Overall, there are five positive and statistically significant coefficients: 1) roommate leadership composite (0.0118), 2) peer SAT verbal (0.4816), 3) peer fitness score (0.1536), 4) sophomore class SAT verbal (0.2903), and 5) junior class leadership composite (0.1045). The sophomore class leadership composite (-0.0748) is small, negative, and statistically significant at the 0.10-level. The positive results for the roommate leadership composite, peer SAT verbal, and peer fitness

---

[18] Because we are not privy to the formula used to calculate the academic composite, we cannot fully decompose the academic composite to exclude SAT scores. Based on our own regressions, we know that the SAT math and verbal scored received different weights and predict less than 50% of the variation in academic composite.

test variables provide evidence of positive peer influence. Furthermore, the positive results for the sophomore class SAT verbal and junior class leadership composite variables provide evidence of positive leadership effects within the squadron.

Although these results provide evidence of positive social spillovers in academic performance, we find the results somewhat unsatisfying.[19] As in Zimmerman (2003) we find the peer effects are linked more closely with SAT verbal scores versus other academic pre-treatment measures. These results also show that other non-academic measures, such as athletic and leadership measures, appear to be linked with positive peer influence; however, it is difficult to theoretically explain why each of these effects should be significant compared to those that are insignificant. Two possible explanations arise. First, the insignificant coefficients may be due to non-linearities in the effects across different types of individuals (i.e., ability, race, or gender). For example, Hoxby (2006) finds strong evidence of non-linearities in peer influence across high versus low achieving students in elementary and middle school. Second, it could be that the positive coefficients on the pre-treatment variables are estimating a purely endogenous effect. Sacerdote (2001) supports this hypothesis in finding that peer effects at Dartmouth are primarily driven through roommate performance versus roommate background characteristics. Expanding upon Manski (1993), if own academic performance is affected by contemporaneous peer academic performance but not pre-treatment peer attributes, and academic performance in general is a function of SAT scores, leadership composite, etc., then pre-treatment peer attributes are indirectly correlated with own academic performance via peer academic performance. A regression of own academic performance on

---

[19] For brevity we do not show the reduced form estimates on athletic performance. Results for these specifications are generally uninteresting, with only one positive and statistically significant effect (junior class leadership composite).

peer pre-treatment characteristics will estimate this indirect correlation. But, absent econometrically identified structural parameters, magnitudes of the underlying peer effects cannot be computed.

To estimate own freshman academic performance as a direct function of peer academic performance, we estimate structural parameters using 2 stage least squares (2SLS) with peer background characteristics as exogenous instruments. Manski (1993) refers to this as an endogenous peer effects model. By assuming peer background characteristics do not affect own freshman academic performance directly, structural parameters of the model are overidentified (Moffitt, 2001).

Estimation of Endogenous Peer & Leadership Effects for Freshmen Students

For freshmen students, we estimate the following model using two-stage least squares (2SLS) with the following explanatory variables:

$$(2) \quad GPA_{isc} = \alpha_0 + \alpha_1 GPA_{isc}^r + \alpha_2 \frac{\sum_{k \neq i} GPA_{ksc}}{n_{sc} - 1} + \alpha_3 \overline{FreshGPA}_{sc-1} + \alpha_4 \overline{FreshGPA}_{sc-2} +$$

$$\alpha_5 \overline{FreshGPA}_{sc-3} + \beta X_{isc} + \varepsilon_{isc}$$

where GPA_{isc} is the freshmen, fall semester, GPA for individual i in squadron s, and graduating

class c. GPA_{isc}^r is the GPA of individual i's roommate[20] and $\frac{\sum_{k \neq i} GPA_{ksc}}{n_{sc} - 1}$ is the average GPA of

all other freshmen peers in squadron s except individual i. As both roommate and squadron classmate GPA are endogenous to our dependent variable, we instrument for GPA_{isc}^r and

---

[20] Average GPA is used for individual with two roommates.

$\dfrac{\sum\limits_{k\neq i} GPA_{ksc}}{n_{sc}-1}$ using all roommate and squadron level peer and upper class average pre-treatment

and demographic characteristics.[21]

$\overline{FreshGPA}_{sc-1}$, $\overline{FreshGPA}_{sc-2}$, and $\overline{FreshGPA}_{sc-3}$, are the average freshmen cumulative GPA for the sophomores, juniors, and seniors in squadron s, respectively. Because these GPAs, high school performance data, and demographic characteristics were known historical data as of time period c, they are formally exogenous with respect to the dependent variable, GPA_{isc}. X_{isc} is the vector of individual specific (pre-treatment) characteristics for individual i and ε_{isc} is the error term. We include graduating class fixed effects to control for unobserved mean differences across years in GPA and we correct all standard errors to reflect clustering at the squadron by class level.

When performing 2SLS estimation, the strength of first stage excluded instruments is of critical importance in obtaining consistent estimates. (Staiger and Stock, 1997; Stock, Wright, and Yogo, 2002; Shea, 1997; Hahn and Hausman, 2003). If instruments are weak, 2SLS estimated coefficients are biased toward inconsistent OLS estimates. Following one definition of weak instruments provided by Stock, Wright, and Yogo (2002), instruments are considered weak if the bias of 2SLS estimates under weak instruments relative to the inconsistency of OLS estimates exceeds 10%. A null hypothesis of weak instruments can be rejected in favor of strong instruments if the F-statistic measuring joint explanatory power of exogenous

---

[21] The complete set of instruments includes roommate and each class's average: academic composite, fitness score, leadership composite, SAT Verbal, SAT Math, black, Hispanic, Asian, female, attended a military preparatory school, and was a recruited athlete. Roommate demographic characteristics are entered as dummy variables and class demographic characteristics are in percentages.

instruments excluded from the final structural equation is sufficiently large, around 10.[22] In the presence of multiple endogenous explanatory variables, individual F-statistics computed for each explanatory variable is insufficient to assess the strength of the instruments should the instruments be sufficiently collinear (Shea, 1997). For our specifications which contain multiple endogenous explanatory variables (roommate and squadron peer effects), we provide the Cragg-Donald weak identification statistic. The relevant critical value for the bias of 2SLS estimates to be 10% of the inconsistency of OLS estimates given our large number of instruments is 11.05.[23] When using the full array of exogenous instruments available, our instruments are not weak, implying the bias of our 2SLS estimates is less than 10% of inconsistency of OLS estimates at a high degree of statistical significance.

