The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point

Author(s): David S. Lyle
Source: *American Economic Journal: Applied Economics*, Vol. 1, No. 4 (October 2009), pp. 69-84
Published by: American Economic Association
Stable URL: https://www.jstor.org/stable/25760182
Accessed: 28-01-2019 13:45 UTC

American Economic Journal: Applied Economics 2009, 1:4, 69–84
http://www.aeaweb.org/articles.php?doi=10.1257/app.1.4.69

The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point[†]

By David S. Lyle*

Understanding how heterogeneity in peer group composition affects academic attainment has important implications for how schools organize students in group settings. The random assignment of cadets to companies at West Point affords an opportunity to investigate this issue empirically. Estimates of the impact of peer group heterogeneity in math SAT scores on freshmen-year academic performance reveals that more heterogeneous peer groups have positive effects on individual grades. High-ability peers account for most of the positive effect, while low-ability peers have no measureable effect. (*JEL* I23, J24, M54)

Human capital acquisition often occurs in group settings such as classrooms, social organizations, athletic teams, or even gangs. Understanding how peers contribute to or detract from learning has drawn considerable attention from educators, parents, sociologists, and economists alike. One unresolved issue with first-order economic implications for how educational institutions organize students in group settings is whether peer group heterogeneity has positive or negative effects on an individual's academic achievement.

Unlike linear-in-means peer effects, where potential gains from increasing the average in one group are offset by decreasing the average in another, peer group heterogeneity effects have the potential to increase education production in aggregate. For example, academic institutions will find it optimal from a production efficiency perspective to segregate students by ability when the loss to high-ability students is greater than the gain to low-ability students. On the other hand, if the loss to high-ability students from interacting with low-ability students is less than the gain that the low-ability students receive from interacting with the high-ability student, schools would find it optimal to mix students by ability.

To date, there is little empirical evidence on how peer group heterogeneity affects individual academic outcomes at the undergraduate level. The fact that we observe higher performing students attending Ivy League colleges and lower performing

* Lyle: Department of Social Sciences, West Point, 607 Cullum Road, West Point, NY 10996 (e-mail: david.lyle@usma.edu). I am grateful to Daron Acemoglu, Josh Angrist, David Autor, Scott Carrell, David L. Dudley, Dean Dudley, Christian Hansen, Larry Katz, Michael Meese, Bruce Sacerdote, John Smith, Casey Wardynski, two anonymous referees, and seminar participants at the Massachusetts Institute of Technology, Syracuse University, and West Point for their valuable comments and suggestions. The views expressed herein are those of the author and do not purport to reflect the position of the US Military Academy, the Department of the Army, or the Department of Defense.

[†] To comment on this article in the online discussion forum, or to view additional materials, visit the articles page at: http://www.aeaweb.org/articles.php?doi=10.1257/app.1.4.69.

students attending community colleges is consistent with the case for segregating students by ability. However, there are at least two other reasons why we may observe segregation in this setting. First, credit market imperfections may prevent lower performing students, who tend to come from disadvantaged socioeconomic backgrounds, from obtaining a more expensive Ivy League education. Second, colleges may impose screening criteria to promote prestige, develop networks, or improve the quality of the labor market signal for their graduates.

The lack of empirical evidence on this subject is mainly a result of several well-documented econometric challenges that confront the identification of peer effects. Charles F. Manski (1993), Bruce Sacerdote (2001), David J. Zimmerman (2003), and David S. Lyle (2007) are a few of the many studies in this growing literature that provide a rich exposition on the selection, endogeneity, and common shock issues that confound the interpretation of estimated peer effects, particularly as it pertains to undergraduate education.

To briefly summarize, selection bias is often present because individuals typically choose their peer group. This makes it difficult to determine the appropriate members of the peer group and also to separate the true peer effect from the selection effect. Endogeneity problems are present when estimating contemporaneous peer effects because an individual and the members of his or her peer group impact each other concurrently. And common shocks, such as a teacher or a classroom configuration, are common treatments that affect the outcomes of all members of the social group. In practice, there are few instances in which it is possible to adequately deal with the identification challenges associated with estimating peer effects, and there are even fewer instances in which it is possible to study the issue of how peer group heterogeneity affects education production.

However, the random assignment of cadets to companies at West Point provides such a rare opportunity. Clearly definable peer groups, random assignment, and reliable measures of pretreatment ability combine to mitigate the selection, endogeneity, and common shock concerns. Each year, the Academy assigns incoming cadets to one of 36 companies. Companies are composed of seniors, juniors, sophomores, and freshmen. The focus of this study is on freshmen, or using West Point terminology, *plebes*. Thus, the term *peer* in this paper refers to the 34 other plebes assigned to an individual plebe's company.

Plebes are randomly assigned to companies, conditional on several observable characteristics: gender, race, recruited athlete, and measures of prior performance and behavior. West Point attempts to equalize company averages across these characteristics, which results in relatively uniform peer group means. However, other moments of the peer group distribution, measures of dispersion, for example, vary considerably. Drawing from Roland Benabou (1996), which links dispersion in peer group distributions to the degree of substitutability of peer ability with individual ability, this paper presents a general education production framework that includes both the first and second moments in the peer distributions.[1] Reinforcing this approach, recent papers by Caroline M. Hoxby and Gretchen Weingarth (2006);

[1] See the Web Appendix for a description of the Benabou model as it relates to the research design in this paper. I provide estimates of a structural model that suggest peer ability is a substitute to own ability.

Gigi Foster (2006); Ralph Stinebrickner and Todd R. Stinebrickner (2006); and Esther Duflo, Pascaline Dupas, and Michael Kremer (2008) make the case for moving beyond linear-in-means estimation to consider how other dimensions of a peer group distribution might affect an individual.

