

Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment[†]

By ROBERT GARLICK*

I study the relative academic performance of students tracked or randomly assigned to South African university dormitories. Tracking reduces low-scoring students' GPAs and has little effect on high-scoring students, leading to lower and more dispersed GPAs. I also directly estimate peer effects using random variation in peer groups across dormitories. Living with higher-scoring peers raises students' GPAs, particularly for low-scoring students, and peer effects are stronger between socially proximate students. This shows that much of the treatment effect of tracking is attributable to peer effects. These results present a cautionary note about sorting students into academically homogeneous classrooms or neighborhoods. (JEL I23, I24, I28, O15, Z13)

Group structures are ubiquitous in education and group composition may have important effects on education outcomes. Students in different classrooms, living environments, schools, and social groups are exposed to different peers, in addition to different education inputs and institutional environments. A growing literature shows that students' peer groups influence their education outcomes.¹ Peer effects play an important role in policy debates around academic tracking, school

*Department of Economics, Duke University, 419 Chapel Drive, 213 Social Sciences Building, Box 90097, Durham, NC 27708 (email: rob.garlick@gmail.com). I am particularly grateful to Jeff Smith, David Lam, John DiNardo, Brian Jacob, and Manuela Angelucci for advice and guidance throughout this project. I also appreciate helpful suggestions from Peter Arcidiacono, Raj Arunachalam, Emily Beam, John Bound, Tanya Byker, Scott Carrell, Susan Godlonton, Andrew Goodman-Bacon, Italo Gutierrez, Brad Hershbein, Stephen Ross, Rebecca Thornton, Adam Wagstaff, Dean Yang, several anonymous referees, and seminar participants at Chicago Harris School, Columbia University, Columbia Teachers College, Cornell University, Duke University, Harvard Business School, LSE, University of Michigan, Michigan State University, Northeastern University, University of Notre Dame, SALDRU, Stanford SIEPR, UC Davis, the World Bank, Yale School of Management, and the ASSA, CSAE, EconEcon, ESSA, MIEDC, NEUDC, PacDev, and SOLE conferences. I received invaluable assistance with student data and institutional information from Jane Hendry, Josiah Mavundla, and Charmaine January at the University of Cape Town. I acknowledge financial support from the University of Michigan's Ford School of Public Policy and Rackham School of Graduate Studies. All errors are my own.

[†]Go to <https://doi.org/10.1257/app.20160626> to visit the article page for additional materials and author disclosure statement or to comment in the online discussion forum.

¹Manski (1993) lays out the identification challenge in studying peer effects: do correlated outcomes within peer groups reflect correlated predetermined characteristics, common institutional factors, or peer effects—causal relationships between students' outcomes and their peers' characteristics? Many education researchers address this challenge using randomized or controlled variation in peer group composition. Peer effects have been documented on standardized test scores (Hoxby 2000), college GPAs (Sacerdote 2001), college entrance examination scores (Ding and Lehrer 2007), cheating (Carrell, Malmstrom, and West 2008), job search (Marmaros and Sacerdote 2002), and major choices (De Giorgi, Pellizzari, and Redaelli 2010).

choice, charter schools, and neighborhood segregation.² Most peer effects studies examine the effect of assignment to or selection into different peer groups for a given group assignment policy or selection process (Sacerdote 2011). This limits the ability of these studies to shed light on the relative effects of different group assignment policies or selection processes.

In this paper, I study the relative effects of two residential group assignment policies—randomization and tracking based on prior academic performance—on the distribution of student outcomes. I contribute to two literatures: optimal peer group design and academic tracking. Comparison of different group assignment policies corresponds to a clear social planning problem: how should students be assigned to groups to maximize some target outcome, subject to a given distribution of student characteristics? Unlike most education interventions, different group assignment policies may improve education outcomes without requiring more education inputs. I study a setting where students in different tracks have different residential peer groups but not different instructors or curricula. I argue that this isolates the role of peer effects in tracking, whereas classroom tracking studies may not separately identify peer effects from teacher or curriculum effects.

I study peer effects under two different group assignment policies at the University of Cape Town in South Africa. First-year students at the university were tracked into dormitories up to 2005 and randomly assigned from 2006 onward. This created residential peer groups that were respectively homogeneous and heterogeneous in prior academic performance. I contrast the distribution of first-year students' academic outcomes under the two policies. I use non-dormitory students in a difference-in-differences design to remove time trends and cohort effects. I show that tracking leads to lower and more dispersed grade point averages (GPAs) than random assignment. Low-scoring students perform substantially worse under tracking than random assignment, while high-scoring students' GPAs are approximately equal under both policies.

To understand the mechanisms driving this result, I use random assignment to dormitories to estimate directly the effect of living with higher- or lower-scoring peers. All students' GPAs are increasing in the prior academic performance of their peers. Peer effects operate largely within race groups, suggesting that social proximity mediates peer effects, but are not systematically stronger when students take the same classes as their dormitory peers. I estimate a reduced-form but theory-guided model of nonlinear peer effects and find that low-scoring students are more sensitive to peer group composition than high-scoring students. The cross-dormitory peer effects estimates imply a negative average effect of tracking, particularly for low-scoring students, suggesting that much of the tracking effect is due to peer effects. However, this analysis does not rule out some role for other mechanisms.

This paper contributes to two literatures in economics: on academic tracking and on peer effects. I contribute to the literature on academic tracking by studying tracking into noninstructional groups. Most existing papers estimate the effect of school or classroom tracking relative to another assignment policy or of assignment to

²See Arnott and Rowse (1987) and Duflo, Dupas, and Kremer (2011) on classroom tracking, Epple and Romano (1998) and Hsieh and Urquiola (2006) on school choice, Angrist et al. (2016) on charter schools, and Bénabou (1996) and Kling, Liebman, and Katz (2007) on neighborhood segregation.

different tracks. However, tracked and untracked groups may differ on resources and instructor behavior, as well as peer group composition.³ Isolating the causal effect of tracking on student outcomes via peer effects, net of these other factors, is difficult. For example, Booij, Leuven, and Oosterbeek (2017); Duflo, Dupas, and Kremer (2011); and Feld and Zölitz (2017) show that low-scoring students perform better in relatively homogeneous classrooms with few high-scoring students but attribute this to entirely different mechanisms.⁴ I study a setting where students in high- and low-track dormitories and students not in dormitories take classes together from the same instructors, minimizing scope for tracking to affect instruction. Variation in dormitory-level attributes might in principle affect student outcomes, but my results are robust to conditioning on these attributes. I cannot rule out the possibility that assignment to low-track dormitories has negative psychological effects on students.

I contribute to the peer effects literature by combining variation in peer group assignment policy with random variation in peer group composition. I show that student outcomes are affected both by residential peers' prior academic performance and by changes in the peer group assignment policy. Both cross-policy and cross-dormitory comparisons show that students with lower prior academic performance are more sensitive to changes in peer group composition. I link this finding to the literatures on optimal group composition in the presence of spillovers (Bénabou 1996) and on marriage matching (Becker 1973). These models apply naturally to the study of peer effects but have received limited attention in this literature other than Graham, Imbens, and Ridder (2010).⁵ The peer effects estimated under random assignment predict effects of tracking that are broadly consistent with the directly estimated effects. This confirms that much of the effect of tracking is due to peer effects. The predictions do not exactly match the directly estimated effects of tracking, which may reflect the challenge of using peer effects for predicting policy effects raised by Bhattacharya (2009); Carrell, Sacerdote, and West (2013); Fruehwirth (2014); Graham, Imbens, and Ridder (2010); and Hurder (2012).

I find that peer effects operate almost entirely within race groups, consistent with results from the United States (Fruehwirth 2013; Hanushek, Kain, and Rivkin 2009; and Hoxby 2000). However, I find that dormitory peer effects are not stronger within than across classes or programs of study. An economics student, for example, is no more strongly affected by other economics students in her dormitory than by noneconomics students in her dormitory. I believe that this finding is novel in the peer effects literature. Taken together, these results suggest that spatial proximity

³ Betts (2011) reviews an extensive literature comparing tracking to alternative assignment policies. A smaller literature studies the effect of assignment to different tracks in an academic tracking system (Abdulkadiroğlu, Angrist, and Pathak 2014; Ajayi 2016; Lucas and Mbiti 2014; Pop-Eleches and Urquiola 2013).

⁴ Duflo, Dupas, and Kremer (2011) attributes the positive effect of tracking to better targeted teaching. This is consistent with multiple studies finding that instruction targeted at students' actual attainment level is more effective than non-targeted instruction (Evans and Popova 2016). However, Booij, Leuven, and Oosterbeek (2017) and Feld and Zölitz (2016) find that instruction does not vary with classroom composition and attribute the positive effect of tracking to better group interaction.

⁵ Both classes of models imply optimal peer group assignment policies given specific nonlinearities in the relationship between GPA and peer attributes. Hoxby and Weingarth (2006) provides a taxonomy of peer effects models with nonlinearities but do not frame their discussion in terms of these models, while multiple papers document evidence of nonlinear peer effects (Burke and Sass 2013; Fruehwirth 2013; Imberman, Kugler, and Sacerdote 2012; and Lavy, Silva, and Weinhardt 2012).

generates peer effects only when students are also socially proximate and likely to interact. But the relevant form of interaction is not direct academic collaboration. Peer effects may instead operate through mechanisms such as time use or transfer of soft skills, consistent with Stinebrickner and Stinebrickner (2006).

I argue that my findings are of general interest for three reasons, even though residential tracking is less widespread than classroom tracking. First, residential segregation by academic performance is not uncommon: 35 of the 50 flagship state universities in the United States offer dormitories or residential areas restricted to students in honors programs. This generates residential peer groups similar to high-track dormitories, though not necessarily groups similar to low-track dormitories. Many residential universities allow students to sort into colleges, dormitories, cooperatives, or fraternities/sororities that may be relatively academically segregated. Second, the few studies that directly compare the magnitude of residential and classroom/study group peer effects find that the former are larger (Hoel, Parker, and Rivenburg 2004; Jain and Kapoor 2015). This suggests that composition of residential peer groups should be an important policy question. Third, neighborhood segregation by income and parental education can generate neighborhood peer groups, and hence social groups, that resemble low-track dormitories. My results are thus relevant to the study of neighborhood effects on human capital formation.