Table 4 presents results for freshmen academic performance and Table 5 presents results for freshmen athletic performance, where we estimate equation (2) replacing all grade point averages (GPAs) with physical education averages (PEAs). Empirical studies have shown Limited Information Maximum Likelihood (LIML) estimation to be more robust to weak instruments than 2SLS (Staiger and Stock, 1997; Stock, Wright, and Yogo, 2002). For comparison , we provide LIML estimates of these specifications in Appendix A.

Results for Freshmen Academic Performance

Table 4, Specification 1, estimates the peer influence at the roommate level only. The positive and significant coefficient (0.1186) on roommate GPA indicates that, on average, an individual's GPA increases 0.07 grade points with a 1-standard deviation increase in roommate GPA. The magnitude of the effect is nearly identical to that found by Sacerdote (2001); however, our vast array of exogenous instruments allows us to correct for the endogeneity of

---

[22] Critical values can be found in Stock, Wright, and Yogo (2002), Table 1.
[23] Critical values can be found in Stock and Yogo (2002), Table 1.

roommates performance using 2SLS estimation.[24]   Results also show that own academic

composite and fitness score are positive and highly significant, while the own leadership

composite is statistically insignificant.   The model estimates that a 1-standard deviation

increase in the academic composite results in a 0.32 increase in GPA and a 1-standard deviation

increase in the fitness score results in a 0.04 increase in GPA.

Next, Specification 2 adds to the model the average GPA of all other freshmen in squadron

s, except individual i (Peer GPA).   The result for the Peer GPA variable (0.6437) is large,

positive, and highly significant. The magnitude of the coefficient on roommate GPA (0.0449)

diminishes and is no longer statistically significant.   The model estimates a 1-standard

deviation increase in peer GPA results in a 0.10 increase in own GPA.   This result provides

strong evidence that the broader peer group of all freshmen within the squadron play a more

important role in academic performance than just that of roommates.   The result shows the

importance of properly identifying the relevant peer group when estimating peer influence.

Hence, previous studies, which have assumed peer group formation at the roommate, dorm

floor, or dorm level, have likely underestimated the magnitude of the peer effects.[25]

To estimate the leadership effects within the squadron, Specifications 3 adds the average

freshmen cumulative GPA of the sophomore, junior and senior class within the squadron.

Results for all three upperclassmen GPA variables are positive, but only the coefficient on the

junior class GPA (0.2283) is statistically significant.   The model estimates a 1-standard

deviation increase in the junior class (freshmen) GPA results in a 0.02 increase in own GPA.

---

[24] Sacerdote (2001) found a 1-point increase in roommate GPA resulted in a 0.120 increase in
own GPA.
[25] In alternate specifications (not shown) we estimate the model without roommate
characteristics to include those students in the squadron in which we were unable to match
roommates. The result for the Peer GPA variable (0.6020) is large, positive, and highly
significant.

Results for the Peer GPA variable remain positive and highly significant. The model estimates that a 1-point increase in peer GPA increases individual GPA by 0.6577 grade points. As a final check, Specification 4 adds to the model a squadron fixed effect, which controls for any unobserved time invariant differences across squadrons (e.g., culture, tradition, etc.). Results for the peer remain positive and highly significant with the coefficient on the peer GPA variable (0.7339) increasing in magnitude. Results for the junior class GPA (0.1188) remains positive, but is no longer statistically significant.

The results in Table 4 provide strong evidence of peer influence in academic performance at the squadron by classmate level and positive leadership effects from the junior class within the squadron. The results are much larger in magnitude than previous studies, which we attribute to proper identification of the relevant peer group in our estimations. Next, we repeat the analysis for freshmen athletic performance

Results for Freshmen Athletic Performance

Table 5, Specification 1, estimates the peer influence at the roommate level only. The positive and significant coefficient (0.0974) on roommate PEA indicates that, on average, an individual's PEA increases 0.05 points with a 1-standard deviation increase in roommate PEA. Results also show that the own academic composite, CFT score, and leadership composite are all positive and significant in predicting athletic performance. The model estimates that a 1-standard deviation increase in the academic composite results in a 0.09 increase in PEA, a 1-standard deviation increase in the CFT score results in a 0.24 increase in PEA, and a 1-standard deviation increase in the leadership composite results in a 0.04 increase in PEA.

Specification 2 adds to the model the Peer PEA variable. Results for this coefficient (0.4164) are positive and highly significant while the magnitude of the coefficient on the

roommate PEA (0.0572) diminishes and is no longer statistically significant. The estimated effect indicates a 1-standard deviation in peer PEA results in a 0.05 increase in own PEA. Again, this result provides further evidence that the broader peer group plays a more important role in predicting performance and exemplifies the importance of proper identification of the relevant peer group when estimating peer effects.

Specifications 3 adds the average freshmen cumulative PEA of the sophomore, junior, and senior classes within the squadron. The Peer PEA variable remains positive and highly significant, with only small changes in the magnitude of the effect. Results again show that the junior class has a positive leadership effect on freshmen performance (0.1446). The model estimates a 1-standard deviation increase in the junior class PEA increases individual PEA by 0.02 grade points. Finally, Specification 4 adds to the model a squadron fixed effect. Results for both the peer PEA variable (0.3940) and the junior class PEA (0.1935) remain positive and highly significant with small changes in the magnitude of the effects.

Results in Tables 4 and 5 provide strong evidence of peer and leadership influences in both academic and athletic performance. Similar to previous studies, we find moderate evidence of peer influence at the roommate level. These roommate effects virtually disappear once we estimate the effects at the proper peer group (squadron) level. Our models estimate that a 1-point increase in peer GPA increases individual GPA by 0.64 to 0.73 grade points and a 1-point increase in the junior class GPA within a squadron increases individual GPA by 0.23 grade points. We also find similar results for athletic performance.

We attribute these results to the proper identification of the relevant peer group when estimating peer effects. Unlike Foster (forthcoming), where peer group formations were assumed to form in dorm "hall-floor wings," the squadron structure at USAFA allows us to

identify the *known* peer group in which students spend a majority of their time. To test this assertion, we next conduct falsification tests by computing artificial or false peer groups using students from different squadrons whose dorm rooms are geographically co-located.