In this study, estimates of several measures of dispersion in the distributions of plebe math SAT scores indicate that more heterogeneous peer groups have positive effects on a plebe's grades. A one-standard deviation increase in the peer group 75–25 differential in math SAT scores increases a company's average math grade by 13 percent of a standard deviation. This finding is robust across different measures of dispersion. In other specifications, estimates indicate that the seventy-fifth percentile, and not the twenty-fifth percentile, in peer math SAT scores accounts for most of the effect.

In the next section, I provide background information on West Point. Section II describes the data, and Section III explains the random assignment of cadets to companies. In Section IV, I present the empirical model and formally discuss the identification assumptions and interpretations. Sections V and VI contain the main results, and Section VII concludes.

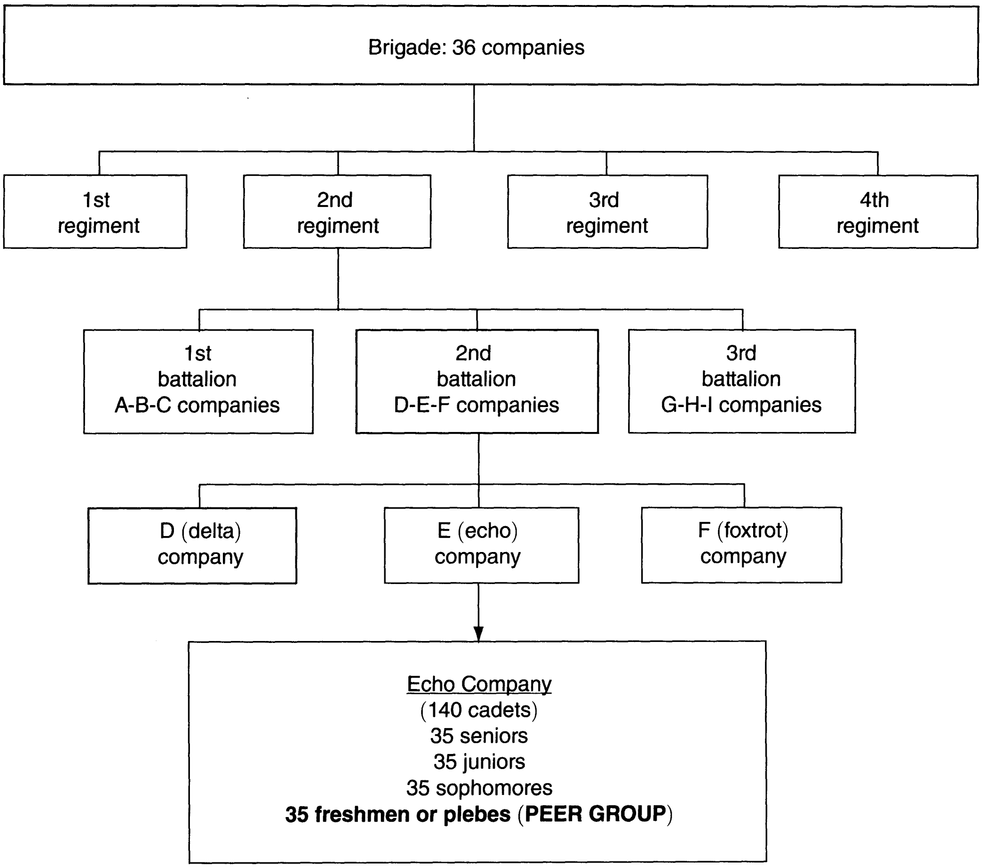
I. West Point

West Point is one of three service academies fully funded by the US Government for the expressed purpose of providing the nation with leaders of character who serve the common defense. Cadets offered admission to the Academy receive a fully funded four-year scholarship. Graduates earn an accredited Bachelor of Science degree and must fulfill a five-year active duty service obligation as an officer in the US Army.

The Corps of Cadets at West Point is organized into one brigade consisting of 36 companies, as seen in Figure 1. The brigade has four regiments, each regiment has three battalions, and each battalion has three companies. Every company has approximately 140 cadets, 35 from each of the four classes. West Point assigns plebes to a company during the summer prior to the start of plebe-year. Plebes remain in the same initially assigned company throughout their plebe-year. For the purposes of this study, the 34 other plebes in a company comprise an individual plebe's peer group.

Companies serve as the dominant social organization. Plebes within the same company eat together, study together, attend mandatory social activities together, perform their military duties together, and live in a section of the barracks together. The hierarchical structure of a company at West Point is similar to a company in an active duty Army unit and is designed to provide leadership training to upperclassmen and to promote teamwork among plebes. Plebes perform many routine duties under the close scrutiny of the upperclassmen such as delivering newspapers and mail, doing laundry, notifying all upperclassmen of formations, keeping the company area in immaculate condition, serving meals to upperclassmen, and memorizing institutional knowledge. The nature of the duties assigned to plebes and the organization of a company forces plebes within the same company to cooperate in order to accomplish their many requirements.

All plebes take the same courses throughout the first year at West Point. English, calculus, history, computer science, behavioral psychology, and chemistry comprise



Term	Description
Company	Made up of all four classes: 35 cadets in each class
Cadet	General term referring to an individual from any of the four classes
Plebe	Freshman cadet
Peer group	A plebe's peer group is comprised of the other 34 plebes in the company

Figure 1. Organization of West Point's Corps of Cadets

the academic grade point average (GPA) during plebe-year. Plebes do not necessarily take classes with other plebes from their company, however, all plebes receive the same program of instruction, complete the same homework assignments, and take the same exams. Since the company is the dominant organization, nearly all homework assignments and exam preparations are conducted between plebes within the same company.