I. Research Design and Data

I study a natural experiment at the University of Cape Town in South Africa, where first-year students are allocated to dormitories using either random assignment or academic tracking. This is a selective research university but admits many students from low-performing high schools. The student population is thus relatively heterogeneous but not representative of South Africa.

Approximately half of the 3,500–4,000 first-year students live in 23 university dormitories. The mean dormitory size is 128 students and the interdecile range is [52, 220]. One dormitory closes in 2005 and another opens in 2007. I exclude seven very small dormitories that each hold fewer than 10 first-year students and one dormitory that is excluded from the randomization policy. This leaves a sample of 14 dormitories in the random assignment period and 15 dormitories in the tracking period. The dormitories provide accommodation, meals, and some organized social activities. Classes and instructors are shared across students from different dormitories and students who do not live in dormitories. Dormitory assignment thus determines the set of residentially proximate peers but not the set of classroom peers. Students are normally allowed to live in dormitories for at most two years. They can move out of their dormitory after one year but cannot change to another dormitory, so their second year peer group is not fully determined by dormitory assignment. Most students live in two-person rooms and the roommate assignment process varies across dormitories. I do not observe roommate assignments. The other half of the incoming first-year students live in private accommodation, typically with family in the Cape Town region.

I observe a dataset of students' dormitory assignments, first year university transcripts, transcripts from high school graduation tests, and demographics (sex, home

language, nationality, and race). I use these data to construct two scalar measures of students' academic performance, one prior to attending the university and one during their first year of university. Each student's first year university transcript lists all courses she took and, for each course, a credit weighting and a score ranging from 0 to 100. I construct a credit-weighted average score and then transform this to have mean zero and standard deviation one in the group of non-dormitory students, separately by year. I call this measure "GPA" and all treatment effects in the paper can be interpreted in standard deviations of GPA. The numerical scores are not typically "curved" and the nominal ceiling score of 100 does not bind.⁶ These features provide some reassurance that my results are not driven by time-varying grading standards or by ceiling effects on the grades of top students. I discuss these potential concerns in online Appendix B.

I also observe students' results in a set of national, content-based high school graduation tests taken by all South African grade 12 students. The tests were developed, graded, and moderated by a national body so are comparable across schools. I observe subject-specific letter grades that I convert into a single score ranging from 0 to 48, using the university's admissions algorithm.⁷ I transform this to have mean zero and standard deviation one in the group of non-dormitory students, separately by year. I call this measure "high school GPA" or "HSGPA" and all measures of prior academic performance in the paper can be interpreted in standard deviations of HSGPA. I interpret this score as a measure of students' prior academic performance that reflects a mix of student ability, prior effort, home and school inputs, and some noise.

Incoming students were tracked into dormitories up until the 2005 academic year. Tracking was based on either HSPGA or scores in "mock" tests taken 2–3 months before the final tests as preparation, depending on when students applied to the university.⁸ The resultant assignments do not partition the distribution of HSGPA for four reasons. First, many assignments were made based on mock test scores, before final test scores were observed by the university. Second, late applicants for admission were waitlisted and assigned to the first available dormitory slot created by an admitted student declining admission. Third, assignment incorporated loose racial quotas, so the threshold HSGPA for assignment to each dormitory tier was higher for white than black students. Fourth, most dormitories were single-sex, creating pairs of female and male dormitories at each track. Under the observed tracking policy, the average student in the top quartile of HSGPA lived with peers scoring

⁶For example, mean percentage scores on economics 1 and mathematics 1 change by respectively 6 and 9 percentage points from year to year, roughly half of a standard deviation. This is not consistent with strictly curved grading. The highest score any student obtains averaged across courses is 97 and the ninety-ninth percentile of student scores is 84.

⁷I observe grades for all six tested subjects for 85 percent of the sample, for five subjects for 6 percent of the sample, and for four or fewer subjects for 9 percent of the sample. I treat the third group of students as having missing scores. I assign the second group of students the average of their five observed grades. I include all three groups of students in the full-sample analysis, but exclude the second and third groups from all analyses that subdivide students by high school test scores. A time-invariant conversion scale is used to convert international students' A-level or international baccalaureate scores into the same 0–48 scale.

⁸The mock tests had the same subject structure as the final tests but were developed and moderated by provincial education departments, rather than the national department. Scores on the mock and final tests are likely to be strongly correlated, but I do not observe the mock test scores.

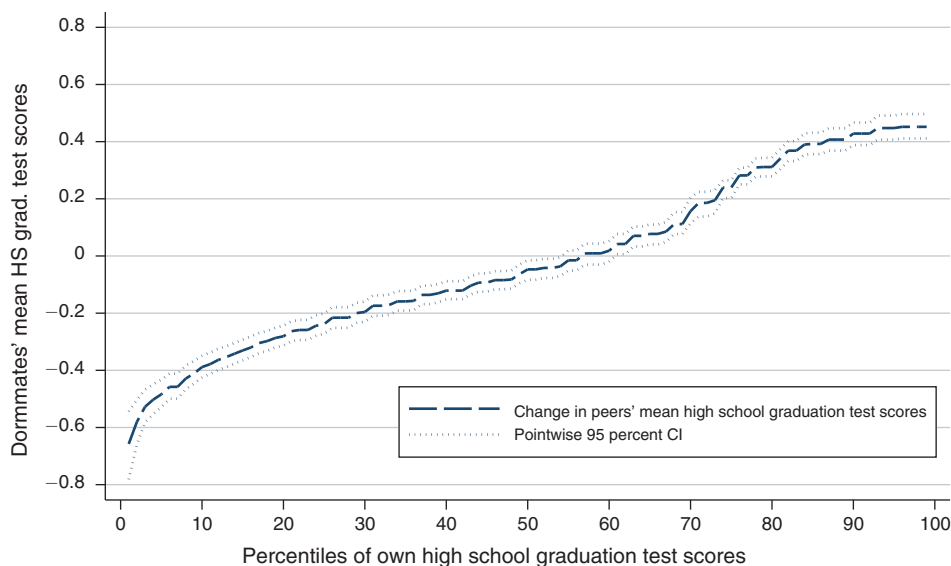


FIGURE 1. EFFECT OF TRACKING ON PEER GROUP COMPOSITION

Notes: The curve shows the fitted values from a student-level local linear regression of mean dormitory HSGPA on students' own HSGPA in the tracking period. The fitted values are evaluated at each percentile of the HSGPA distribution. The fitted values are shifted down by 0.20 standard deviations, the mean value of mean dormitory HSGPA under random assignment. So the curve shows the change in mean dormitory HSGPA from the random assignment to tracking period for students at each percentile of HSGPA. The dotted lines show pointwise 95 percent confidence intervals from 1,000 iterations of a percentile bootstrap.

0.92 standard deviations higher on HSGPA than the average student in the bottom quartile. The difference would be larger under a race-blind, sex-blind tracking policy with no waitlisting that used final HSGPA for assignments.⁹

From 2006 onward, incoming students were randomly assigned to dormitories. The policy change reflected concern by university administrators that tracking was inequalitarian and contributed to social segregation by income. Assignment used a random number generator with ex post changes to avoid racial imbalance. In a 2009 interview, the staff member responsible for assignment recalled making only “occasional” changes. Most dormitories remained single-sex. One small dormitory was excluded from the randomization because it did not provide meals and charged a lower accommodation fee. Students could request to live in this dormitory. I exclude this dormitory from my sample in the randomization period but all results are robust to including it.

The policy change induced a large change in students' peer groups. Figure 1 shows how tracking affected the relationship between students' own HSGPA and their peers' HSGPA. For example, students in the top decile lived with peers who

⁹This hypothetical tracking policy might have different effects to the observed tracking policy but strong assumptions are needed to characterize this difference. For example, I show in Section IV that peer effects differ within and across race groups. Given this result, we can only characterize the different effects of different types of tracking policies by modeling the joint distribution of dormitory sizes, race group sizes, and race-specific HSGPA. I do not attempt to do so in this paper.

TABLE 1—EFFECTS OF TRACKING ON PEER GROUP COMPOSITION

	Tracked dorm students (1)	Randomized dorm students (2)	Difference (1) – (2) (3)
Variance of HSGPA in dormitory	0.740	0.857	–0.117
Interquartile range of HSGPA in dormitory	1.244	1.436	–0.192
Fraction of dormmates of own sex	0.900	0.934	–0.034
Fraction of dormmates of own race	0.436	0.412	0.024
Fraction of dormmates of own language	0.379	0.321	0.058
Fraction of dormmates in own program	0.284	0.276	0.008

Notes: Table 1 reports summary statistics on dormitory composition for tracked and randomly assigned dormitory students in columns 1 and 2, respectively. The statistic “fraction of dormmates of own sex” is defined at the student level and equals, for a female/male student, the proportion of female/male students in her/his dormitory. The other three “fraction of ...” statistics are defined in the same way. All differences in column 3 are significantly different from 0 at the 5 percent level.

scored approximately 0.5 standard deviations of HSGPA higher under tracking than random assignment; students in the bottom decile lived with peers who scored approximately 0.3 standard deviations of HSGPA lower. This is the identifying variation I use to study the effect of tracking.

Table 1 shows that the within-dormitory dispersion of HSGPA is substantially lower under tracking (rows 1–2). The within-dormitory variance and interquartile range remain quite high under tracking because the distribution of HSGPA has a long left tail. The within-dormitory variance under random assignment is significantly less than one because most dormitories are sex-segregated and female students have on average higher HSGPA than male students. Relative to random assignment to all mixed-sex dormitories, this decreases the variance of HSGPA within dormitories and increases the variance of mean HSGPA across dormitories. Peer groups under tracking are slightly more homogeneous in terms of race, language, and program of study (rows 4–6). However, cross-dormitory and cross-policy variation in these proportions is not robustly associated with differences in students’ outcomes and all results in the paper are robust to controlling for these proportions.