Falsification Tests

The unique dorm structure at USAFA provides the opportunity to empirically test for false peer effects. All 4,200 students at USAFA live in one of only two dorm halls. Squadrons 1-21 reside in Vandenberg Hall and squadrons 22-36 reside in Sijan Hall. While all members of a respective squadron are geographically co-located in the same area of the dorm, squadrons located in the same dorm hall and floor are adjacent to one another with no visible partitions. Therefore, to test for the importance of proper identification of the relevant peer group, we are able to construct false peer groups of students whose dorm rooms are located in the same section of the dorm hall, but are not necessarily in the same squadron. We construct these groups using student dorm room assignments at the start of the fall semester. Each dorm room is identified by the hall (Vandenberg or Sijan), floor (2, 3, 5, and 6), section (A to G), and room number. In total, there are 39 identifiable dorm/floor/sections with which we construct false peer groups. These groupings are analogous to hall-floor wings as defined by Foster (forthcoming). During the three years in our sample, 92.3% of the hall/floor/sections contain students from different squadrons. We construct and test for two separate false peer groups: 1) freshmen students in the same hall/floor/section, and 2) all students within the same hall/floor/section.

Table 6 presents results for this analysis for freshmen student outcomes. Specifications 1-2 show results for academic outcomes and Specifications 3-4 show results for athletic outcomes. In all four specifications, the average performance (GPA or PEA) of the false peer

group has no statistically significant effect on individual performance. Similar to results found by Foster (forthcoming), these results show that geographic proximity of individuals alone does not generate positive peer effects. Hence, the results provide further empirical evidence of the importance of *properly* identifying the relevant peer group when estimating peer effects.

Estimation of Peer & Leadership Effects for Sophomore Students

With evidence of positive peer and leadership effects in freshmen academic and athletic performance, we look for persistence of freshman peer effects in sophomore performance. It is possible to statistically separate freshman peer effects from sophomore peer effects on sophomore performance because all students are (conditionally) randomly assigned to a new squadron at the beginning of their sophomore year.

For sophomore academic performance we again estimate a purely endogenous peer effect using 2SLS on the following model:

$$(3) \quad GPA_{isc} = \lambda_0 + \lambda_1 \frac{\sum_{k \neq i} GPA_{ksc}}{n_{sc} - 1} + \lambda_2 \frac{\sum_{k \neq i} GPA_{ks-1,c}}{n_{s-1c} - 1} + \lambda_3 \overline{FreshGPA}_{s-1,c-2} + \lambda_4 \overline{FreshGPA}_{sc-1} +$$
$$\lambda_5 \overline{FreshGPA}_{sc-2} + \beta X_{isc} + v_{isc}$$

where, GPA_{isc} is the sophomore, fall semester, grade point average for individual i in squadron s, and graduating class c. As roommates are not randomly assigned for sophomore students, we are unable to estimate roommate level peer effects. $\dfrac{\sum_{k \neq i} GPA_{ksc}}{n_{sc} - 1}$ is the average GPA of all other sophomores in squadron s except individual i and $\dfrac{\sum_{k \neq i} GPA_{ks-1,c}}{n_{s-1c} - 1}$ is the average (freshmen) GPA for all other classmates in individual i's freshman year squadron. As both

$$\frac{\sum\limits_{k \neq i} GPA_{ksc}}{n_{sc} - 1} \text{ and } \frac{\sum\limits_{k \neq i} GPA_{ks-1,c}}{n_{s-1c} - 1}$$ are endogenous with respect to the dependent variable, we

instrument using all current and previous year squadron average pre-treatment and

demographic characteristics. $\overline{FreshGPA}_{s-1,c-2}$ is the average cumulative GPA of the junior

class in individual *i's* previous freshmen squadron and $\overline{FreshGPA}_{sc-1}$ and $\overline{FreshGPA}_{sc-2}$ are

the average freshmen cumulative GPA of the junior and senior class in individual *i's* current

squadron. Because these GPAs are all historical data relative to the dependant variable, they

are by definition exogenous with respect to the dependent variable. X_{ic} is the vector of

individual specific (pre-treatment) characteristics for individual *i*. We also include an indicator

variable for whether individual *i* had greater than a 3.499 GPA or greater than a 3.14 MPA

during their freshmen year as we know the assignment algorithm seeks to spread students with

high freshman year performance uniformly throughout all squadrons.[26] v_{isc} is the error term.

Again, we include graduating class year fixed effects and correct all standard errors to reflect

clustering at the squadron by class level.

 Table 7 presents results for sophomore academic performance and Table 8 reports results

for sophomore athletic performance.

Results for Sophomore Academic Performance

 Table 7, Specification 1 estimates the peer influence at the (current) squadron by classmate

level only. Results are similar to those found in the freshmen specifications, with the peer GPA

variable exhibiting a large positive and statistically significant coefficient (0.4742). The model

estimates a 1-standard deviation increase in peer GPA results in a 0.05 increase in own GPA.

---

[26] Our empirical estimates show that this selection mechanism reduced the variance in average Peer GPA across squadrons. Controlling for this observable selection mechanism in the regressions; therefore, provides unbiased estimates of the peer effect.

Next, to test for persistence in the peer and leadership effects from the freshman year squadron, Specifications 2 and 3 sequentially add to the model the freshman peer GPA and junior class GPA from individual i's freshmen year squadron.[27] For Specification 3, the positive and statistically significant coefficients for both the peer GPA (0.4440) and previous peer GPA (0.2957) indicate that both the current and previous peer groups exhibit positive peer influence. The coefficient on the previous peer GPA is roughly one-half the magnitude of that estimated during the freshmen year, indicating a likely diminishing of the previous peer group's influence. Additionally, the positive and significant coefficient on the previous year's junior class freshmen GPA (0.1612) also indicates persistence in the leadership effects from the previous year. Finally, Specification 4 adds to the model the current squadron junior and senior class (freshmen) GPA to test for leadership effects. Results for both variables are statistically insignificant, indicating the upperclassmen in the new squadron play a diminished role during the sophomore year.

We next estimate the effects during the sophomore year on athletic performance by estimating equation (3) replacing all GPA measures with the PEA.

Results for Sophomore Athletic Performance

Table 8 reports results for sophomore athletic performance. Results for Specification 1 show the peer PEA variable exhibits a moderately large positive and statistically significant coefficient (0.3815). The model estimates a 1-standard deviation increase in peer PEA results in a 0.04 increase in own PEA. Next, Specifications 2-4 sequentially adds to the model the previous peer PEA, previous junior class PEA, current junior class PEA and current senior class PEA. In all three specifications the previous peer PEA variable exhibits a positive and

---

[27] We instrument for the previous year peer GPA with the previous year squadron level pre-treatment characteristics.

statistically significant coefficient. For Specification 4, the estimated coefficient (0.2463) indicates a 1-standard deviation increase in the explanatory variable results in a 0.03 increase in own PEA. Similar to results found in academic performance, the magnitude of the effect is less than that estimated effect during the freshmen year, providing further evidence of a significant, but diminished effect from the previous peer group. Results for all the leadership variables from both the previous and current squadron are small, negative, and statistically insignificant.