II. Data Description

The data for this study are from the Office of Economic and Manpower Analysis (OEMA), West Point, NY. The data contain admissions files and personnel records

TABLE 1—COMPANY-LEVEL AND INDIVIDUAL-LEVEL SUMMARY STATISTICS

Panel A. Plebe-year academic attainment	Observations	Mean	SD	Minimum	Maximum
Math grade plebe year (individual)	6,309	2.70	0.80	0.50	4.30
Academic GPA plebe year (individual)	6,870	2.67	0.54	0.19	4.10
Math grade plebe year (company)	252	2.70	0.23	2.10	3.15
Academic GPA plebe year (company)	252	2.67	0.12	2.33	2.97
Panel B. Math SAT scores (pretreatment characteristics)	Companies	Mean	SD	Minimum	Maximum
Math SAT score	252	636.7	10.6	599.0	661.8
Variance of math SAT score	252	4,448.8	1,122.3	1,989.5	8,725.2
75–25 math SAT score differential	252	94.0	21.2	40.0	180.0
Seventy-fifth percentile of math SAT score	252	685.0	15.4	640.0	730.0
Twenty-fifth percentile of math SAT score	252	591.0	15.8	550.0	620.0
Panel C. Random assignment scrambling controls	Companies	Mean	SD	Minimum	Maximum
Female	252	0.118	0.029	0.038	0.214
Black	252	0.065	0.033	0.000	0.179
Hispanic	252	0.045	0.029	0.000	0.154
Recruited football players	252	0.076	0.035	0.000	0.188
Other recruited athletes	252	0.146	0.047	0.000	0.344
Attended the West Point Prep School	252	0.156	0.034	0.063	0.259
College Entrance Exam Rank (CEER)	252	608.0	6.0	586.3	625.9
Whole Candidate Score (WCS)	252	6,032.7	38.0	5,911.9	6,167.4

Notes: Data are from the Office of Economic & Manpower Analysis, West Point, NY. Data include personnel, admissions, and performance files for the graduating classes of 1992–1998. Differences in individual-level math grade and individual-level GPA sample size are due to missing math grades from the West Point database (a result of changing data management systems). The 252 company-level observations come from the 36 companies across 7 years. The CEER score is a weighted average of SAT, ACT, and high school rank. WCS is an aggregated score from pretreatment activities and performance.

for cadets in the graduating classes of 1992–1998 and include approximately 7,000 plebes as part of 252 companies (36 companies across 7 years). I organize the data in Table 1 into three categories: academic attainment at West Point, pretreatment measures of academic ability, and randomization controls.

In panel A, I present individual-level and company-level summary statistics for the two outcome variables used in this study: plebe-year math grade and academic GPA. Grades range from 0 to 4.3 points. For example, a 4.3 equates to an A+, a 4.0 equates to an A, and a 3.7 equates to an A−. Plebe math grades have a mean of 2.70 points (C+) and a standard deviation of 0.23 points across companies. The plebe-year academic GPA has a similar mean of 2.67 points (C+), but a smaller standard deviation of 0.12 points across companies.

Panel B contains summary statistics for measures of the company-level math SAT distribution.[2] The math SAT scores represent a measure of quantitative ability obtained prior to entering West Point and correspond naturally to math grades

[2] All SAT scores were taken prior to the 1995 renormalization.

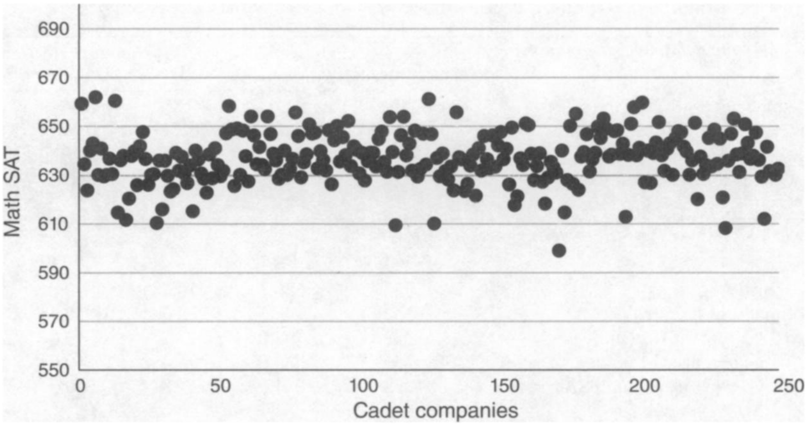


FIGURE 2. COMPANY AVERAGE MATH SAT

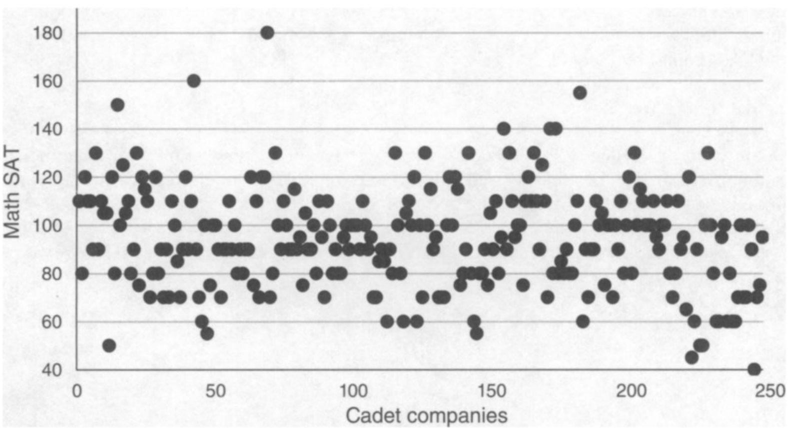


FIGURE 3. COMPANY 75–25 DIFFERENTIAL MATH SAT

Note: Axes span identical ranges (140 SAT points) so that both figures are visually comparable.