My research design compares the students’ first year GPAs between the tracking period (2004 and 2005) and the random assignment period (2007 and 2008). I define tracking as the “treatment” even though it is the earlier policy. I omit 2006 because first-year students were randomly assigned to dormitories while second-year students continued to live in the dormitories into which they had been tracked. GPA differences between the two periods may reflect cohort effects as well as peer effects. In particular, benchmarking tests show a downward trend in the academic performance of incoming first-year students at South African universities over this time period (Higher Education South Africa 2009). I therefore use a difference-in-differences design that compares the time change in dormitory students’ GPAs with the time change in non-dormitory students’ GPAs over the same period:

(1)

$$GPA_{id} = \delta_0 + Dorm_{id} \cdot \delta_1 + Track_{id} \cdot \delta_2 + DormTrack_{id} \cdot \delta_3$$
$$+ \mathbf{X}_{id} \cdot \boldsymbol{\delta} + \mu_d + \epsilon_{id},$$

TABLE 2—SUMMARY STATISTICS AND BALANCE TESTS

	Number of students (1)	Mean (2)	SD (3)	Min (4)	Max (5)	Normalized double- difference (6)	<i>p</i> -value for zero double- difference (7)
University GPA	14,598	0.056	0.943	−4.427	2.641		
High school GPA	12,410	0.090	1.011	−4.031	3.553	−0.039	0.276
Mean dorm HSGPA, randomization	28	0.244	0.180	−0.214	0.653		
Mean dorm HSGPA, tracking	30	0.002	0.771	−2.269	1.243		
Dormitory size	58	128	63	44	258	−0.134	0.610
Female	14,598	0.515	0.500			−0.062	0.061
Black	14,598	0.317	0.465			−0.030	0.359
White	14,598	0.425	0.494			−0.017	0.611
Other race	14,598	0.258	0.437			0.050	0.135
English-speaking	14,598	0.715	0.451			0.099	0.003
International	14,598	0.142	0.349			0.028	0.389
Tracked	14,598	0.504	0.500				
Graduated high school in tracking period	14,002	0.517	0.500			−0.113	0.000
Lives in dormitory	14,598	0.508	0.500				
Graduated from high school in Cape Town	13,324	0.413	0.492			−0.016	0.645

Notes: Table 2 reports summary statistics for the main outcome (university GPA), student characteristics determined prior to university admission, dormitory size, and mean dormitory HSGPA under each assignment policy. Minimum and maximum values are not shown for binary variables. Each normalized double-difference in column 6 equals the second difference between dormitory and non-dormitory students between the tracking and random assignment periods, divided by the variable's standard deviation. Each *p*-value in column 7 is derived from a test for a zero double-difference, i.e., a test for the parallel trends assumption holding for this variable.

where i and d index students and dormitories, $Dorm$ and $Track$ are indicator variables respectively equal to 1 for students living in dormitories and for students enrolled in the tracking period, \mathbf{X}_{id} is a row vector of students' demographic covariates and HSGPA, and μ_d is an indicator equal to one for students living in dormitory d and zero for non-dormitory students and students living in other dormitories.

The variable δ_3 equals the average treatment effect of tracking on dormitory students under a “parallel trends” assumption: that dormitory and non-dormitory students would have experienced the same mean time trend in GPAs if the assignment policy had remained constant. If the observed time trend for non-dormitory students does not equal the counterfactual time trend for dormitory students, then δ_3 does not equal the average treatment effect of tracking on dormitory students. If the change in assignment policy affects students through mechanisms other than peer effects, δ_3 recovers the correct treatment effect but its interpretation changes. I present a variety of robustness checks and falsification tests in online Appendix B that support the parallel trends assumption and peer effects interpretation. Note that the difference-in-differences design identifies only a treatment on the treated effect; the effect tracking would have on non-dormitory students is not identified without stronger assumptions.

The assumption of parallel trends in mean GPA is not directly testable. But I can test the related assumption that dormitory and non-dormitory students experience parallel time trends in covariates determined prior to university. I report these trends in Table 2 and show more detailed balance tests in online Appendix A. There are statistically significant differences in three of the ten covariates—sex, language, and period of high school graduation—but the deviations from parallel trends are small

in magnitude, with a mean absolute value of 0.06 standard deviations. There is a strong correlation between living in a dormitory and graduating from a high school outside of Cape Town, as the university gives these students priority for dormitory accommodation.¹⁰ I use this admission rule in Section II to construct a geographic instrument for whether or not students live in a dormitory. This relationship is stable through time, providing reassurance that students did not strategically choose whether or not to live in dormitories in response to the dormitory assignment policy change. This reflects the limited information available to prospective students about the dormitory assignment policy: the change was not announced in the university's admissions materials or in internal, local, or national media.

Omitting students living in the one non-randomly assigned dormitory during the random assignment period generates some mechanical violations of the parallel trends assumption. Students in this dormitory are disproportionately likely to have low HSGPA and to be black, male, and enroll in the university several years after graduating from high school. These students are omitted from the sample of dormitory students in the random assignment period but not in the tracking period. If students living in this dormitory in both periods are included in the sample, the mean absolute value of the double-differences drops to 0.04 standard deviations and only one covariate—language—violates the parallel trends assumption.

Even though the time trends in preuniversity characteristics are similar for dormitory and non-dormitory students, their levels are different: dormitory students have higher average HSGPA and are more likely to be black, speak a language other than English, and to be international students. This is not necessarily a problem for the difference-in-differences design. But I account for these differences by including in \mathbf{X}_{id} HSGPA, HSGPA squared, language, nationality, sex, race, and all two-way interactions between these variables.

II. Average Effects of Tracking

Tracked dormitory students obtain GPAs 0.13 standard deviations lower than randomly assigned dormitory students (Table 3, column 1). This finding is robust to conditioning on HSGPA, demographic covariates, and dormitory fixed effects using regression and/or re-weighting and to different strategies to account for missing HSGPA data (columns 2–5). The point estimate in my preferred specification is 0.12 standard deviations with a 95 percent confidence interval of $[-0.20, -0.03]$.¹¹

¹⁰ I do not observe students' home addresses, which are used for the university's dormitory admissions. Instead, I match records on students' high schools to a public database of high school GIS codes. I then determine whether students attended high schools in or outside the Cape Town metropolitan area. This is an imperfect proxy of their eligibility to live in a dormitory for three reasons: long commutes and boarding schools are fairly common, the university allows students from low-income neighborhoods on the outskirts of Cape Town to live in dormitories, and a small number of Cape Town students with medical conditions or very high graduation test scores are permitted to live in the dormitories.

¹¹ The bootstrapped standard errors used throughout the paper allow clustering at the dormitory-year level. Non-dormitory students are treated as individual clusters, yielding 58 large clusters and approximately 7,200 singleton clusters. As a robustness check, I also test if the treatment effect estimates equal zero using a wild cluster bootstrap (Cameron, Gelbach, and Miller 2008). The wild bootstrap p -values are 0.076 for the basic regression model (Table 3, column 1) and < 0.001 for the model with dormitory fixed effects and student covariates (Table 3, column 3). I

TABLE 3—AVERAGE TREATMENT EFFECT OF TRACKING ON TRACKED STUDENTS' GPAS

	(1)	(2)	(3)	(4)	(5)
Tracking \times dormitory	−0.134 (0.073)	−0.101 (0.040)	−0.117 (0.044)	−0.137 (0.063)	−0.125 (0.064)
Tracking	0.000 (0.023)	0.002 (0.021)	−0.012 (0.021)	0.039 (0.050)	0.002 (0.051)
Student covariates		\times	\times	\times	\times
Missing data indicators			\times		\times
Dormitory fixed effects		\times	\times	\times	\times
Re-weighting				\times	\times
Adjusted R^2	0.006	0.255	0.231	0.238	0.227
Number of dormitory-year clusters	58	58	58	58	58
Number of dormitory students	7,410	6,571	7,410	6,571	7,410
Number of non-dormitory students	7,188	6,685	7,188	6,685	7,188

Notes: Table 3 reports results from regressing GPA on indicators for living in a dormitory, the tracking period, and their interaction. Columns 2–5 report results controlling for dormitory fixed effects and student covariates: sex, language, nationality, race, a quadratic in HSGPA, and all pairwise interactions. Columns 2 and 4 report results excluding students with missing HSGPA from the sample. Columns 3 and 5 report results including all students, with missing HSGPA replaced with zeros and controlling for a missing test score indicator. Columns 4 and 5 report results from propensity score-weighted regressions that re-weight all groups to have the same distribution of observed student covariates as tracked dormitory students. Standard errors in parentheses are from 1,000 bootstrap iterations clustering at the dormitory-year level, stratifying by dormitory status and assignment policy, and re-estimating the weights on each iteration. Results are similar with linear and cubic functions of the covariates as controls, which respectively drop the squared test score and two-way interaction terms and add a cubed test score and three-way interaction terms.

How large is a treatment effect of -0.12 standard deviations? This is substantially smaller than the black-white GPA gap at this university (0.46 standard deviations) but larger than the male-female GPA gap (0.09).¹² The effect size is marginally smaller when students are strategically assigned to squadrons at the US Airforce Academy (Carrell, Sacerdote, and West 2013) and marginally larger when Kenyan primary school students are tracked into classrooms (Duflo, Dupas, and Kremer 2011). These results provide a consistent picture about the plausible average short-run effects of alternative group assignment policies. The effect sizes are not very large but are substantial relative to many other education interventions. However, the direction of effects are not consistent across settings (for example, Duflo, Dupas, and Kremer (2011) find a positive average effect of classroom tracking).