The results shown in Table 7 and Table 8 provide evidence that both the current and previous peer group formations play an important role in both academic and athletic performance. The previous peer group's effect appears to diminish in size, but persists the following year after reassignment to a new peer group. Unfortunately, our data do not contain performance information beyond the sophomore year, so we are unable to estimate the persistence in the peer influence in later years.

The Social Multiplier

Manski (1993), Becker & Murphy (2000), Glaeser, Sacerdote, & Scheinkman (2003), Graham (2004), and others discuss the existence and estimation of a social multiplier in the presence of positive social spillovers. In the context of our paper, the social multiplier exists as peers (and leaders) with higher (or lower) performance impact one another. The modeling of a purely endogenous peer effect makes estimating the social multiplier particularly straightforward.

To estimate the social multiplier in freshmen academic performance we sum equation (2) over all individuals, divide by N, and solve for average GPA to yield:

$$(4) \quad \overline{GPA}_{sc} = \left(\frac{\alpha_0}{1-\alpha_1-\alpha_2}\right) + \left(\frac{\alpha_3}{1-\alpha_1-\alpha_2}\right)\overline{FreshGPA}_{sc-1} + \left(\frac{\alpha_4}{1-\alpha_1-\alpha_2}\right)\overline{FreshGPA}_{sc-2} + \left(\frac{\alpha_5}{1-\alpha_1-\alpha_2}\right)\overline{FreshGPA}_{sc-3} + \left(\frac{\beta}{1-\alpha_1-\alpha_2}\right)\overline{X}_{sc} + \varepsilon_{sc}$$

Hence, if an entire freshmen squadron cohort's *academic composite* increases by υ, the total effect on average squadron GPA will be $\upsilon * \left(\frac{\beta}{1-\alpha_1-\alpha_2}\right)$. The own effect is $\upsilon * \beta$ and the social multiplier is $\left(\frac{1}{1-\alpha_1-\alpha_2}\right)$. Using estimates from Table 4, Specification 3, our results indicate a social multiplier of $\frac{1}{1-0.0420-0.6577} = 3.33$. That is, if all freshmen in a squadron have a 1-standard deviation increase from the mean in the HS academic composite (295.6), the estimated increase in average squadron GPA would be 1.08, with 0.32 of the increase due to the increase in own academic ability and the remaining 0.76 due to social spillovers.

Our model and results also imply that these social spillovers apply to the leadership effects of the upperclassmen in the squadron. Using the estimated effect of the junior class GPA in Specification 3, our model estimates a 1-standard deviation increase in the freshmen cumulative GPA of all juniors in the squadron (0.599), would result in an increased average GPA of 0.46 for the (current) freshmen class. Again, this total effect can be separated into the direct effect (0.14) and that caused by the social spillover (0.32).

VI. Conclusion

We examine a data set of students from the graduating classes of 2000 – 2007 at the United States Air Force Academy for evidence of peer and leadership effects in academic and athletic

performance. Institutional characteristics of the Air Force Academy create a natural experiment for estimating peer influence. The overwhelming majority of entering students do not know anybody currently enrolled at USAFA. Sibling students are deliberately separated. The appointment process, by which each member of the U.S. Congress and Senate nominate candidates from their congressional district or state, insures geographic diversity. Squadrons are groups of approximately 120 students who live in adjacent dorm rooms, dine together, compete in intramural sports together and perform military training together. Through random roommate and squadron assignment of freshmen students and a random reshuffle into new squadrons at the start of the sophomore year, our data allow us to identify peer and leadership influences at three distinct peer-group levels: roommate pairs, squadron classmates, and squadron upperclassmen.

Using the squadron as the peer group, we find peer effects of much larger magnitude than those found in the previous literature. For freshmen students, our models estimate that a 1-point increase in peer grade point average (GPA) increases individual GPA by 0.64 to 0.73 grade points. Additionally, we find evidence of positive leadership effects from the upper class "supervisors" within the squadron. A 1-point increase in the junior class GPA within a squadron increases individual GPA by 0.23 grade points. Both the peer and leadership effects continue into the sophomore year after reassignment to a new squadron, providing evidence of persistence in the effects. Our results demonstrate the importance of properly identifying the relevant peer group when estimating peer effects. Because squadron placements are random, there are no self-selection effects in our peer estimates. Additionally, the use of high school leadership and academic composites, standardized test scores, and pre-entrance physical fitness scores allow us to correct for the endogeneity of individual and peer outcomes.

Bibliography

Becker, Gary S. and Kevin M. Murphy. (2000). Social Economics: Market Behavior in a Social Environment. Cambridge, MA: Harvard University Press.

Betts J.R. & Zau A. (2004). Peer Groups and Academic Achievement: Panel Evidence from Administrative Data. Unpublished Manuscript.

Boozer, M.A. & Cacciola, S.E. (2001). Inside the 'Black Box' of Project STAR: Estimation of Peer Effects Using Experimental Data. Unpublished manuscript.

Burke, M.A. & Sass, T.R. (2004). Classroom Peer Effects and Student Achievement. Presented at the American Economic Association Annual Meetings, Jan 2005.

Carrell, Scott E., Malmstrom, F.V., & West, J.E., (2005). Peer Effects in Academic Cheating. Presentation to NBER Higher Education Working Group.

Dorans, Neil J., (1999) Correspondences Between ACT and SAT I Scores. College Board Report No. 99-1, ETS RR No. 99-2.

Foster, Gigi, (2006). It's not your peers, and it's not your friends: some progress towards understanding educational peer effects. Forthcoming, Journal of Public Economics.

Glaeser, E.L., Sacerdote, B.L., and Scheinkman, J.A., (2003). The Social Multiplier. Journal of the European Economic Association, 1, 345 - 353

Graham, Bryan S. (2004). Identifying Social Interactions through Excess Variance Contrasts. Working paper.

Hahn, J. & Hausman J.H. (2003). Weak Instruments: Diagnosis and Cures in Empirical Econometrics. Recent Advances in Econometric Methodology for Policy Analysis, 93, 2, 118-123.

Hanushek, E.A., Kain, J.F., Markham, J.M. & Rivkin, S.G. (2003). Does Peer Ability Affect Student Achievement? Journal of Applied Econometrics, 18, 527-544.

Hoxby, Caroline M., and Weingarth, G., (2006). Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects. Working Paper.