and academic performance measured by the overall plebe-year GPA. The average math SAT score is approximately 640 points with a standard deviation of about 10 points.[3] I also provide summary statistics for two measures of dispersion in the company math SAT distributions: the variance and the 75–25 differential. Despite comparatively equal company-level average SAT scores, as shown in Figure 2, some companies have a high concentration of cadets in the tails and other companies have a high concentration of cadets in the middle of the math SAT distributions. The sizeable variation in measures of dispersion, as shown in Figure 3, relative to the minimal variation in means across companies (Figure 2), is an important feature

[3] I decompose the total variation in math SAT scores into the within and between company components. As expected with the random assignment process, the within variation explained 98 percent of the total variation.

of the experimental design in this paper. I provide more discussion on this issue in Section V.

Company-level statistics for the randomization controls used for assigning cadets to companies are in panel C of Table 1. Approximately 12 percent of the Corps of Cadets is made up of females. Blacks and Hispanics combine to account for about 11 percent of each class. A little more than 22 percent of incoming cadets are recruited for one of the 20 NCAA Division-One athletic programs at the Academy. Also, 16 percent attended the United States Military Academy Prep School the year before entering West Point. The College Entrance Exam Rank (CEER) is a weighted average between the high school graduation ranking of the cadet and the SAT/ACT scores. The range of this ranking is from 0 to 800 points, with a mean of approximately 610 points. The Whole Candidate Score (WCS) aggregates assigned values to various high school activities and pretreatment performance outcomes. For example, playing varsity high school basketball or being a member of a high school student council contributes points to the WCS. The WCS ranges from 0 to 8,000 points with a mean of about 6,000 points.

III. Social Groups and Random Assignment

The critical identification assumption for this experiment is that the assignment of cadets to companies at West Point is random, conditional on the individual level controls listed in panel C of Table 1. The following description of the assignment process and some brief empirical analysis support this assumption.

West Point assigns a random number to each incoming plebe in a process known at the Academy as *scrambling*.[4] The goal of scrambling is to produce companies with comparable average characteristics. West Point initially assigns plebes to one of the companies based on their random number, and then shuffles plebes between companies in an attempt to further equalize the means of the eight characteristics. All subsequent reassignments of plebes are a function of the scrambling controls and the random number.

To assess the extent to which scrambling differs from the initial random assignment, I provide a comparison of math SAT scores that resulted from the actual scrambling process with those that I construct from a purely random assignment process. Table 2 contains descriptive statistics from the actual scrambling process (the same as in panel B of Table 1), and descriptive statistics from a purely random assignment process.[5] In general, the means and standard deviations are highly comparable across the company average and all measures of dispersion in company-level math SAT scores. Given that the scrambling process attempts to further equalize the company averages, we should expect to see a slightly larger standard deviation of

[4] USMA publication 98-007, "Evaluation of Scrambling in the Corps of Cadets 1962–1998" provides a detailed description of the scrambling process. Discussions with managers in charge of scrambling from the Institutional Research and Analysis, Office of Policy Planning and Analysis, West Point, NY, also confirm my description of the process.

[5] See the notes in Table 2 for a detailed description of the scrambling and purely random assignment processes.

TABLE 2—COMPARISON OF PURELY RANDOM AND SCRAMBLED COMPANY-LEVEL MATH SAT SCORES

		Scrambling		Purely random	
Company level statistics	Companies	Mean	SD	Mean	SD
Average math SAT score	252	636.7	10.6	636.6	11.6
Variance of math SAT score	252	4,448.8	1,122.3	4,434.6	1,097.4
75–25 math SAT score differential	252	94.0	21.2	91.6	20.1
75th percentile of the math SAT score	252	685.0	15.4	683.6	16.5
25th percentile of the math SAT score	252	591.0	15.8	592.0	17.9

Notes: The 252 company-level observations come from the 36 companies across 7 years. The slight difference in average SAT scores between the scrambling and purely random assignment process are a result of differences in company populations. The scrambling process at West Point has two stages. In the first stage, West Point conducts ten purely random assignment draws and selects the draw that provides the most uniform company means with regard to the scrambling controls: gender, race, athlete, prep school, CEER, and WCS. In the second stage, West Point moves a few cadets around by hand to further balance company means along these dimensions. The scrambling columns contain descriptive statistics from the final assignments. In the purely random assignment columns, I provide descriptive statistics from a replication of stage one in which I conduct ten random-assignment draws and display the draw with the most uniform means. However, all ten draws produce nearly identical descriptive statistics. See Table 1 notes for further description.

TABLE 3—THE RANDOM ASSIGNMENT OF PEERS AT WEST POINT
OUTCOME VARIABLE: PEER GROUP AVERAGE MATH SAT

	(1)	(2)
Math SAT	−0.011	0.004
	(0.002)	(0.004)
R^2	0.06	0.07
Observations	6,870	6,870
Scrambling controls	No	Yes

Notes: Standard errors in parentheses account for clustering at the company and year level. OLS estimates reflect regressions of average peer Math SAT on individual Math SAT. All specifications include year dummies and a constant. Individual-level random assignment scrambling controls (gender, race, recruited athlete, prep school, CEER, and WCS) are included as indicated. See Table 1 notes for sample description.

the company-level math SAT score for the random process, 11.6, as compared to the scrambling process, 10.6.