In online Appendix B, I consider five alternative explanations that might have generated the mean GPA difference between tracked and randomly assigned dormitory students. I argue that none of these explanations are likely to account for this difference. I provide a brief overview of this analysis here. First, I show that there is limited evidence of time-varying selection into dormitory status. The relationship between dormitory status and observed covariates is stable between the tracking and random assignment periods, except for the imbalance induced by dropping the one non-randomly assigned dormitory in the random assignment period.

also account for the possibility of persistent dormitory-level shocks with a wild bootstrap clustered at the dormitory level. These p -values are 0.189 and < 0.001 for the models in Table 3, columns 1 and 3.

¹²This magnitude is robust to using different GPA scales. Using raw GPA, the treatment effect is -2.22 points on a 0–100 scale, while the black-white and male-female gaps in raw GPA are 6.74 and 1.21, respectively.

My results are robust to accounting for selection on unobserved covariates using a sensitivity analysis in the spirit of Altonji, Elder, and Taber (2005). I also account for selection on unobserved covariates by instrumenting dormitory status with an indicator for attending a high school outside Cape Town. The university gives priority for dormitory accommodation to students from outside Cape Town. Students living outside Cape Town are 60 percentage points more likely to live in dormitories (F -stat = 1,411) and the instrumented treatment effect of tracking is -0.13 standard deviations. Second, I use a longer time series of data to show that GPA trends prior to the policy change are approximately parallel (Figure 2, top panel). I observe high school location but not dormitory assignments for this longer time series, so I use an intention-to-treat approach for this trend analysis. I also show that the trends in the share of high school graduates from in and outside Cape Town who qualify for admission to the university are parallel over an eight-year period around the policy change (Figure 2, bottom panel). Together, these results indicate that the composition of potential and actual dormitory and non-dormitory students is similar through time.

Third, I consider the possibility of spillover effects between dormitory and non-dormitory students. I cannot directly test for spillover effects. But I show that the most plausible model of spillovers generates testable predictions for non-dormitory students' GPAs that do not match the data. This analysis of time-varying selection, differential time trends, and spillover effects supports the hypothesis that the change in assignment policy drives the main results.

Fourth, I explore the nature of the grading system that generates the GPA measures. I show that the results are unlikely to be driven by "curving," truncation of GPA, or changes in course selections. This suggests that the results reflect effects on learning, not some peculiarity of the grading system. Fifth, I explore whether the treatment effect of tracking may be due to some mechanism other than peer effects: interaction effects between student characteristics and dormitory amenities, discriminatory grading, or negative psychological effects of assignment to low-track dormitories. The interaction effects and discriminatory grading hypotheses yield additional testable implications that do not match the data. This analysis of the grading system and alternative mechanisms shows that the change in assignment policy affects learning outcomes, and does so through peer effects.

III. Heterogeneous Effects of Tracking

Tracking changes peer groups in different ways for different students: high-scoring students live with higher-scoring peers and low-scoring students live with lower-scoring peers. The effects of tracking are thus likely to vary systematically with students' HSGPAs. I explore this heterogeneity in two ways. I first estimate conditional average treatment effects for different subgroups of students. I then estimate quantile treatment effects of tracking, which show how tracking changes the full distribution of GPAs.

I begin by estimating equation (1) fully interacted with an indicator for students with above-median HSGPA. Above- and below-median students' GPAs fall respectively 0.01 and 0.24 standard deviations under tracking (standard errors 0.07 and

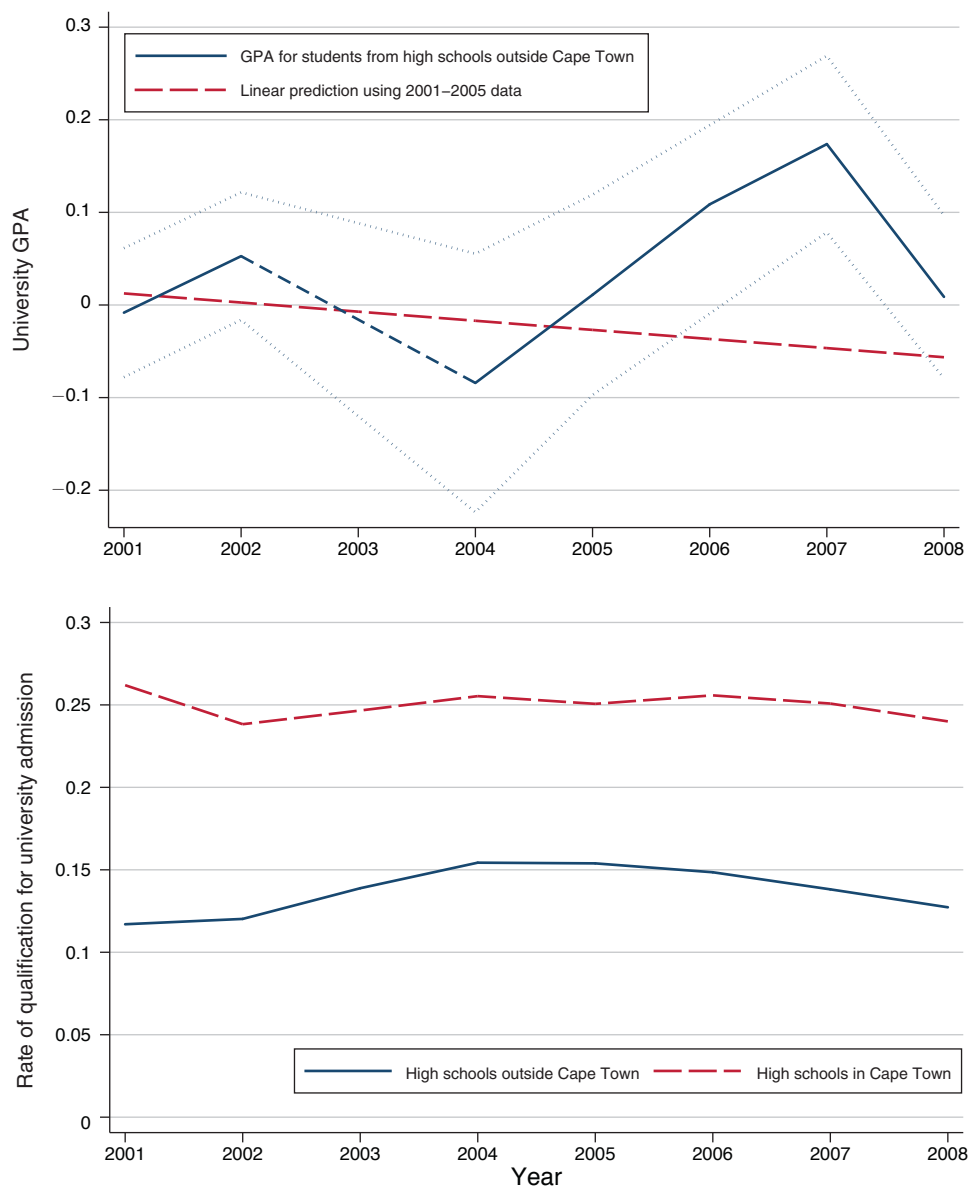


FIGURE 2. LONG-TERM TRENDS IN STUDENT ACADEMIC PERFORMANCE

Notes: The top panel shows mean GPA for first-year university students from high schools outside Cape Town. The time series covers the tracking period (2001–2005) and the random assignment period (2006–2008). Mean GPA for students from Cape Town high schools is, by construction, zero in each year. Data for 2003 is missing and replaced by a linear imputation. The dotted lines show a 95 percent confidence interval constructed from 1000 iterations of a percentile bootstrap stratifying by assignment policy and dormitory status. The bootstrap resamples dormitory-year clusters for 2004–2008, the only years in which dormitory assignments are observed. The bottom panel shows the proportion of grade 12 students whose high school graduation test score qualified them for admission to university. The mean qualification rate for high schools in Cape Town is 0.138 in the tracking period (2001–2005) and 0.133 in the random assignment period (2007–2008). The mean qualification rate for high schools outside Cape Town is 0.250 in the tracking period (2001–2005) and 0.245 in the random assignment period (2007–2008). The second difference is 0.001 (bootstrap standard error 0.009) or, after weighting by the number of grade 12 students enrolled in each school, 0.007 (standard error 0.009).

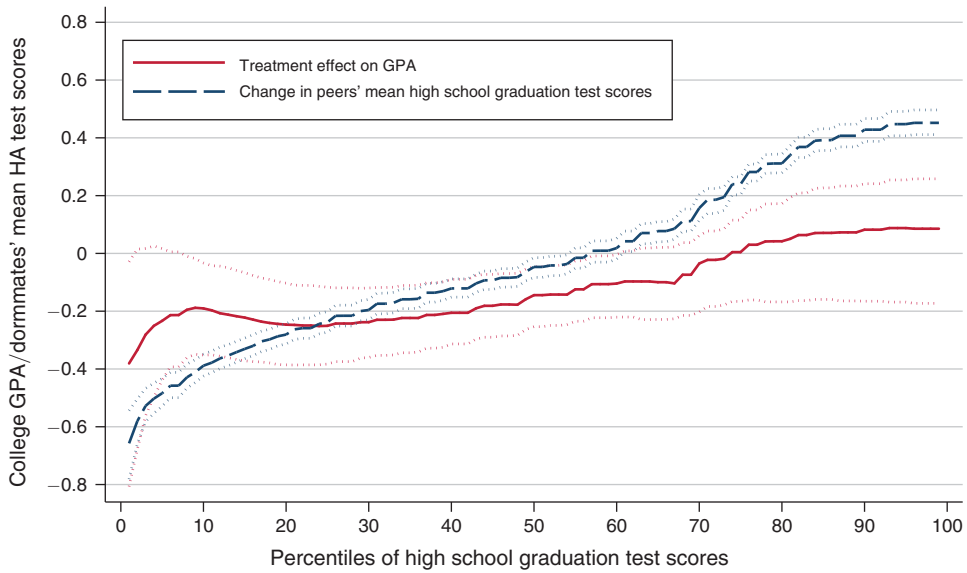


FIGURE 3. EFFECTS OF TRACKING ON GPA BY HIGH SCHOOL TEST SCORES

Notes: The curve labeled “Treatment effect on GPA” is constructed in three steps. First, I estimate a student-level local linear regression of GPA on students’ own HSGPAs. I estimate the regression separately for each of the four groups (tracking/randomization policy and dormitory/non-dormitory status). Second, I calculate the second difference between the fitted values at each percentile of the HSGPA distribution. Third, I construct pointwise 95 percent confidence intervals from a percentile bootstrap with 1,000 iterations, clustering at the dormitory-year level and stratifying by assignment policy and dormitory status. I calculate the second difference within each bootstrap iteration, which allows for any covariance between the local linear estimates for the four samples. The dashed line shows the effect of tracking on mean peer group composition, discussed in the note below Figure 1.