Kremer, Michael, and Levy, Dan M. (2003). Peer Effects and Alcohol Use Among College Students. NBER Working Paper 9876.

Malmstrom, F.V. (March 2006). The 1965 USAFA Cheating Scandal, Checkpoints, (34)4, 36-41.

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

Moffitt, Robert A. (2001). Policy Interventions, Low-Level Equilibria, and Social Interactions. In Social Dynamics, edited by Seven N. Durlauf and H. Peyton Youg. Brookings Institution Press, Washington D.C.

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

Shea, J. (1997). Instrument Relevance in Multivariate Linear Models: A Simple Measure, The Review of Economics and Statistics, 79(2), 348-352.

Staiger, D. & Stock J.H. (1997). IV Regression with Weak Instruments. Econometrica, 65, 3, 557-86.

Stock, J.H., Wright J.H., & Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in GMM. Journal of Business and Economics Statistics, 20, 4, 518-29.

Stock, J.H., and Yogo, M. (2002). Testing for Weak Instruments in Linear IV Regression, Technical Working Paper 284, National Bureau of Economic Research

Stanard, C.I. & Bowers, W.J. (1970). The college fraternity as an opportunity structure for meeting academic demands. Social Problems, 17, 371-390.

U.S. News and World Report (2006). America's Best Colleges 2006. Accessed on 19 June 2006 at http://www.usnews.com/usnews/rankguide/rghome.htm

Vidgor, J. & Nechyba, T. (2004). Peer Effects in North Carolina Public Schools. Unpublished manuscript.

 Zimmerman, D.J. (2003). "Peer Effects in Academic Outcomes: Evidence From a Natural Experiment. The Review of Economics and Statistics, 85,1, 9–23

Table 1: Summary Statistics for Freshman Entering Classes of 2005-2007

Variable	Obs	Mean	Std. Dev.	Min	Max
Grade Point Average (GPA) (fall semester)	3407	2.88	0.62	0.28	4.00
Physical Education Average (PEA) (fall semester)	2878	2.52	0.52	0.65	4.00
SAT Math	3489	665.47	63.88	440	800
SAT Verbal	3489	631.95	67.00	330	800
Academic Composite	3489	3,245.30	295.63	2,252.00	4,023.00
Fitness Score	3489	459.70	96.88	215.00	745.00
Leadership Composite	3490	1,724.16	182.42	900.00	2,370.00
Black	3490	0.05	0.22	0	1
Hispanic	3490	0.06	0.24	0	1
Asian	3490	0.05	0.23	0	1
Female	3490	0.18	0.38	0	1
Recruited Athlete	3490	0.28	0.45	0	1
Military Preparatory School	3490	0.21	0.41	0	1
Freshmen Roommate GPA	2165	2.89	0.55	0.47	4.00
Freshmen Roommate PEA	1977	2.51	0.47	0.80	4.00
Freshmen Roommate Academic Composite	2171	3,248.88	261.87	2,304.00	3,973.00
Freshmen Roommate Fitness Score	2171	458.07	83.81	245.00	735.00
Freshmen Roommate Leadership Composite	2171	1,720.47	160.21	900.00	2,295.00
Peer GPA (squadron by class)	3490	2.88	0.15	2.41	3.37
Peer PEA (squadron by class)	3490	2.52	0.11	2.14	2.82
Peer Academic Composite (squadron by class)	3489	3,245.30	57.55	3,091.77	3,445.44
Peer Fitness Score (squadron by class)	3489	459.70	18.25	411.40	513.45
Peer Leadership Composite (squadron by class)	3490	1,724.16	31.70	1,613.23	1,809.58

Table 2: Own pre-treatment characteristics regressed on peer pre-treatment characteristics

Variable	Academic Composite			Leadership Composite			Fitness Score (CFT)		
Class/year	Freshmen Roommate	Freshmen Squadron	Sophomore Squadron	Freshmen Roommate	Freshmen Squadron	Sophomore Squadron	Freshmen Roommate	Freshmen Squadron	Sophomore Squadron
Specification	1	2	3	4	5	6	7	8	9
Class of 2000	NA	NA	-0.513 (-1.64)	NA	NA	-0.359 (-1.44)	NA	NA	-0.234 (-0.80)
Class of 2001	NA	NA	-0.293 (-1.65)	NA	NA	-0.011 (-0.06)	NA	NA	-0.424* (-1.75)
Class of 2002	-0.077 (-1.40)	-1.386** (-2.66)	-0.055 (-0.22)	-0.049 (-0.55)	-0.414 (-1.40)	-1.005** (-2.11)	-0.139* (-1.95)	-1.192** (-2.72)	-0.239 (-0.99)
Class of 2003	NA	NA	-1.303*** (-3.17)	NA	NA	-0.230 (-1.07)	NA	NA	-0.703* (-1.86)
Class of 2004	-0.042 (-0.75)	-1.184*** (-2.76)	-0.451 (-1.60)	-0.023 (-0.41)	-0.555 (-1.24)	-0.033 (-0.13)	-0.067 (-1.27)	-1.392*** (-2.82)	-0.094 (-0.42)
Class of 2005	-0.0892 (-1.30)	-0.187 (-0.53)	0.037 (0.15)	0.014 (0.24)	-0.574 (-1.50)	0.051 (0.27)	0.049 (0.70)	-0.110 (-0.44)	-0.002 (-0.01)
Class of 2006	-0.026 (-0.49)	0.124 (0.56)	-0.295 (-0.89)	0.067 (0.84)	0.038 (0.17)	-0.062 (-0.28)	0.002 (0.06)	-0.0004 (-0.00)	-0.432 (-1.12)
Class of 2007	-0.016 (-0.27)	-0.102 (-0.39)	-0.479 (-1.26)	-0.020 (-0.37)	0.094 (0.42)	-0.124 (-0.46)	0.138** (2.43)	-0.213 (-0.80)	-0.289 (-1.03)
Total Observations	3,475	5,785	8,427	3,475	5,786	8,429	3,472	5,783	8,426

Each coefficient represents a separate regression where the individual (pre-treatment) characteristic is regressed on the peer characteristic. No other controls are included in each regression. * Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level. t-statistic in parentheses, robust standard errors are clustered by class by squadron. For individuals with two roommates, the explanatory variables represent the average of the two roommates. For the squadron specifications, the explanatory variables are the average of all classmates in the squadron.