Estimates in Table 3 further support my characterization of the assignment process. I regress peer average math SAT scores on corresponding individual level math SAT scores to determine if a plebe's background predicts the background of his peer group. The peer average math SAT score is the average math SAT score of the plebes in a company minus the individual plebe. Estimates in column 1 are from a regression of average peer math SAT score on individual math SAT score. There is a small negative correlation, as would be expected, given the equalizing intent of the scrambling process.[6] When I include the scrambling controls in column 2, the

[6] Since the peer average math SAT is constructed without the own math SAT score, plebes with higher SAT scores will likely be assigned to companies with lower average peer math SAT scores, and vice versa. As discussed in Jonathan Guryan, Kory Kroft, and Matt Notowidigdo (2007), the size of the negative bias decreases as the population size from which peers are drawn increases. In this study, the peer population is approximately 1,400 cadets with peer group sizes of approximately 35 cadets. To mitigate any residual negative bias, I also include peer group means as controls.

point estimate is smaller in absolute value by an order of magnitude and is no longer statistically significant. Therefore, to account for the conditional randomization process, I include individual-level controls for the eight scrambling variables in all specifications.

IV. Empirical Framework

Drawing from Benabou (1996) and the existing literature on peer effects, I estimate variations of a model that prescribes individual academic outcomes as a function of own ability, average peer ability, and measures of dispersion in peer ability.

$$(1) \quad Y_{ict} = \kappa + \theta_t + \lambda\, Z_{ict-1} + \delta\, \overline{Z}_{pt-1} + \gamma\, \hat{Z}_{pt-1} + \tau\, X_{ict-1} + \eta\, \overline{X}_{pt-1} + \varepsilon_{ict}.$$

The left-hand-side variable Y_{ict} is the academic outcome of interest (math grade or GPA) for plebe i, in company c, in year t (plebe-year). On the right-hand side of the equation, κ is a constant, θ_t are year dummies for 1993–1998, and λ denotes the effect of own pretreatment $(t-1)$ measures of academic ability (math SAT). δ represents the effect of average math SAT score for the peer group (p), where $\overline{Z}_{pt-1}$ contains the average peer math SAT score. γ represents the effect of varying measures of the peer group's distribution of math SAT scores, where $\hat{Z}_{pt-1}$ contains either the variance, the 75–25 differential, or the seventy-fifth and twenty-fifth percentiles of the peer group distribution.[7] τ and η denote estimates of the scrambling controls at the individual (X_{ict-1}) and peer level $(\overline{X}_{pt-1})$, respectively; and, ε_{ict} represents other potential determinants of individual-level academic attainment.

In most settings, estimates for coefficients of interest δ and γ would be biased by selection, common shocks, and endogeneity due to correlations between $\overline{Z}_{pt-1}$ and ε_{ict}, and between $\hat{Z}_{pt-1}$ and ε_{ict}. However, the random assignment of cadets to companies, and the use of only pretreatment characteristics for all right-hand-side variables, suggest $E[\overline{Z}_{pt-1}\varepsilon_{ict}] = E[\hat{Z}_{pt-1}\varepsilon_{ict}] = 0$. Random assignment is likely to negate the selection component of ε_{ict}, and it is unlikely that any shocks to pretreatment characteristics are common to members of the newly assigned social group. Finally, using only pretreatment measures of academic ability mitigates the endogeneity problem by exploiting the timing structure on the peer effect of interest. Y_{ict} cannot influence $\overline{Z}_{pt-1}$ or $\hat{Z}_{pt-1}$.

As explained above, exogenous variation in measures of peer-group dispersion that results from random assignment is central to my identification strategy. However, the nature of random assignment also makes having an appropriately sized peer group important. For example, randomly assigning individuals to small peer groups comprised of only a few plebes would produce measures of dispersion that

[7] Note p equals all plebes in company (c) minus plebe (i).

TABLE 4—LINEAR-IN-MEANS SPECIFICATIONS
(*Outcome variable: individual plebe math grade or plebe GPA*)

	Math grades (1)	GPA (2)
Panel A. Mean		
Peer group mean math SAT/100	−0.029	0.070
	(0.088)	(0.052)
R^2	0.29	0.43
Observations	6,309	6,870
Panel B. Mean interactions		
Peer group mean math SAT × top 25 percent/100	−0.028	0.072
	(0.092)	(0.053)
Peer group mean math SAT × middle 50 percent/100	−0.029	0.073
	(0.094)	(0.054)
Peer group mean math SAT × bottom 25 percent/100	−0.028	0.072
	(0.095)	(0.055)
R^2	0.29	0.43
Observations	6,309	6,870

Notes: Standard errors in parentheses account for clustering at the company and year level. OLS estimates in panel A reflect regressions of individual academic outcomes as listed in the column headings on the peer mean math SAT score. Panel B interacts the peer mean math SAT with dummies for if a cadet falls in the bottom 25 percent, middle 50 percent, or top 25 percent of the math SAT distribution for each company. All specifications also include year dummies, a constant, and individual-level random scrambling controls: gender, race, recruited athlete, prep school, CEER, and WCS. See Table 1 notes for sample description.

are highly susceptible to outliers. At the other extreme, randomly assigning individuals to large peer groups would produce measures of dispersion that are highly comparable, which reduces the ability to identify peer effects. In this paper, peer sizes range from 30 to 34 plebes, making for a large enough distribution to have well-defined measures of dispersion, but small enough to provide enough power to identify reasonably sized peer effects.

In general, the design of this experiment addresses the main obstacles confronting the identification of peer effects. Therefore, ordinary least squares (OLS) estimates of δ and γ from models of the form in equation (1) are apt to provide credible evidence for the degree to which average peer ability and peer group heterogeneity affect educational attainment in this setting. Since the key right-hand-side variables, $\overline{Z}_{pt-1}$ and $\hat{Z}_{pt-1}$, vary by peer group, all standard errors are clustered at the company by year level using Huber-White robust standard errors.