0.07; p -value of difference 0.026). However, above- and below-median students experience “treatments” of similar magnitude: they have residential peers with HSGPAs on average 0.24 standard deviations higher and 0.25 standard deviations lower under tracking. This is not consistent with a linear response to changes in mean peer quality.¹³ Either low-scoring students are more sensitive to changes in their mean peer group composition or GPA depends on some measure of peer quality other than mean HSGPA.

The near-zero treatment effect on above-median students is perhaps surprising. Aggregating all above-median students together may hide positive effects on very high-scoring students. I therefore estimate treatment effects throughout the distribution of HSGPA. Figure 3 shows that tracking reduces GPA through more than half of the distribution. The negative effects in the left tail are considerably larger than the positive effects in the right tail, though they are not statistically different. Figure 3 also shows the change in mean peer HSGPA (from Figure 1). I reject equality of the treatment effects on GPA and changes in peer HSGPA in the right but not the left

¹³ If student GPA is a linear function of mean peer HSGPA, then the ratio $\Delta GPA / \Delta HSGPA$ should be constant. In particular, this ratio should be equal for above- and below-median students. I reject this hypothesis with a cluster bootstrap p -value of 0.070. This result further motivates my analysis of nonlinear peer effects in Section V.

tail. These results reinforce the finding that low-scoring students are substantially more sensitive to changes in peer group composition than high-scoring students. Tracking may have a small positive effect on students in the top quartile but this effect is imprecisely estimated.

There is stronger evidence of heterogeneity across HSGPA than demographic covariates. Treatment effects are larger on black than white students: -0.20 versus -0.11 standard deviations. However, this difference is not significant ($p = 0.488$) and is smaller after conditioning on HSGPA. I also estimate a quadruple-differences model allowing the effect of tracking to differ across four race/academic subgroups (black/white \times above-/below-median HSGPA). The point estimates show that tracking affects below-median students more than above-median students within each race group and affects black students more than white students within each HSGPA group. However, these differences are very small and statistically insignificant. Treatment effects do not differ by sex: tracking lowers female and male GPAs by 0.14 and 0.13 standard deviations, respectively (p -value of difference 0.931). I conclude that prior academic performance is the primary dimension of treatment effect heterogeneity.

I estimate quantile treatment effects of tracking on the tracked students using a nonlinear difference-in-differences model, which shows how tracking changes the full GPA distribution (Athey and Imbens 2006). I discuss the model and results in detail in online appendices C and D, respectively. The quantile treatment effects are large and negative in the bottom quintile (0.2–1.1 standard deviations), small and negative for most of the distribution (≤ 0.2 standard deviations), and small and positive in the top ventile (≤ 0.2 standard deviations). The negative effects in the bottom two quintiles are significantly different to zero. This reinforces the conclusion that the negative average effect of tracking is driven by large negative effects on students with low academic performance, whether that performance is measured in terms of university GPA or HSGPA.

I also estimate differences in measures of dispersion between the observed and GPA counterfactual distributions. The literature on academic tracking emphasizes inequality concerns (Betts 2011) and the effect of tracking on outcome dispersion is sometimes obvious from quantile treatment effects. However, this is the first study of which I am aware to quantify the effect of tracking on outcome dispersion. The standard deviation, interquartile range, and interdecile range of GPA are respectively 18, 13, and 18 percent higher under tracking.

IV. Effects of Random Variation in Dormitory Composition

The principal research design uses cross-policy variation by comparing tracked and randomly assigned dormitory students. My second research design uses cross-dormitory variation in peer group composition induced by random assignment. I first use a standard test to confirm the presence of residential peer effects, providing additional evidence that the cross-policy results are not driven by confounding factors. I then document differences in dormitory-level peer effects within and between demographic and academic subgroups, providing some information about mechanisms.

I first estimate the standard linear-in-means peer effects model (Manski 1993):

$$(2) \quad GPA_{id} = \alpha_0 + HSGPA_{id} \cdot \alpha_1 + \overline{HSGPA}_d \cdot \alpha_2 + \mathbf{X}_{id} \cdot \boldsymbol{\alpha} + \mu_d + \epsilon_{id},$$

where $HSGPA_{id}$ and \overline{HSGPA}_d are individual and mean dormitory high school graduation test scores, \mathbf{X}_{id} is a vector of student demographic covariates, and μ_d is a dormitory fixed effect. The parameter of interest, α_2 , measures the average gain in GPA from a one standard deviation increase in the mean HSGPA of one's residential peers. I interpret α_2 as a treatment effect that may combine multiple peer effects mechanisms. For example, peer effects may operate through the GPA production technology (peers with higher HSGPA have higher ability and hence raise the marginal product of students' own effort) or through behavioral responses (peers with higher HSGPA have higher propensity to study and hence lower the marginal disutility of students' own study effort). These mechanisms are not identified without specifying a full model of student behavior and GPA production, which I do not attempt here.

Random dormitory assignment ensures that \overline{HSGPA}_d is uncorrelated with individual students' unobserved characteristics so α_2 can be consistently estimated by least squares.¹⁴ However, random assignment also means that mean HSGPA is equal in each dormitory in expectation. Note that α_2 is identified using sample variation in scores across dormitories due to finite numbers of students in each dormitory (Angrist 2014). This variation is relatively low: the range and standard deviation of dormitory means are respectively 11 and 18 percent of the range and standard deviation of individual scores (Table 2).

I estimate equation (2) using the sample of all dormitory students in the random assignment period and report the results in Table 4. I find that $\hat{\alpha}_2 = 0.22$. This means that a one standard deviation increase in \overline{HSGPA} , corresponding to a 0.18 standard deviation increase of HSGPA in the individual distribution, increases college GPA by 0.04 standard deviations. Moving a student from the dormitory with the lowest observed mean high school graduation test score to the highest would increase her GPA by 0.19 standard deviations. This effect size is at the sixty-fifth percentile of the peer effects estimates in papers reviewed by Sacerdote (2011), though some of the larger estimates in those papers are imprecisely estimated. This effect size is larger than most papers studying peer effects in higher education, perhaps because I measure peer effects using scores on a content-based high school graduation test and much of the prior literature measures peer effects using ACT or SAT scores. Stinebrickner and Stinebrickner (2006) compare peer effects from high school GPA and SAT scores, and find the former are larger. They argue this is because study time is an important mechanism for peer effects and GPA or content-based test scores are more strongly associated with peers' study behavior than SAT scores. However, α_2 is fairly imprecisely estimated with 95 percent confidence interval

¹⁴The observed dormitory assignments are consistent with randomization. I regress each baseline covariate on a vector of dormitory indicators and test the hypothesis that the coefficients on the dormitory indicators are all equal. The bootstrap p -values for this test are 0.748 for HSGPA, 0.825 for black, 0.883 for white, 0.962 for other races, 0.886 for English-speaking, and 0.879 for international. I also conduct a joint test for equality of all six covariates across all dormitories and fail to reject equality ($p = 0.980$).

TABLE 4—PEER EFFECTS ON GPA FROM RANDOM ASSIGNMENT TO DORMITORIES

	(1)	(2)	(3)	(4)	(5)	(6)
Own HSGPA	0.332 (0.015)	0.373 (0.021)	0.371 (0.021)	0.328 (0.016)	0.321 (0.018)	0.331 (0.018)
Own HSGPA squared		0.143 (0.014)	0.141 (0.014)			
Mean dormitory HSGPA	0.216 (0.112)	0.213 (0.108)	0.263 (0.163)			
Mean dormitory HSGPA squared		0.303 (0.212)	−0.052 (0.279)			
Own × mean dormitory HSGPA		−0.126 (0.072)	−0.124 (0.072)			
Mean dormitory HSGPA for own race				0.166 (0.077)		
Mean dormitory HSGPA for other races				−0.029 (0.089)		
Mean dormitory HSGPA for own program					0.102 (0.038)	0.058 (0.056)
Mean dormitory HSGPA for other programs					0.180 (0.073)	0.168 (0.082)
<i>p</i> -value of test against equivalent linear model		0.000	0.000			
<i>p</i> -value for equal race effects				0.071		
<i>p</i> -value for equal program effects					0.205	0.258
Student covariates	×	×	×	×	×	×
Dormitory fixed effects	×		×	×	×	×
Program fixed effects						×
Adjusted R^2	0.248	0.270	0.277	0.248	0.249	0.261
Number of students	3,048	3,048	3,048	3,048	3,048	3,048
Number of dormitory-year clusters	28	28	28	28	28	28

Notes: Table 4 reports results from estimating the linear-in-means peer effects model in equation (2) (column 1), the quadratic-in-means peer effects model in equation (4) (columns 2–3), and the linear-in-subgroup-means model in equation (3) (columns 4–6). All columns control for students' sex, language, nationality, and race. All columns except 2 include dormitory fixed effects and column 6 includes program of study fixed effects. Estimates of the linear-in-means and linear-in-subgroup-means models are not sensitive to including student covariates or dormitory fixed effects. Estimates of the quadratic-in-means models are sensitive to including dormitory fixed effects (contrast columns 2 and 3) but not to including student covariates. The sample is all dormitory students in the random assignment period with non-missing HSGPA. Standard errors in parentheses are from 1000 bootstrap iterations clustering at the dormitory-year level.