Table 3: Freshmen GPA on Roommate and Squadron Pre-treatment Characteristics

Variable	1	2	3	4
Roommate SAT Verbal	-0.0064 (-0.28)	0.0085 (0.38)	-0.0061 (-0.26)	-0.0023 (-0.10)
Roommate SAT Math	-0.0227 (-0.89)	-0.0189 (-0.69)	-0.0144 (-0.54)	-0.0175 (-0.65)
Roommate Academic Composite		0.0007 (0.12)	0.0018 (0.31)	0.0014 (0.25)
Roommate Fitness Score		0.0160 (1.11)	0.0124 (0.87)	0.0135 (0.95)
Roommate Leadership Composite		0.0126** (1.99)	0.0118* (1.94)	0.0127** (2.04)
Peer SAT Verbal (other freshmen in squadron)			0.4218** (2.47)	0.4816*** (3.13)
Peer SAT Math (other freshmen in squadron)			-0.0106 (-0.06)	-0.0146 (-0.08)
Peer Academic Composite (other freshmen in squadron)			-0.0328 (-0.69)	-0.0333 (-0.73)
Peer Fitness Score (other freshmen in squadron)			0.1153 (1.37)	0.1536* (1.89)
Peer Leadership Composite (other freshmen in squadron)			0.0166 (0.28)	0.0294 (0.50)
Sophomore Class SAT Verbal				0.2903** (2.10)
Sophomore Class SAT Math				0.1913 (1.21)
Sophomore Class Academic Composite				-0.0072 (-0.21)
Sophomore Class Fitness Score				-0.0602 (-0.66)
Sophmore Class Leadership Composite				-0.0748* (-1.66)
Junior Class SAT Verbal				-0.0241 (-0.16)

	(1)	(2)	(3)	(4)
Junior Class SAT Math				0.0908
				(0.51)
Junior Class Academic Composite				0.0109
				(0.27)
Junior Class Fitness Score				0.0967
				(0.99)
Junior Class Leadership Composite				0.1045***
				(2.72)
Senior Class SAT Verbal				0.0702
				(0.53)
Senior Class SAT Math				-0.0454
				(-0.24)
Senior Class Academic Composite				-0.0188
				(-0.47)
Senior Class Fitness Score				0.0674
				(0.67)
Senior Class Leadership Composite				-0.0448
				(-0.93)
SAT Verbal (own)	0.0994***	-0.0498**	-0.0485**	-0.0439**
	(4.96)	(2.45)	(2.42)	(2.14)
SAT Math (own)	0.2829***	0.0223	0.0236	0.0210
	(10.47)	(0.77)	(0.84)	(0.75)
Academic Composite (own)		0.1090***	0.1089***	0.1086***
		(20.86)	(20.74)	(21.30)
Fitness Score (own)		0.0449***	0.0462***	0.0483***
		(3.74)	(3.86)	(4.05)
Leadership Composite (own)		0.0012	0.0011	0.0021
		(0.18)	(0.15)	(0.31)
Observations	2,167	2,166	2,166	2,166
R^2	0.2035	0.3409	0.3447	0.3551
Control Variables	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female

* Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level. t-statistic in parentheses, robust standard errors are clustered by class by squadron.

Table 4: Freshmen Academic Outcomes on Roommate and Squadron Outcomes --
2SLS estimation

Variable	1	2	3	4
Freshmen Roommate GPA (maximum if two)	0.1186*** (3.29)	0.0449 (1.26)	0.0420 (1.17)	0.0377 (1.06)
Peer GPA (other freshmen in squadron)		0.6437*** (7.92)	0.6577*** (8.11)	0.7339*** (9.18)
Sophomore Class GPA (freshmen cumulative)			0.0522 (0.45)	-0.0820 (-0.61)
Junior Class GPA (freshmen cumulative)			0.2283*** (2.83)	0.1188 (1.16)
Senior Class GPA (freshmen cumulative)			0.0358 (0.34)	-0.0139 (-0.11)
Academic Composite (own)	0.1090*** (21.82)	0.1091*** (21.87)	0.1091*** (21.87)	0.1098*** (22.40)
Fitness Score (own)	0.0455*** (3.82)	0.0448*** (3.74)	0.0453*** (3.81)	0.0451*** (3.83)
Leadership Composite (own)	0.0002 (0.04)	-0.0008 (-0.11)	-0.0011 (-0.16)	-0.0007 (-0.10)
Observations	2,160	2,160	2,160	2,160
R^2	0.3389	0.3485	0.3496	0.3546
F-statistic on Roommate GPA first stage excluded instruments	31.84	32.58	29.51	33.63
F-statistic on Peer GPA first stage excluded instruments		6.40	6.41	12.88
Cragg-Donald weak identification statistic	19.52	18.90	18.84	18.52
Control Variables	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, Squadron FE

* Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the
0.01 level. z-statistic in parentheses, robust standard errors are clustered by class by
squadron. First stage excluded instruments include all roommate and squadron level
peer class, mentor class, supervisor class and leader class average pre-treatment and
demographic characteristics. All specifications fail to reject the Hansen-Sargan joint
null hypothesis that the instruments are uncorrelated with the error term, and that the
excluded instruments are correctly excluded from the estimated equation.

Table 5: Freshmen Physical Education Outcomes on Roommate and Squadron Outcomes -- 2SLS estimation

Variable	1	2	3	4
Freshmen Roommate PEA (maximum if two)	0.0974** (1.97)	0.0573 (1.12)	0.0563 (1.11)	0.0666 (1.32)
Peer PEA (other freshmen in squadron)		0.4164*** (4.45)	0.4275*** (4.71)	0.3940*** (5.04)
Sophomore Class PEA (freshmen cumulative)			0.0262 (0.34)	0.1046 (1.26)
Junior Class PEA (freshmen cumulative)			0.1446* (1.67)	0.1935** (2.26)
Senior Class PEA (freshmen cumulative)			0.0036 (0.04)	0.0274 (0.27)
Academic Composite (own)	0.0298*** (5.22)	0.0294*** (5.08)	0.0294*** (5.06)	0.0294*** (5.15)
Fitness Score (own)	0.2459*** (21.61)	0.2469*** (21.64)	0.2469*** (21.66)	0.2494*** (21.18)
Leadership Composite (own)	0.0236*** (3.78)	0.0233*** (3.75)	0.0232*** (3.74)	0.0251*** (4.08)
Observations	1,659	1,659	1,659	1,659
R^2	0.3269	0.3229	0.3232	0.3320
F-statistic on Roommate PEA first stage excluded instruments	17.27	17.27	18.74	31.25
F-statistic on Peer PEA first stage excluded instruments		6.91	10.63	21.27
Cragg-Donald weak identification statistic	11.15	10.49	10.49	10.48
Control Variables	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M

* Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level. z-statistic in parentheses, robust standard errors are clustered by class by squadron. First stage excluded instruments include all roommate and squadron level peer class, mentor class, and supervisor class pre-treatment and demographic characteristics. All specifications fail to reject the Hansen-Sargan joint null hypothesis that the instruments are uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation.