V. Linear-in-Means and Peer Group Heterogeneity

I begin this portion of the analysis by estimating a standard, linear-in-means peer effect specification. Table 4 contains estimates from a specification as in equation (1), where Y_{ict} is the plebe math grade or GPA; Z_{it-1} as a cubic in own math SAT score; $\overline{Z}_{pt-1}$ is the average peer math SAT score; and I omit the measure of dispersion, $\hat{Z}_{pt-1}$. Estimates in the math grades columns show that average peer group math SAT scores have no statistically significant effect on plebe math grades or plebe GPA. In the GPA columns, I interact the average peer group math SAT score with dummy

TABLE 5—ESTIMATES OF PEER GROUP VARIANCE
(*Outcome variable: individual plebe math grade or plebe GPA*)

	Math grades		GPA	
	(1)	(2)	(1)	(2)
Peer group variance in math SAT/10,000	0.159	0.168	0.081	0.088
	(0.075)	(0.082)	(0.047)	(0.048)
Peer group mean math SAT/100	−0.002	−0.003	0.083	0.097
	(0.087)	(0.096)	(0.053)	(0.056)
R^2	0.29	0.29	0.43	0.43
Observations	6,309		6,870	
Average peer group scrambling controls	No	Yes	No	Yes

Notes: Standard errors in parentheses account for clustering at the company and year level. OLS estimates reflect regressions of individual academic outcomes on the designated measure of dispersion in peer math SAT scores. All specifications include year dummies, a constant, a cubic in own math SAT, average peer math SAT, and individual-level random scrambling controls: gender, race, recruited athlete, prep school, CEER, and WCS. Average peer group scrambling controls are added as indicated. See Table 1 notes for sample description.

variables that indicate whether a plebe belongs to the top 25 percent, middle 50 percent, or bottom 25 percent of the math SAT distribution in each company. These estimates reveal a relatively uniform and insignificant mean effect across the distribution of math SAT scores.

This small and insignificant mean effect is consistent with findings in other studies on peer effects at the undergraduate level. Using pretreatment measures of academic ability, Sacerdote (2001) finds no statistically significant effects for roommates at Dartmouth College. Zimmerman (2003) finds small, positive, and significant effects in only one of three measures of academic ability that he tests in his study on roommates at Williams College. Foster (2006) finds little evidence of peer effects on undergraduate performance at the University of Maryland, and Stinebrickner and Stinebrickner (2006) find no effect on freshmen at Berea College.

There are several possible interpretations of the zero mean effect in this paper.[8] Within the context of this particular experiment, and as shown in Figure 2, it could be that peer group means do not have enough variation to identify an effect due to the equalizing intent of the random scrambling process. However, consistent with the comparable literature, it could also be that the average math SAT score for a peer group has no effect on a plebe's academic performance. Hoxby and Weingarth (2006), Stinebrickner and Stinebrickner (2006), and Foster (2006) argue that the scant evidence of peer effects found in the literature from linear-in-means specifications suggest looking beyond "average peer effects."

Accordingly, I turn to the full specification in equation (1). The first measure of dispersion, $\hat{Z}_{pt-1}$, that I test is the variance of the peer group distributions. Table 5 contains estimates of γ_{var} for plebe math grades and GPA, respectively. For both math and GPA, column 1 contains the main specification. There is a positive and significant effect of the variance, but, again, no significant effect of the mean. A one

[8] The linear-in-means estimates in this paper are identical to the estimates in Lyle (2006).

TABLE 6—ESTIMATES OF THE 75–25 DIFFERENTIAL AND THE 75TH AND 25TH PERCENTILE
(*Outcome variable: individual plebe math grade or plebe GPA*)

	75–25 differential of the peer math SAT distribution			
	Math grades		GPA	
	(1)	(2)	(1)	(2)
Panel A				
Peer group 75–25 math SAT differential/100	0.142	0.142	0.081	0.080
	(0.040)	(0.042)	(0.024)	(0.024)
Peer group mean math SAT/100	−0.031	−0.031	0.068	0.083
	(0.087)	(0.095)	(0.052)	(0.055)
R^2	0.29	0.29	0.43	0.43
Observations	6,309		6,870	
Average peer group scrambling controls	No	Yes	No	Yes

	75th and 25th percentiles of the peer math SAT distribution			
	Math grades		GPA	
	(1)	(2)	(1)	(2)
Panel B				
Peer group seventy-fifth math SAT percentile/100	0.270	0.272	0.184	0.189
	(0.088)	(0.089)	(0.055)	(0.054)
Peer group twenty-fifth math SAT percentile/100	−0.030	−0.025	0.008	0.018
	(0.086)	(0.093)	(0.051)	(0.054)
Peer group mean math SAT/100	−0.261	−0.277	−0.116	−0.121
	(0.181)	(0.192)	(0.112)	(0.117)
R^2	0.29	0.29	0.43	0.43
Observations	6,309		6,870	
Average peer group scrambling controls	No	Yes	No	Yes

Notes: Standard errors in parentheses account for clustering at the company and year level. OLS estimates in panel A columns reflect regressions of individual academic outcomes on the 75–25 differential of the peer math SAT distribution. Estimates in panel B reflect regressions of individual academic outcomes on the twenty-fifth and seventy-fifth percentile of the peer math SAT distribution. All specifications include year dummies, a constant, a cubic in own math SAT, average peer math SAT, and individual-level random scrambling controls: gender, race, recruited athlete, prep school, CEER, and WCS. Average peer group scrambling controls added as indicated. See Table 1 notes for sample description.

standard deviation increase in the variance of peer group math SAT score improves company average math grades by 7.8 percent of a standard deviation and company average GPA by 8.3 percent of a standard deviation.

To test whether these estimates are sensitive to the inclusion of other aggregate peer group characteristics that may also be correlated with academic outcomes, I include average peer group measures of the eight scrambling controls in column 2.[9] This has little effect on the estimates of either the peer group variance or the peer group mean, providing further evidence of the robustness of these findings.

One concern with using the variance as a measure of dispersion is that it is sensitive to observations at the extremes. An alternative measure of dispersion that is less

[9] All specifications control for individual-level random scrambling controls.

susceptible to outliers is the 75–25 differential in the peer group math SAT distribution (seventy-fifth math SAT percentile minus the twenty-fifth math SAT percentile). As an additional robustness check, I include estimates for γ_{75-25} in panel A of Table 6.

Column 1 contains the preferred specification and column 2 adds the average peer group measures of the eight scrambling controls. The signs on the estimates of γ_{75-25} are consistent with those of γ_{var}, and the point estimates are not sensitive to the inclusion of additional aggregate controls. Estimates of γ_{75-25} imply that a one standard deviation increase in the peer group 75–25 differential in math SAT scores positively affects company average plebe-year math grades by 13 percent of a standard deviation and company average GPA by 15.6 percent of a standard deviation. The effects of the 75–25 differential are slightly larger than the variance effects in Table 5.

On balance, these estimates indicate that plebes with greater peer group heterogeneity in math SAT scores achieve higher plebe math grades and GPAs than plebes with less peer group heterogeneity in math SAT scores.[10] This result is robust to different measures of dispersion. Moreover, the estimates are stable when other aggregate peer group characteristics are included in the model. In the context of the Hoxby and Weingarth (2006) peer effect models, the positive effect of peer group heterogeneity provides support for the "Rainbow" model, where students are best off when forced to deal with all other types of students. From a policy perspective, this result implies that if West Point maintains its current admissions standards, there are no linear-in-means effects, and West Point continues to seek relatively uniform companies, then it can improve educational production efficiency by also maximizing the dispersion in peer ability.

VI. Distributional Effects

The finding that greater heterogeneity in peer group composition has positive effects on academic attainment raises the question of whether having a higher upper or lower bottom tail of the distribution influences this outcome. For instance, the positive effect of γ_{75-25} may be due to plebes having a higher seventy-fifth or a lower twenty-fifth math SAT percentile of their peer group. To determine if one tail of the distribution is driving the 75–25 differential result, I include the seventy-fifth percentile and the twenty-fifth percentile of the peer math SAT distribution as represented by $\hat{Z}_{pt-1}$ in equation (1).

Panel B of Table 6 (math grades percentiles and GPA percentiles) contains these specifications and is organized similar to panel A. For both math grades and GPA, having a higher seventy-fifth percentile of the peer math SAT distribution has positive and significant effects on individual academic attainment, while the twenty-fifth percentile has no significant effect. A one standard deviation increase in the seventy-fifth percentile of the peer math SAT distribution increases company average math

[10] Jacob Vigdor and Thomas Nechyba (2005), using data on fifth graders in North Carolina, also find a positive and significant increase in standardized test scores for children assigned to classrooms with a wider variance in lagged test scores.

TABLE 7—QUANTILE REGRESSION FOR THE 75–25 DIFFERENTIAL AND THE 75TH AND 25TH PERCENTILE
(*Outcome variable: individual plebe math grade or plebe GPA*)

	75th math quantile (1)	50th math quantile (2)	25th math quantile (3)	75th math quantile (1)	50th math quantile (2)	25th math quantile (3)
	Math grades			GPA		
Panel A. 75–25 differential of the peer math SAT distribution						
Peer group 75–25 math SAT differential/100	0.063	0.116	0.158	0.077	0.099	0.096
	(0.044)	(0.041)	(0.049)	(0.030)	(0.031)	(0.033)
Peer group mean math SAT/100	−0.018	−0.078	0.050	0.073	0.033	0.058
	(0.089)	(0.083)	(0.100)	(0.062)	(0.062)	(0.065)
Observations	6,309	6,309	6,309	6,870	6,870	6,870
	Math grades			GPA		
Panel B. 75th and 25th percentiles of the peer math SAT distribution						
Peer group seventy-fifth math SAT percentile/100	0.076	0.196	0.277	0.169	0.190	0.191
	(0.107)	(0.096)	(0.105)	(0.067)	(0.073)	(0.083)
Peer group twenty-fifth math SAT percentile/100	−0.055	−0.048	−0.054	−0.003	−0.024	−0.011
	(0.096)	(0.086)	(0.094)	(0.060)	(0.066)	(0.076)
Peer group mean math SAT/100	−0.033	−0.210	−0.192	−0.074	−0.131	−0.108
	(0.197)	(0.176)	(0.194)	(0.123)	(0.135)	(0.153)
Observations	6,309	6,309	6,309	6,870	6,870	6,870

Notes: Standard errors in parentheses. Quantile regression estimates for either the seventy-fifth, fiftieth, or twenty-fifth grade quantiles are listed as column headings These estimates reflect quantile regressions of individual academic outcomes on measures of dispersion as listed in the panel headings. All specifications include year dummies, a constant, a cubic in own math SAT, average peer math SAT, and individual-level random scrambling controls: gender, race, recruited athlete, prep school, CEER, and WCS. See Table 1 notes for sample description.

grades by 18 percent of a standard deviation and company average GPA by 26 percent of a standard deviation.