$= [-0.00, 0.44]$ so the magnitude should be interpreted with caution. This may reflect the limited variation in \overline{HSGPA}_d .

The linear-in-means model can be augmented to allow the effect of residential peers to vary within and across sub-dormitory groups. Specifically, I explore within- and across-race peer effects by estimating

$$\begin{aligned}
 (3) \quad GPA_{ird} = & \alpha_0 + HSGPA_{ird} \cdot \beta_1 + \overline{HSGPA}_{rd} \cdot \beta_2 \\
 & + \overline{HSGPA}_{-rd} \cdot \beta_3 + \mathbf{X}_{ird} \cdot \boldsymbol{\beta} + \mu_d + \epsilon_{ird}.
 \end{aligned}$$

For student i of race r in dormitory d , \overline{HSGPA}_{rd} and \overline{HSGPA}_{-rd} denote the mean high school graduation test scores for other students in dormitory d of, respectively, race r and all other race groups. Here, $\hat{\beta}_2$ and $\hat{\beta}_3$ equal 0.16 and -0.03 , respectively

(Table 4, column 4). The large and marginally significant difference ($p = 0.071$) shows that peer effects operate primarily within race groups. I interpret this as evidence that spatial proximity does not automatically generate peer effects. Instead, peer groups are formed through a combination of spatial proximity and proximity along other dimensions such as race, which remains highly salient in South Africa.¹⁵ This indicates that interaction patterns by students mediate residential peer effects; in which case, estimates may not be policy-invariant.

I also explore the content of the interaction patterns that generate residential peer effects by estimating equation (3) using program of study groups instead of race groups.¹⁶ The estimated within- and across-program peer effects are respectively 0.10 and 0.18 (standard errors 0.04 and 0.07). These results show that within-program peer effects are not systematically stronger than cross-program peer effects.¹⁷ This result is not consistent with peer effects being driven by direct academic collaboration such as joint work on problem sets or studying together for the same examinations. My data cannot directly identify the mechanisms that generate the observed peer effects. But informal interviews with students at the university suggest two mechanisms through which peer effects operate: time allocation over study and leisure activities, and transfers of tacit knowledge such as study skills, norms about how to interact with instructors, and strategies for navigating academic bureaucracy. This is consistent with prior findings of strong peer effects on study time (Stinebrickner and Stinebrickner 2006) and social activities (Duncan et al. 2005).

Combining the race- and program-level peer effects results indicates that spatial proximity alone does not generate peer effects. Some direct interaction is also necessary and is more likely when students are also socially proximate. However, the relevant form of the interaction is not direct academic collaboration.

V. Can Peer Effects Account for the Effects of Tracking?

The results in Section IV show that peer effects are quantitatively important in this setting. I now explore whether these peer effects can account for the estimated treatment effect of tracking. The linear-in-means model restricts average GPA to be invariant to any group reassignment: moving a strong student to a new group has equal but oppositely signed effects on her old and new peers' average GPA. So the linear-in-means model cannot generate the negative average treatment effect of

¹⁵I find a similar result using language instead of race to define subgroups. This pattern could also arise if students sort into homogeneous racial or linguistic geographic within-dormitory groups by choosing nearby rooms within their assigned dormitories. As I do not observe roommate assignments, I cannot test this mechanism.

¹⁶I divide students into six programs, corresponding to the faculty in which they are registered: commerce, engineering, health sciences, humanities and social sciences, law, and science. Some students take courses exclusively within their faculty (engineering, health sciences) while students in other faculties take courses across faculties.

¹⁷This result is robust to conditioning on program fixed effects (column 6 of Table 4). I obtain similar results using course-specific grades as the outcome and allowing residential peer effects to differ at the course level. For example, I estimate equations (2) and (3) with introductory microeconomics grades as an outcome. I find that there are strong peer effects on grades in this course ($\hat{\alpha}_2 = 0.34$, standard error 0.14) but they are not driven primarily by other students in the same course ($\hat{\beta}_2 = -0.01$ and $\hat{\beta}_3 = 0.28$, standard errors 0.14 and 0.15). These, and other course-level regression results, are consistent with the main results but the smaller sample sizes yield less precise estimates that are somewhat sensitive to the inclusion of covariates.

tracking estimated using cross-policy variation. I therefore estimate a more general production function that permits nonlinear peer effects:

$$\begin{aligned}
 (4) \quad GPA_{id} = & \gamma_0 + HSGPA_{id} \cdot \gamma_1 + \overline{HSGPA}_d \cdot \gamma_2 + HSGPA_{id}^2 \cdot \gamma_{11} \\
 & + \overline{HSGPA}_d^2 \cdot \gamma_{22} + HSGPA_{id} \times \overline{HSGPA}_d \cdot \gamma_{12} \\
 & + \mathbf{X}_{id} \cdot \gamma + \mu_d + \varepsilon_{id}.
 \end{aligned}$$

This is a parsimonious specification that permits average outcomes to vary over assignment policies and aligns with theoretical models of matching markets and neighborhood segregation.¹⁸ As with model (2), I do not attach a structural interpretation to this model, so the parameters may combine multiple peer effects mechanisms. The key parameters of the model are γ_{12} and γ_{22} . Parameter γ_{12} indicates whether own and mean peer HSGPA are technological complements or substitutes in GPA production. If $\gamma_{12} < 0$, the GPA gain from high-scoring peers is larger for low-scoring students.¹⁹ In classic binary matching models, this parameter determines whether positive or negative assortative matching is output-maximizing (Becker 1973). In matching models with more than two agents, γ_{12} is not sufficient to characterize the output-maximizing set of matches. Parameter γ_{22} indicates whether GPA is a concave or convex function of mean peer HSGPA. If $\gamma_{22} < 0$, total output is higher when mean test scores are identical in all groups. If $\gamma_{22} > 0$, total output is higher when some groups have very high means and some groups have very low means. This parameter has received relatively little attention in the peer effects literature but features prominently in some models of neighborhood effects (Bénabou 1996; Graham, Imbens, and Ridder 2010). Tracking will deliver lower total GPA than random assignment if both parameters are negative and vice versa. If the parameters have different signs, the average effect of tracking is ambiguous.²⁰

¹⁸This specification assumes that GPA depends on peer characteristics only through the dormitory mean. I consider two alternative specifications that deliver similar results. First, the estimated coefficients from equations (2), (3), and (4) are similar but less precise if I replace dormitory-year means with medians. Second, the estimated coefficients from all models are similar if I control for the dormitory-year standard deviation of HSGPA. The coefficient on the standard deviation when this is included in model (2) is negative (−0.12 to −0.16) but not statistically significant. This specification is motivated by previous work that finds a relationship between individual outcomes and the variance of peer characteristics (Sacerdote 2011). See Carrell, Sacerdote, and West (2013) for an alternative parameterization and Graham (2011) for background discussion.

¹⁹Technological complementarity/substitutability is based on the sign of the cross-partial derivative with respect to attributes exogenous to the theoretical model. HSGPA is determined prior to peer exposure and so fulfills this condition. This does not correspond to strategic complementarity/substitutability, which is based on the sign of the cross-partial derivative with respect to own and peer endogenous inputs, such as effort. I cannot directly test strategic complementarity/substitutability without stronger assumptions. However, the interviews discussed in Section IV suggest a role for study effort or time use as a peer effects mechanism, and HSGPAs are likely to be positively correlated with unobserved study effort. γ_{12} may thus reflect both peer effects from fixed ability and from effort choices, where HSGPA is a proxy for both.

²⁰To derive this result, note that $E[\overline{HSGPA}_d | HSGPA_{id}] = HSGPA_{id}$ under tracking and $E[HSGPA_{id}]$ under random assignment. Hence, $E[HSGPA_{id} \overline{HSGPA}_d]$ and $E[HSGPA_{id}^2]$ both equal $E[HSGPA_{id}^2]$ under tracking and $E[HSGPA_{id}]^2$ under random assignment. Plugging these results into equation (4) for each assignment policy yields $E[GPA_{id} | \text{Tracking}] - E[GPA_{id} | \text{Randomization}] = \sigma_{HSGPA}^2(\gamma_{22} + \gamma_{12})$. This simple demonstration assumes an infinite number of students and dormitories but a similar result holds with a finite population.

Estimates from equation (4) are shown in Table 4, columns 2 and 3. Note that $\hat{\gamma}_{12}$ is negative and marginally statistically significant across all specifications. The point estimate of -0.12 (standard error 0.07) implies the GPA gain from an increase in mean peer HSGPA is 0.2 standard deviations larger for students at the twenty-fifth percentile of the individual HSGPA distribution than students at the seventy-fifth percentile. This is consistent with the Section III result that low-scoring students are hurt more by tracking than high-scoring students are helped. However, the sign of $\hat{\gamma}_{22}$ flips from positive to negative with the inclusion of dormitory fixed effects. It is thus unclear whether GPA is concave or convex in mean peer HSGPA.

I draw three conclusions from these results. First, there is clear evidence of nonlinear peer effects from the cross-dormitory variation generated under random assignment. Wald tests, Akaike information criteria, and Bayesian information criteria prefer quadratic models to linear models, with or without conditioning on student demographics and dormitory fixed effects. The in-sample mean squared prediction error is also lower for each quadratic model than the corresponding linear model. Second, the results from the fixed effects specification (column 3 of Table 4) are qualitatively consistent with the negative average treatment effect of tracking. Third, however, peer effects estimates using randomly induced cross-dormitory variation are sensitive to the support of the data. Using dormitory fixed effects reduces the standard deviation of \overline{HSGPA}_d from 0.18 to 0.11 . This leads to different conclusions about the curvature of the GPA model in columns 2 and 3 of Table 4.

Can the peer effects estimated from equation (4) account for the observed treatment effects of tracking? I answer this question by comparing observed GPA^{Track} to predicted GPA using the coefficient estimates reported in Table 4, column 3: $\widehat{GPA}^{Track} = \mathbf{W}^{Track} \cdot \hat{\theta}^{Random}$, where \mathbf{W} and θ are respectively the stacked regressors and coefficients from equation (4). The mean difference between the observed and predicted values is a consistent estimator of the average treatment effect of tracking on the tracked students (ATET) if the regression model is correctly specified (Fortin, Lemiux, and Firpo 2011).²¹ Comparing this estimate of the ATET to the difference-in-differences estimate evaluates how well peer effects models can explain the effects of tracking. The ATET from this prediction exercise is -0.08 standard deviations (standard error 0.28), compared to the difference-in-differences estimate of -0.12 . The treatment effects for students with below- and above-median HSGPA are -0.13 and -0.03 standard deviations, respectively (standard errors 0.31 and 0.27). The prediction exercise is accurate for above-median students but understates the negative effect on below-median students, which is -0.24 in the difference-in-differences framework. The results are similar using the coefficient estimates from the quadratic peer effects model without demographic controls or dormitory fixed effects.

²¹ The observed and predicted values can also differ if the mean values of any unobserved determinants of standardized GPA differ between the two periods. Any mean time trends that are common to the dormitory and non-dormitory students are already accounted for because GPA in each period is standardized with reference to the non-dormitory students.

The observed and predicted effects of tracking may differ for three interrelated reasons. First, tracking may affect GPAs through some alternative mechanism, besides peer effects. Second, model (4) may be misspecified because it does not account for behavioral changes in peer interactions induced by tracking. For example, within-dormitory friendships will be more common under tracking if students have a preference for academically homogenous social groups. Third, model (4) may be misspecified because it does not extrapolate to high- and low-track dormitories where peer effects might operate differently.

I explore the relative importance of the extrapolation explanation, versus the alternative mechanism and behavioral change explanations, by comparing prediction error within and outside the support of the dormitory means observed under random assignment. First, I estimate the quadratic peer effects model on the sample of randomly assigned dormitory students. Second, I use the coefficient estimates to predict GPA for tracked dormitory students and calculate the squared prediction error for each student. Third, I calculate the mean squared prediction error for three subgroups of tracked students: those in dormitories with mean peer HSGPA below, within, and above the range observed under random assignment. I refer to these as low-, mid-, and high-track dormitories, respectively. Fourth, I calculate squared prediction error for each randomly assigned dormitory student using leave-one-out estimation. The mean squared prediction error for randomly assigned dormitory students (0.440 with standard error 0.026) provides a benchmark for the “normal” prediction error without any policy change.²² The mean squared prediction error for tracked students in mid-track dormitories (0.612 with standard error 0.051) adds prediction error due to alternative mechanisms and behavioral change. The mean squared prediction error for tracked students in low- and high-track dormitories (1.196 with standard error 2.141) adds prediction error due to extrapolation.

These results show that both extrapolation and at least one of the alternative mechanism and behavioral change explanations are quantitatively important, with extrapolation error being more important. Mean squared prediction error is 39 percent higher for tracked than randomly assigned dormitory students within the original support of dormitory means. Mean squared prediction error is 172 percent higher outside the original support for dormitory means, verifying that the model estimated under random assignment is less accurate when extrapolating outside the support of the data. Figure 4 shows that the higher prediction error from extrapolation is due mainly to low-track dormitories, rather than high-track dormitories. However, outside-support mean squared prediction error is very imprecisely estimated so the latter result should be interpreted with caution. The relative importance of extrapolation error decreases if the one non-randomly assigned dormitory in the random assignment period is included in the sample; excluding this dormitory from the random assignment estimation sample omits this fixed effect from the prediction model in the tracking period, increasing extrapolation error. But even when this dormitory is included, prediction error is 55 percent higher outside the original support for dormitory means.

²² Standard errors are estimated by bootstrapping the entire prediction process: estimating the quadratic peer effects model using randomly assigned students and calculating squared prediction error for each randomly assigned and tracked student. The bootstrap algorithm resamples dormitory-year clusters over 1,000 iterations.

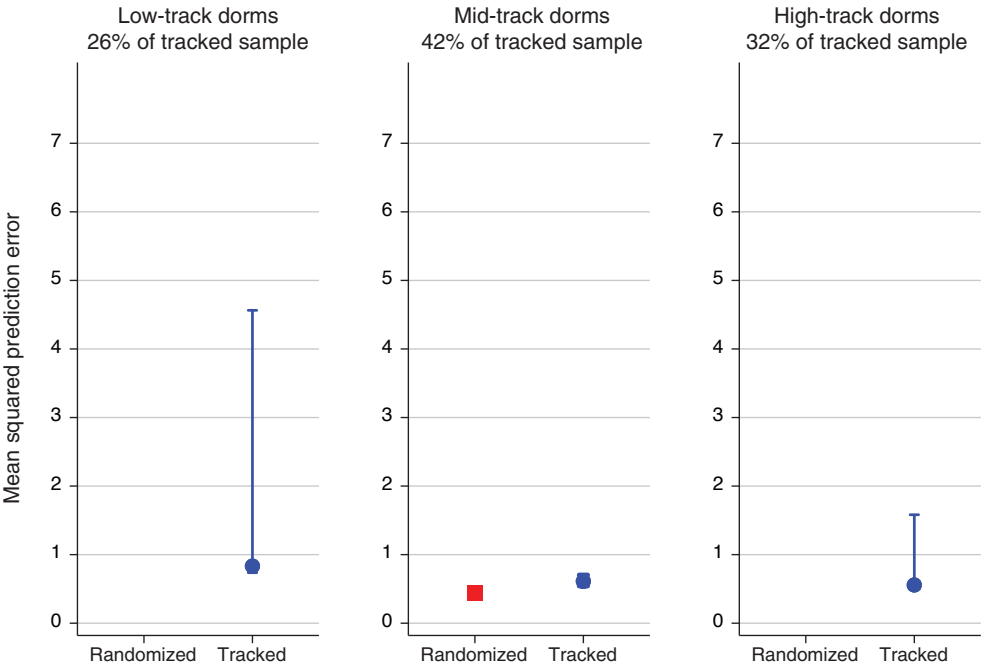


FIGURE 4. MEAN SQUARED PREDICTION ERROR BY DORMITORY TRACK, FOR TRACKED AND RANDOMLY ASSIGNED DORMITORY STUDENTS

Notes: Circles show mean squared prediction error for tracked students. Squares show mean squared prediction error for randomly assigned students. Squared prediction error is the difference between observed GPA and predicted GPA using the coefficient estimates from estimating the quadratic peer effects model (equation (4)) on the sample of randomly assigned dormitory students. The prediction error for randomly assigned dormitory students is derived from leave-one-out models so all errors are out-of-sample predictions. The 90 percent confidence intervals are from a percentile bootstrap algorithm that resamples dormitory-year observations, estimates the quadratic peer effects model, and estimates individual prediction errors 1,000 times. Low-, mid-, and high-track dormitories are defined as those with mean high school graduation test scores below, within, and above the observed range under random assignment.

I do not observe data that can measure differences in the way peers interact under tracking and random assignment. But my results, and the existing peer effects literature, suggest some possible candidates. I show in Section IV that peer effects operate primarily within race (and language) groups. This implies that social interactions mediate peer effects based on group assignment, consistent with Carrell, Sacerdote, and West (2013). The peer effects model in equation (4) does not allow for this behavior, but the prediction error patterns are similar for the linear-in-means model (2), the quadratic-in-means model (4), a quadratic-in-own-race-means model integrating models (3) and (4), and models that depend on both the mean and standard deviation of peer HSGPA for either dormitory or dormitory-race groups. But none of these models necessarily captures behavioral responses such as higher effort; Fruehwirth (2014) shows that effort adjustment can make it difficult to use estimated peer effects for prediction.

Similarly, I do not observe data that can measure effects of tracking that operate through mechanisms other than peer effects. I discuss and reject several mechanisms in online Appendix B: discriminatory grading, correlated policy changes, time-varying

dormitory characteristics, or interactions between time-invariant dormitory characteristics and student characteristics. The larger-than-predicted negative effect of tracking on low-scoring students may occur because students are discouraged or suffer lower confidence when they learn they have been assigned to a low-track dormitory. This is consistent with the stereotype threat literature in social psychology (Steele and Aronson 1995). However, dormitory assignment probably provided students with limited information about their academic rank because high school graduation test results are published in newspapers and the university publishes the minimum HSGPA required for admission to specific programs of study. I conclude that peer effects account for much of the negative effect of tracking, unmodeled differences in peer interaction between tracking and random assignment may also be important, and stereotype threat may exacerbate the negative effects on low-scoring students.

VI. Conclusion

This paper describes the effect of tracking relative to random dormitory assignment on student GPAs at the University of Cape Town in South Africa. I show that tracking lowered mean GPA and increased GPA inequality. This result occurred because living with high-scoring peers has a large positive effect on low-scoring students' GPAs and little effect on high-scoring students' GPAs. These peer effects arise largely through interaction with own-race peers and the relevant form of interaction does not appear to be direct academic collaboration. I present an extensive set of robustness checks supporting a causal interpretation for these results.

My findings show that different peer group assignment policies can have substantial effects on students' academic outcomes. Academic tracking into residential groups, and perhaps other noninstructional groups, may generate a substantially worse distribution of academic outcomes than random assignment. My findings suggest that policymakers can change the distribution of students' academic performance by rearranging the groups in which these students interact without changing the marginal distribution of inputs into the education production function. This is attractive in any setting but particularly in resource-constrained developing countries. While the external validity of any result is always questionable, my findings may be particularly relevant to universities serving a diverse student body that includes both high performing and academically underprepared students. This is particularly relevant to selective universities with active affirmative action programs.

My results do not permit a welfare ranking of the two policies. Tracking clearly harms low-scoring students but some (imprecise) results suggest a positive effect on high-scoring students. Changing the assignment policy may thus entail a transfer from one group of students to another and, as academic outputs are not directly tradeable, Pareto-ranking the two policies may not be possible. Non-measured student outcomes may also be affected by different group assignment policies. For example, high-scoring students' GPAs may be unaffected by tracking because the rise in their peers' academic proficiency induces them to substitute time away from studying toward leisure. In future work, I plan to study the long-term effects of tracking on graduation rates, time-to-degree, and labor market outcomes and may use survey evidence on networks and time use. This will permit a more comprehensive

evaluation of the two group assignment policies. One simple revealed preference measure of student welfare under the two policies is the proportion of dormitory students who stay in their dormitory for a second year. Tracking reduces this rate for students with above- and below-median high school test scores by 0.4 percentage points and 6.7 percentage points, respectively (standard errors 3.5 and 3.8). Low-scoring dormitory students may thus be aware of the negative effect of tracking and respond by leaving the dormitory system early.

Despite these provisos, my findings shed light on the importance of peer group assignment policies. I provide what appears to be the first cleanly identified evidence on the effects of noninstructional tracking. This complements the small literature that cleanly identifies the effect of instructional tracking. Booij, Leuven, and Oosterbeek (2017) and Duflo, Dupas, and Kremer (2011) find that students with both high and low prior academic performance obtain better outcomes in tracked rather than randomly assigned classrooms.²³ They attribute this respectively to better group interaction and more targeted instruction in tracked classrooms. Residential and classroom tracking can generate respectively negative and positive results if peer effects operate differently outside and within classrooms. For example, classroom peer groups may be more relevant for academic collaboration, while residential peer groups may be more relevant for study behavior and time use. Low-track classrooms may facilitate targeted instruction and productive academic collaboration, while low-track dormitories facilitate distracting behavior and direct time away from studying. This is consistent with my finding that peer effects are not stronger when dormmates are in the same program of study. This is also consistent with the general pattern that residential peer effects in universities are larger for non-academic outcomes (e.g., drinking) than academic outcomes (Sacerdote 2011). If this hypothesis is correct, my results may be important for the study of neighborhood effects on human capital formation.

The examination of peer effects under random assignment also points to fruitful avenues for future research. As in Carrell, Sacerdote, and West (2013), peer effects estimated under random assignment do not exactly predict the effects of a new assignment policy and residential peer effects appear to be mediated by students' patterns of interaction. Combining peer effects estimated under different group assignment policies with detailed data on social interactions and explicit models of network formation may provide additional insights.

REFERENCES

- Abdulkadiroğlu, Atila, Joshua Angrist, and Parag Pathak.** 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82 (1): 137–96.
- Ajayi, Kehinde.** 2016. "Student Performance and the Effects of School Quality versus School Fit." Paper presented at UNU-WIDER Human Capital and Growth Conference, Helsinki, Finland, June 7.

²³ Booij, Leuven, and Oosterbeek (2017) use stratified random assignment to generate classrooms that look like low-, mid-, and high-track classrooms. This strategy ensures that they observe "tracking-like" dormitories under random assignment. However, their prediction may be problematic if student behavior is systematically different under different group assignment policies, conditional on peer group composition. Duflo, Dupas, and Kremer (2011) directly compare tracked and randomly assigned classrooms in different schools, which avoids this problem.

- Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber.** 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1): 151–84.
- Angrist, Joshua D.** 2014. "The Perils of Peer Effects." *Labour Economics* 30: 98–108.
- Angrist, Joshua D., Sarah R. Cohodes, Susan M. Dynarski, Parag A. Pathak, and Christopher R. Walters.** 2016. "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice." *Journal of Labor Economics* 34 (2): 275–318.
- Arnott, Richard, and John Rowse.** 1987. "Peer group effects and educational attainment." *Journal of Public Economics* 32 (3): 287–305.
- Athey, Susan, and Guido W. Imbens.** 2006. "Identification and Inference in Nonlinear Difference-in-Differences Models." *Econometrica* 74 (2): 431–97.
- Becker, Gary S.** 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4): 813–46.
- Bénabou, Roland.** 1996. "Equity and Efficiency in Human Capital Investment: The Local Connection." *Review of Economic Studies* 63 (2): 237–64.
- Betts, Julian R.** 2011. "The Economics of Tracking in Education." In *Handbook of the Economics of Education*, Vol. 3, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 341–81. Amsterdam: North-Holland.
- Bhattacharya, Debopam.** 2009. "Inferring Optimal Peer Assignment from Experimental Data." *Journal of the American Statistical Association* 104 (486): 486–500.
- Booij, Adam S., Edwin Leuven, and Hessel Oosterbeek.** 2017. "Ability Peer Effects in University: Evidence from a Randomized Experiment." *Review of Economic Studies* 84 (2): 547–78.
- Burke, Mary A., and Tim R. Sass.** 2013. "Classroom Peer Effects and Student Achievement." *Journal of Labor Economics* 31 (1): 51–82.
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller.** 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90 (3): 414–27.
- Carrell, Scott E., Frederick V. Malmstrom, and James E. West.** 2008. "Peer Effects in Academic Cheating." *Journal of Human Resources* 43 (1): 173–207.
- Carrell, Scott E., Bruce I. Sacerdote, and James E. West.** 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81 (3): 855–82.
- De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli.** 2010. "Identification of Social Interactions through Partially Overlapping Peer Groups." *American Economic Journal: Applied Economics* 2 (2): 241–75.
- Ding, Weili, and Steven F. Lehrer.** 2007. "Do Peers Affect Student Achievement in China's Secondary Schools?" *Review of Economics and Statistics* 89 (2): 300–312.
- Dufo, Esther, Pascaline Dupas, and Michael Kremer.** 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101 (5): 1739–74.
- Duncan, Greg J., Johanne Boisjoly, Michael Kremer, Dan M. Levy, and Jacque Eccles.** 2005. "Peer Effects in Drug Use and Sex among College Students." *Journal of Abnormal Child Psychology* 33 (3): 375–85.
- Epple, Dennis, and Richard E. Romano.** 1998. "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects." *American Economic Review* 88 (1): 33–62.
- Evans, David K., and Anna Popova.** 2016. "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews." *World Bank Research Observer* 31 (2): 242–70.
- Feld, Jan, and Ulf Zölitz.** 2017. "Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects." *Journal of Labor Economics* 35 (2): 387–428.
- Fortin, Nicole, Thomas Lemieux, and Sergio Firpo.** 2011. "Decomposition Methods in Economics." In *Handbook of Labor Economics*, Vol. 4A, edited by Orley Ashenfelter and David Card, 1–102. Amsterdam: North-Holland.
- Fruehwirth, Jane Cooley.** 2013. "Identifying peer achievement spillovers: Implications for desegregation and the achievement gap." *Quantitative Economics* 4 (1): 85–124.
- Fruehwirth, Jane Cooley.** 2014. "Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box." *Review of Economics and Statistics* 96 (3): 514–23.
- Garlick, Robert.** 2018. "Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment: Dataset." *American Economic Journal: Applied Economics*. <https://doi.org/10.1257/app.20160626>.
- Graham, Bryan S.** 2011. "Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers." In *Handbook of Social Economics*, Vol. 1, edited by Jesse Benhabib, Alberto Bisin, and Matthew O. Jackson, 965–1052. Amsterdam: North-Holland.

- Graham, Bryan S., Guido W. Imbens, and Geert Ridder. 2010. "Measuring the Effects of Segregation in the Presence of Social Spillovers: A Nonparametric Approach." National Bureau of Economic Research (NBER) Working Paper 16499.
- Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2009. "New Evidence about *Brown v. Board of Education*: The Complex Effects of School Racial Composition on Achievement." *Journal of Labor Economics* 27 (3): 349–83.
- Higher Education South Africa. 2009. *National Benchmark Tests Project and Standards for National Examination and Assessment Systems: Department of Higher Education*. Basic Education Portfolio Committee. Cape Town, South Africa, August.
- Hoel, Jessica, Jeffrey Parker, and Jon Rivenburg. 2004. "Peer Effects: Do First-Year Classmates, Roommates, and Dormmates Affect Students Academic Success?" https://www.reed.edu/economics/parker/Peer_Effects_HEDS.pdf.
- Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." National Bureau of Economic Research (NBER) Working Paper 7867.
- Hoxby, Caroline M., and Gretchen Weingarth. 2006. "Taking Race out of the Equation: School Reassignment and the Structure of Peer Effects." Unpublished.
- Hsieh, Chang-Tai, and Miguel Urquiola. 2006. "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program." *Journal of Public Economics* 90 (8–9): 1477–1503.
- Hurder, Stephanie. 2012. "Evaluating Econometric Models of Peer Effects with Experimental Data." Unpublished.
- Imberman, Scott A., Adriana D. Kugler, and Bruce I. Sacerdote. 2012. "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees." *American Economic Review* 102 (5): 2048–82.
- Jain, Tarun, and Mudit Kapoor. 2015. "The Impact of Study Groups and Roommates on Academic Performance." *Review of Economics and Statistics* (97) 1: 44–54.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.
- Lavy, Victor, Olmo Silva, and Felix Weinhardt. 2012. "The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools." *Journal of Labor Economics* 30 (2): 367–414.
- Lucas, Adrienne M., and Isaac M. Mbiti. 2014. "Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya." *American Economic Journal: Applied Economics* (6) 3: 234–63.
- Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60 (3): 531–42.
- Marmaros, David, and Bruce Sacerdote. 2002. "Peer and social networks in job search." *European Economic Review* 46 (4–5): 870–79.
- Pop-Eleches, Christian, and Miguel Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses." *American Economic Review* 103 (4): 1289–1324.
- Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *Quarterly Journal of Economics* 116 (2): 681–704.
- Sacerdote, Bruce. 2011. "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In *Handbook of the Economics of Education*, Vol. 3, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 249–77. Amsterdam: North-Holland.
- Steele, Claude M., and Joshua Aronson. 1995. "Stereotype threat and the intellectual test performance of African Americans." *Journal of Personality and Social Psychology* 69 (5): 797–811.
- 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.