Table 6: Freshmen Peer Falsification Tests - 2SLS Results

Variable	1	2	3	4
Outcome	Academic		Athletic	
False Peer 1	0.0878		-0.0754	
(freshmen students in the same dorm section)	(0.60)		(-0.54)	
False Peer 2		-0.0255		0.0548
(all students in the same dorm section)		(-0.07)		(0.28)
Academic Composite	0.1104***	0.1104***	0.0286***	0.0285***
(own)	(24.69)	(24.55)	(6.71)	(6.67)
Fitness Score	0.0485***	0.0487***	0.2430***	0.2429***
(own)	(4.91)	(4.91)	(27.86)	(27.72)
Leadership Composite	0.0017	0.0019	0.0251***	0.0251***
(own)	(0.31)	(0.34)	(5.50)	(5.51)
Observations	3,367	3,367	2,846	2,846
R^2	0.3446	0.3428	0.3279	0.3276
Control Variables	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M

* Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level. z-statistic in parentheses, robust standard errors are clustered by class by squadron.

Table 7: Sophomore Academic Outcomes on Current and Freshmen Squadron Outcomes -- 2SLS estimation

Variable	1	2	3	4
Peer GPA	0.4742***	0.4372***	0.4440***	0.4502***
(other sophomores in current squadron)	(4.66)	(4.25)	(4.34)	(4.00)
Previous Peer GPA		0.2938***	0.2957***	0.2817***
(freshmen in previous squadron)		(3.87)	(3.88)	(3.46)
Previous Junior Class GPA			0.1608*	0.1616**
(junior's in previous squadron)			(1.95)	(1.96)
Junior Class GPA				-0.0377
(junior's in current squadron)				(-0.52)
Senior Class GPA				0.0618
(senior's in current squadron)				(0.92)
Academic Composite	0.0711***	0.0714***	0.0715***	0.0716***
(own)	(16.46)	(16.39)	(16.37)	(16.39)
Fitness Score	0.0245***	0.0233***	0.0235***	0.0236***
(own)	(2.84)	(2.67)	(2.69)	(2.71)
Leadership Composite	0.0038	0.0031	0.0028	0.0028
(own)	(0.85)	(0.67)	(0.62)	(0.62)
Observations	3,097	3,097	3,097	3,097
R^2	0.4488	0.4490	0.4495	0.4497
F-statistic on Peer GPA first stage excluded instruments	4.57	4.57	4.83	3.97
F-statistic on Previous Peer GPA first stage excluded instruments		109.77	113.01	108.29
Cragg-Donald weak identification statistic	15.95	15.88	15.88	16.79
Control Variables	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA

* Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level. z-statistic in parentheses, robust standard errors are clustered by class by squadron. First stage excluded instruments include all roommate and squadron level current and previous year peer class, mentor class, supervisor class, and leader class pre-treatment and demographic characteristics. All specifications fail to reject the Hansen-Sargan joint null hypothesis that the instruments are uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation.

Table 8: Sophomore Athletic Outcomes on Current and Freshmen Squadron Outcomes -- 2SLS estimation

Variable	1	2	3	4
Peer PEA	0.3815***	0.3689***	0.3687***	0.3936***
(other sophomores in current squadron)	(3.85)	(3.78)	(3.79)	(4.05)
Previous Peer PEA		0.2478**	0.2488**	0.2463**
(freshmen in previous squadron)		(2.40)	(2.37)	(2.35)
Previous Junior Class PEA			-0.0426	-0.0422
(junior's in previous squadron)			(-0.41)	(-0.40)
Junior Class PEA				-0.0324
(junior's in current squadron)				(-0.62)
Senior Class PEA				0.0164
(junior's in current squadron)				(0.29)
Academic Composite	0.0174***	0.0173***	0.0172***	0.0172***
(own)	(4.36)	(4.34)	(4.34)	(4.31)
Fitness Score	0.2029***	0.2027***	0.2027***	0.2027***
(own)	(23.43)	(23.70)	(23.69)	(23.66)
Leadership Composite	0.0204***	0.0200***	0.0200***	0.0200***
(own)	(4.41)	(4.34)	(4.33)	(4.34)
Observations	2,801	2,801	2,801	2,801
R^2	0.2352	0.2329	0.2330	0.2323
F-statistic on Peer PEA first stage excluded instruments	4.68	4.68	4.73	4.93
F-statistic on Previous Peer PEA first stage excluded instruments		113.87	128.78	126.97
Cragg-Donald weak identification statistic	19.10	19.08	19.02	19.09
Control Variables	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA

* Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level. z-statistic in parentheses, robust standard errors are clustered by class by squadron. First stage excluded instruments include all roommate and squadron level current and previous year peer class, mentor class, supervisor class, and leader class pre-treatment and demographic characteristics. All specifications fail to reject the Hansen-Sargan joint null hypothesis that the instruments are uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation.

Appendix A: LIML Estimation

Table A1: Freshmen Academic Outcomes on Roommate and Squadron Outcomes -- LIML estimation

Variable	1	2	3	4
Freshmen Roommate GPA (maximum if two)	0.1262*** (3.22)	0.0461 (1.22)	0.0429 (1.14)	0.0386 (1.04)
Peer GPA (other freshmen in squadron)		0.6520*** (7.86)	0.6658*** (8.07)	0.7482*** (9.27)
Sophomore Class GPA (freshmen cumulative)			0.0532 (0.45)	-0.0808 (-0.60)
Junior Class GPA (freshmen cumulative)			0.2282*** (2.85)	0.1205 (1.17)
Senior Class GPA (freshmen cumulative)			0.0361 (0.35)	-0.0106 (-0.08)
Academic Composite (own)	0.1090*** (21.85)	0.1091*** (21.88)	0.1091*** (21.87)	0.1098*** (22.39)
Fitness Score (own)	0.0455*** (3.82)	0.0448*** (3.74)	0.0453*** (3.81)	0.0451*** (3.83)
Leadership Composite (own)	0.0002 (0.04)	-0.0008 (-0.12)	-0.0011 (-0.17)	-0.0007 (-0.11)
Observations	2,160	2,160	2,160	2,160
R^2	0.3382	0.3483	0.3494	0.3540
F-statistic on Roommate GPA first stage excluded instruments	31.84	32.58	29.51	33.63
F-statistic on Peer GPA first stage excluded instruments		6.40	6.41	12.88
Cragg-Donald weak identification statistic	19.52	18.90	18.84	18.52
Control Variables	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, Squadron FE

* Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level. z-statistic in parentheses, robust standard errors are clustered by class by squadron. First stage excluded instruments include all roommate and squadron level peer class, mentor class, supervisor class and leader class average pre-treatment and demographic characteristics. All specifications fail to reject the Hansen-Sargan joint null hypothesis that the instruments are uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation.