It is also useful to explore the effects of increased heterogeneity across the distribution of academic outcomes. To address this issue, I estimate specifications using quantile regression instead of OLS. Panel A of Table 7 contains quantile regression estimates where the variable of interest is the peer group 75–25 peer math SAT differential. Panel B of Table 7 contains quantile regression estimates where the variable of interest is the seventy-fifth and twenty-fifth peer math SAT percentiles. Estimates of quantiles of the math grade distribution are on the left and estimates of quantiles of the GPA are on the right. Column 1 contains estimates at the seventy-fifth quantile, column 2 contains estimates at the fiftieth quantile, and column 3 contains estimates at the twenty-fifth quantile of the plebe math grade or GPA distribution.

The magnitude of the estimates in panel A of Table 7 are comparable to the magnitude of the estimates in panel A of Table 6. They reveal relatively uniform, positive, and significant effects of heterogeneity across the distribution of academic outcomes. The point estimates for math grades in panel A indicate a stronger effect at the twenty-fifth quantile than at the seventy-fifth quantile, although these are not statistically different from each other. In general, the estimates in panel A suggest that all parts of the academic outcome distribution may benefit, to some degree, from greater heterogeneity.

Likewise, estimates in panel B of Table 7 are comparable to estimates in panel B of Table 6. For both math grades and GPA, the twenty-fifth math SAT percentile has no significant effect across any part of the distribution of academic outcomes. However, for math grades, the seventy-fifth math SAT percentile has positive and significant effects at the twenty-fifth and fiftieth quantile, and insignificant effects at the seventy-fifth quantile, although they are not statistically different. The point estimates suggest that having a higher seventy-fifth percentile of the math SAT distribution may be more helpful at the lower tail than at the upper tail of the math grade distribution. Estimates for GPA show a positive and relatively uniform effect of the seventy-fifth math SAT percentile across the GPA distribution.

VII. Conclusions

This paper investigates how peer group heterogeneity affects individual academic outcomes. These findings have important implications for how educational institutions organize students in group settings to improve production efficiency. The environment at West Point provides a unique opportunity to estimate how peer group heterogeneity affects educational attainment. The random assignment of plebes to peer groups and the exclusive use of pretreatment measures of ability as the peer effects of interest overcomes the well-documented empirical problems associated with identifying peer effects.

Estimates of measures of dispersion in peer group math SAT distributions reveal that more heterogeneous peer groups have positive effects on academic outcomes. This finding is robust across several measures of dispersion. A one standard deviation increase in the peer group 75–25 differential in math SAT scores increases company average math grades by 13 percent of a standard deviation and company average GPA by 15.6 percent of a standard deviation. This effect is relatively uniform across the distribution of academic outcomes and provides evidence in support of the Hoxby and Weingarth (2006) "Rainbow" model where heterogeneity promotes learning.

For the estimates of the 75–25 differential in peer group math SAT scores, the seventy-fifth percentile, but not the twenty-fifth percentile, accounts for most of this effect. A one standard deviation increase in the seventy-fifth percentile of the peer group math SAT distribution increases company average math grades by 18 percent of a standard deviation and company average GPA by 26 percent of a standard deviation. This result is consistent with findings from Duflo, Dupas, and Kremer (2008) in which students benefit from having academically stronger peers. While probably also consistent with the view of most educators, this experiment affords a rare opportunity to provide empirical evidence for this commonly held belief.

On the whole, these findings imply that if West Point continues to maintain its current admissions standards and relatively equal peer group means in their assignment process, then maximizing heterogeneity in peer group ability can lead to improved production efficiency. Although West Point is a unique setting, these findings both contribute to the literature on peer effects in higher education among the nation's elite institutions (Dartmouth, Williams, etc.) and raise several important questions for future research. For example, in this experiment, peer means are relatively uniform across individuals, and there are few outliers in the lower tail of the ability distribution at West Point. Other settings, where means vary to a greater extent, and where the lower tail is longer, may produce different results. There may also be other factors, such as signaling and networks, that create larger economic effects when schools segregate students by ability that could offset the gains found by mixing individuals in this study.

REFERENCES

Acemoglu, Daron. 2002. "The Theory of Human Capital Investments." In *Lecture Notes for Graduate Labor Economics, 14.662.* Part 1, Chapter 4. http://www.chass.utoronto.ca/~siow/2801/acemoglu_labor_notes.pdf.

Benabou, Roland. 1996. "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance." *American Economic Review*, 86(3): 584–609.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2008. "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." National Bureau of Economic Research Working Paper 14475.

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

Guryan, Jonathan, Kory Kroft, and Matt Notowidigdo. 2007. "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments." National Bureau of Economic Research Working Paper 13422.

Hoxby, Caroline M., and Gretchen Weingarth. 2006. "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects." http://www.hks.harvard.edu/inequality/Seminar/Papers/Hoxby06.pdf.

Lyle, David S. 2007. "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point." *Review of Economics and Statistics*, 89(2): 289–99.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531–42.

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

Schneider, Dianne, and Neil J. Dorans. 1999. "Concordance Between SAT® I and ACT™ Scores for Individual Students." The College Board Office of Research and Development Report RN-07. http://professionals.collegeboard.com/profdownload/pdf/concordance_between_s_10502.pdf.

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

United States Military Academy. 1998. "Evaluations of Scrambling in the Corps of Cadets, 1962–1998: 98-007." West Point, NY: United States Military Academy Printing Office.

Vigdor, Jacob, and Thomas Nechyba. 2005. "Peer Effects in North Carolina Public Schools." In *Schools and the Equal Opportunity Problem*, ed. Ludger Woessmann and Paul E. Peterson, 73–102. Cambridge, MA: MIT Press.

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