Table A2: Freshmen Physical Education Outcomes on Roommate and Squadron Outcomes -- LIML estimation

Variable	1	2	3	4
Freshmen Roommate PEA (maximum if two)	0.1077** (2.03)	0.0621 (1.15)	0.0610 (1.13)	0.0722 (1.36)
Peer PEA (other freshmen in squadron)		0.4280*** (4.52)	0.4393*** (4.78)	0.4123*** (5.22)
Sophomore Class PEA (freshmen cumulative)			0.0256 (0.33)	0.0998 (1.21)
Junior Class PEA (freshmen cumulative)			0.1417* (1.65)	0.1875** (2.20)
Senior Class PEA (freshmen cumulative)			0.0018 (0.02)	0.0238 (0.23)
Academic Composite (own)	0.0297*** (5.21)	0.0293*** (5.06)	0.0294*** (5.05)	0.0294*** (5.14)
Fitness Score (own)	0.2456*** (21.42)	0.2468*** (21.54)	0.2467*** (21.55)	0.2493*** (21.08)
Leadership Composite (own)	0.0235*** (3.77)	0.0232*** (3.74)	0.0232*** (3.73)	0.0250*** (4.06)
Observations	1,659	1,659	1,659	1,659
R^2	0.3252	0.3219	0.3221	0.3303
F-statistic on Roommate PEA first stage excluded instruments	17.27	17.27	18.74	31.25
F-statistic on Peer PEA first stage excluded instruments		6.91	10.63	21.27
Cragg-Donald weak identification statistic	11.15	10.49	10.49	10.48
Control Variables	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M

* Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level. z-statistic in parentheses, robust standard errors are clustered by class by squadron. First stage excluded instruments include all roommate and squadron level peer class, mentor class, and supervisor class pre-treatment and demographic characteristics. All specifications fail to reject the Hansen-Sargan joint null hypothesis that the instruments are uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation.

Table A3: Sophomore Academic Outcomes on Current and Freshmen Squadron Outcomes -- LIML estimation

Variable	1	2	3	4
Peer GPA (other sophomores in current squadron)	0.5200*** (4.93)	0.4710*** (4.44)	0.4772*** (4.53)	0.4955*** (4.90)
Previous Peer GPA (freshmen in previous squadron)		0.2991*** (3.87)	0.3009*** (3.88)	0.2856*** (3.44)
Previous Junior Class GPA (junior's in previous squadron)			0.1608* (1.95)	0.1616** (1.96)
Junior Class GPA (junior's in current squadron)				-0.0313 (-0.45)
Senior Class GPA (senior's in current squadron)				0.0661 (1.02)
Academic Composite (own)	0.0711*** (16.46)	0.0715*** (16.39)	0.0716*** (16.37)	0.0716*** (16.41)
Fitness Score (own)	0.0245*** (2.84)	0.0233*** (2.67)	0.0236*** (2.69)	0.0237*** (2.70)
Leadership Composite (own)	0.0039 (0.85)	0.0031 (0.67)	0.0028 (0.62)	0.0028 (0.62)
Observations	3,097	3,097	3,097	3,097
R^2	0.4477	0.4481	0.4486	0.4485
F-statistic on Peer GPA first stage excluded instruments	4.57	4.57	4.83	3.97
F-statistic on Previous Peer GPA first stage excluded instruments		109.77	113.01	108.29
Cragg-Donald weak identification statistic	15.95	15.88	15.88	16.79
Control Variables	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA

* Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level. z-statistic in parentheses, robust standard errors are clustered by class by squadron. First stage excluded instruments include all roommate and squadron level current and previous year peer class, mentor class, supervisor class, and leader class pre-treatment and demographic characteristics. All specifications fail to reject the Hansen-Sargan joint null hypothesis that the instruments are uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation.

Table A4: Sophomore Athletic Outcomes on Current and Freshmen Squadron Outcomes -- LIML estimation

Variable	1	2	3	4
Peer PEA	0.4047***	0.3879***	0.3880***	0.4151***
(other sophomores in current squadron)	(4.04)	(3.93)	(3.95)	(4.24)
Previous Peer PEA		0.2529**	0.2541**	0.2517**
(freshmen in previous squadron)		(2.42)	(2.40)	(2.37)
Previous Junior Class PEA			-0.0446	-0.0443
(junior's in previous squadron)			(-0.43)	(-0.42)
Junior Class PEA				-0.0295
(junior's in current squadron)				(-0.58)
Senior Class PEA				0.0172
(junior's in current squadron)				(0.31)
Academic Composite	0.0174***	0.0173***	0.0172***	0.0172***
(own)	(4.35)	(4.33)	(4.33)	(4.31)
Fitness Score	0.2029***	0.2027***	0.2027***	0.2027***
(own)	(23.40)	(23.69)	(23.67)	(23.65)
Leadership Composite	0.0204***	0.0200***	0.0200***	0.0200***
(own)	(4.41)	(4.34)	(4.33)	(4.33)
Observations	2,801	2,801	2,801	2,801
R^2	0.2344	0.2322	0.2322	0.2314
F-statistic on Peer PEA first stage excluded instruments	4.68	4.68	4.73	4.93
F-statistic on Previous Peer PEA first stage excluded instruments		113.87	128.78	126.97
Cragg-Donald weak identification statistic	19.10	19.08	19.02	19.09
Control Variables	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA	graduation class, recruited athlete, black, Hispanic, Asian, preparatory school, female, SAT-V, SAT-M, >3.499 freshmen GPA & >3.14 freshmen MPA

* Significant at the 0.10 level, ** Significant at the 0.05 level, *** Significant at the 0.01 level. z-statistic in parentheses, robust standard errors are clustered by class by squadron. First stage excluded instruments include all roommate and squadron level current and previous year peer class, mentor class, supervisor class, and leader class pre-treatment and demographic characteristics. All specifications fail to reject the Hansen-Sargan joint null hypothesis that the instruments are uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation.