ISOLATING PEER EFFECTS IN THE RETURNS TO COLLEGE SELECTIVITY

NICOLÁS DE ROUX*
EVAN RIEHL+

SEPTEMBER 2019

ABSTRACT. This paper asks how a student's classmates affect her returns to college. We exploit a "tracking" admission system at a selective Colombian university that led to large differences in mean classmate ability for students in the same programs. In a regression discontinuity design, we find that students in higher-ability classes were more likely to fail courses and drop out, and had lower earnings one decade later. Testable predictions from a human capital model with peer externalities show that individuals learned less in more able classrooms. Our findings suggest that exposure to higher-ability college peers can harm an individual's career trajectory.

*de Roux: Department of Economics, Universidad de los Andes, Calle 19A No 1-37 Este, W 823, Bogotá, Colombia (email: nicolas.de.roux@uniandes.edu.co).
+Riehl (corresponding author): Department of Economics, Cornell University, 266 Ives Hall, Ithaca, NY 14853, United States (email: eriehl@cornell.edu).

Do disadvantaged students benefit from attending top colleges? There is compelling evidence that admission to a selective college raises graduation rates and earnings (Hoekstra, 2009; Cohodes and Goodman, 2014), and in some cases estimated returns are even larger for low-income or minority students (Dale and Krueger, 2002). Since disadvantaged students often have lower achievement levels, however, other work argues that they may be "overmatched" at top colleges, and thus more likely to drop out or switch majors (Arcidiacono et al., 2016). The debate on this question persists because it is hard to separate *quality* effects of selective colleges—such as greater resources—from *match* effects that arise when students are less prepared than their peers (Arcidiacono and Lovenheim, 2016). Students who choose to attend a more selective college typically experience both effects simultaneously.

This paper exploits a natural experiment to isolate how the match between a student and her college peers affects course grades, graduation rates, and earnings. We use data from a selective public university in Colombia that "tracked" students into different cohorts by admission exam scores. From this we develop a regression discontinuity design that estimates the effects of a large increase in mean classmate ability for students in the same college and major. We find that marginal admits to higher-ability cohorts were more likely to fail first-year courses and drop out of college, and had lower earnings one decade later.

To interpret these effects, we develop a model that embeds classmate externalities in a standard human capital framework. The model's testable implications show that the most compelling explanation for our results is that individuals learned less in more able classes. We provide suggestive evidence that class composition affected both teaching and student effort. In sum, our paper highlights the importance of match effects in college, and it shows that exposure to more able peers can in some cases harm an individual's career trajectory.

Our empirical setting is a large flagship university in the country of Colombia. Students apply to separate majors at the flagship, and admission is based solely on national exam scores. Related work has used similar settings to estimate the earnings return to selective institutions or majors (Hastings et al., 2013; Kirkebøen et al., 2016). In these papers, affected students attend schools with higher-ability peers, but also different resources and courses of study. Thus the estimates reflect both quality and match effects of selective programs.

A unique admissions system at the flagship allows us to isolate the match between a student and her college peers. From 2000–2003, several architecture, business, and engineering programs used admission scores to "track" students into separate fall and spring cohorts. The 60 highest-scoring applicants were admitted to a fall cohort, and the next 60 applicants were admitted to a spring cohort. This affected the mean ability of a student's classmates because admission cohorts take core courses together. Since the flagship is highly selective, 85 percent of applicants accepted their admission offer with no discontinuity at the tracking

threshold. As a result, marginal admits to the fall cohort took first-year courses in which the mean student scored eight percentile points higher on the national admission exam.

We use a regression discontinuity design that compares students admitted to the same program, but who took courses with different mean classmate ability. We collected data on students' admission scores and grades in each flagship course, and match these to administrative records on other colleges attended and formal sector earnings.

Our main finding is that students in more able classes had lower grades, graduation rates, and post-college earnings. Marginal admits to higher-ability cohorts were five percentage points more likely to fail first-year courses, and eight percentage points less likely to complete the program. Most students who dropped out did not enroll in another college program. One decade after applying, these students had 11 percent lower daily earnings than those just below the tracking threshold, consistent with their lower educational attainment.

To understand the mechanisms for these results, we develop a model that illustrates how an individual's classmates can affect her returns to college. For this we augment a standard human capital model (Becker, 1964) with two types of classmate externalities. First, we allow a student's *learning* to vary with mean classmate ability, consistent with the linear-in-means model of Manski (1993). The learning externality can be *a priori* either positive or negative, as it is the net effect of peer interactions (Sacerdote, 2001), teaching responses (Duflo et al., 2011), and other classmate influences on human capital accumulation.

Second, classmates can affect an individual's returns through *information* externalities. We allow the program a student enrolls in to provide information on her unknown ability. A peer externality arises because students have higher *expected* ability when they enroll in more selective programs. This can affect a professor's decision on how to set grades, and thus whether to pass or fail a given student. It can also affect an individual's labor market returns if employers use program identity as a signal of ability (MacLeod and Urquiola, 2015).

In this framework, we show that our results are most consistent with negative learning externalities from more able classmates. In particular, positive learning spillovers and information externalities do not explain why a given student would be more likely to fail in a higher-ability class. If ability is positively related to learning and professors grade benevolently, both of these externalities *reduce* the probability of failing because students have higher expected returns in more able classes. Our model predicts that more higher-ability classmates increase the likelihood of failing only if they reduce a student's learning.

We derive testable implications of our model and show evidence that individuals learned less in more able classes. Marginal admits to high-ability cohorts also received lower grades in

advanced courses, and they had lower earnings even conditional on graduation.[1] Our model shows that such effects arise only if classmates directly affect individual learning. We also show that our results cannot be explained by tracking effects on the timing of enrollment.

Finally, we provide suggestive evidence that learning externalities resulted from both professor and student responses to class composition. The negative effects appear only in programs that used tracking admissions for four years; we find no effects in programs that used tracking for a single year. Thus learning externalities were larger in programs where professors had more time to adjust teaching practices. Further, we find negative effects only for male students. This is consistent with evidence that changes in relative classroom position have a greater impact on effort for men than for women (Murphy and Weinhardt, 2018).

Our paper contributes to work on the returns to college or major selectivity. There is compelling evidence that admission to selective college programs raises graduation rates and earnings (Hoekstra, 2009; Hastings et al., 2013; Goodman et al., 2014; Kirkebøen et al., 2016; Chetty et al., 2017; Canaan and Mouganie, 2018; Hoxby, 2018; Zimmerman, 2019), and in some cases the returns are largest for students from disadvantaged backgrounds (Dale and Krueger, 2002, 2014; Saavedra, 2009; Zimmerman, 2014). Arcidiacono and Lovenheim (2016) summarize this work as demonstrating that there are positive *quality* effects of selective colleges, defined as factors that produce better outcomes for all students.

Other work focuses on the *match* between a student and her college peers. Arcidiacono et al. (2011, 2014, 2016) find that affirmative action can reduce major completion and graduation rates for students whose achievement is far below that of their peers. Riehl (2019) shows that similar effects can arise if college admission exams become less informative. Andrews et al. (2016) and Dillon and Smith (2018) find evidence of a complementarity in earnings between student ability and college quality. These papers suggest that students can be "overmatched" in programs with higher-ability peers, but the results often rely on strong assumptions about unobservable determinants of college choice.[2]

We use the same identification strategy as in work that finds benefits to college selectivity, but find large negative effects in isolating the match component of these returns. If match effects are also important in other settings, our results suggest that selective programs have a large positive value added beyond peer mechanisms (MacLeod et al., 2017). This could stem from higher expenditures (Bound et al., 2010; Deming and Walters, 2018) or access to certain labor markets (MacLeod and Urquiola, 2018). Since our results exclude quality

---

[1] There is an inherent selection issue in analyzing both of these outcomes because marginal admits to higher-ability cohorts were less likely to graduate. We implement a selection correction method in the spirit of Lee (2009) using the fact that drop-out rates are strongly related to students' first-year grades.

[2] Other reduced form work exploits affirmative actions bans but finds no direct evidence of mismatch (Cortes, 2010; Backes, 2012; Hinrichs, 2012). Related research examines the causes of "overmatch" or "undermatch" (Hoxby and Avery, 2013; Smith et al., 2013; Dillon and Smith, 2017; Cortes and Lincove, 2019).

effects, they do not necessarily imply that programs like affirmative action can harm students. But they show that negative effects may be more likely when admission policies lead to large achievement gaps between an individual and her college peers.

Our paper also relates to work on peer effects in college. This work exploits random assignment of students to roommates (Sacerdote, 2001; Zimmerman, 2003; Stinebrickner and Stinebrickner, 2006), social groups (Foster, 2006; Lyle, 2007), military squadrons (Carrell et al., 2009, 2013), and tutorial groups (Booij et al., 2017), and mostly finds either zero or positive effects of higher-ability peers on a student's GPA (Sacerdote, 2011). We find negative effects of more able peers in a less-studied reference group: the college classroom.[3] This is consistent with work showing that a student's relative rank affects schooling outcomes (Elsner et al., 2018; Murphy and Weinhardt, 2018; Ribas et al., 2018).[4] Our empirical design is most similar to Duflo et al. (2011), who find no achievement differences for students near a tracking threshold in Kenyan primary schools. Our results highlight the importance of student drop-out for understanding peer effects in college classrooms.

The paper proceeds as follows. Section 1 develops a model of the returns to college with classmate externalities. Section 2 describes the flagship's tracking admissions and our identification strategy. Section 3 shows how tracking affected peer composition and grades in first-year courses. Section 4 presents effects on graduation and earnings. Section 5 concludes.

1. Model

This section develops a model that shows how classmates can affect a student's returns to college. We define two types of peer externalities—learning and information—and embed them into a standard human capital model. We show how these externalities affect a professor's incentives for which students to pass. Lastly, we derive predictions for peer effects in grades, graduation rates, and earnings under different classmate externalities.

1.1. **Peer learning externalities.** Students arrive at college with heterogeneous ability. We let θ_{it} denote individual i's skill at the start of year t, where $t = 0$ is the first year of college. θ_{i0} is thus a student's skill at the beginning of college. We use the terms "skill" and "ability" interchangeably; both refer to human capital as defined by labor market productivity.

Students accumulate skill by taking courses. For simplicity we assume there is one class per year, and thus we index courses by t. Skill accumulation is given by the recursive equation

$$(1) \qquad \theta_{i,t+1} = \theta_{it} + v_{it},$$

---

[3] There is more evidence on classroom peer effects at the K–12 level (Hoxby, 2000; Hoxby and Weingarth, 2005; Lavy et al., 2012). This work often exploits natural variation in cohort composition, which is less likely to be exogenous in higher education because students have more flexibility in choosing schools and courses.
[4] These papers use identification strategies that isolate the role of student effort. We consider a broader set of mechanisms that matter in college classrooms, including professor behavior and information externalities.

where v_{it} is individual's i skill return from course t. A student's skill at the start of the next year is her previous year's skill plus the value added from her course that year.

We suppose that the skill return to college, v_{it}, is given by the linear specification:

$$(2) \qquad v_{it} = \mu + \gamma\theta_{it} + \beta\bar{\theta}_t.$$

The parameter μ is the mean return to the course, and γ is the effect of individual skill, θ_{it}, on course value added. We assume that $\gamma > 0$, i.e., the return to college is increasing in individual ability. Below this assumption will yield the prediction that more able students obtain more schooling on average—a well-documented empirical regularity.

The parameter β allows an individual's skill return to vary with the mean ability of her classmates, which we denote by $\bar{\theta}_t$. This allows for a linear-in-means *peer learning externality* (Manski, 1993). Classroom learning spillovers could arise for many reasons, including peer interactions (Sacerdote, 2001), changes in instruction (Duflo et al., 2011), and student effort responses (Murphy and Weinhardt, 2018). The reduced form parameter β reflects the net effect of these mechanisms. Its sign is *a priori* ambiguous and is a key question in our analysis. We can shut down learning externalities by considering the case $\beta = 0$.[5]

1.2. **Peer information externalities.** We also allow classmates to affect individual outcomes through information channels. For this we begin by assuming that students do not perfectly know their pre-college skill, θ_{i0}. This assumption is motivated by high drop-out rates at the university in our analysis, where almost half of enrollees do not graduate.

Students have priors on their ability from the skill distribution in the first-year class. We assume ability at the start of the first course is normally distributed, $\theta_{i0} \sim N(\bar{\theta}_0, \sigma^2)$, and students know the mean, $\bar{\theta}_0$, and the variance, σ^2, of this distribution.[6] This assumption is motivated by our empirical setting, in which a selective university admits small cohorts to each major, and these cohorts take initial courses together. In this context, students may have expectations on their own skill from knowledge of the admission process.

Students learn about their skill through course grades. At the end of each class, students receive a grade, g_{it}, that is a noisy measure of their current skill:

$$(3) \qquad g_{it} = \theta_{i,t+1} + \epsilon_{it}^g,$$

where $\epsilon_{it}^g \overset{i.i.d.}{\sim} N(0, \sigma_g^2)$ and the error variance, σ_g^2, is known. Skill accumulation is recursive, so a student's grade in course t is a signal of her skill at the start of the next year, $t+1$. We

---

[5] The parameter β corresponds to what we identify in our empirical analysis: the returns to a large increase in mean classmate ability, $\bar{\theta}_t$. β is unlikely to characterize all potential classmate externalities. In the model we assume $\beta > -\gamma$, which ensures that the mean skill return, $E[v_{it}|\bar{\theta}_t] = \mu + (\gamma + \beta)\bar{\theta}_t$, is increasing in $\bar{\theta}_t$.

[6] Appendix B.4 shows that this classroom ability distribution can be microfounded by assuming college programs are perfectly selective based on Gaussian admission scores (MacLeod and Urquiola, 2015).

refer to g_{it} as a *raw grade* because it measures skill on an absolute scale. For example, g_{it} might represent the number of correct answers on coursework and exams. Below we specify how professors convert raw grades into normalized scores, as they are typically reported.

In this setup, a student's expectations for her skill after one year of college are

$$(4) \qquad E[\theta_{i1}|\bar{\theta}_0, g_{i0}] = \frac{\rho_0}{\rho_0 + \rho_g}\bar{\theta}_0^{End} + \frac{\rho_g}{\rho_0 + \rho_g}g_{i0},$$

where $\bar{\theta}_0^{End} = E[\theta_{i1}|\bar{\theta}_0] = \mu + (1 + \gamma + \beta)\bar{\theta}_0$ denotes mean skill at the *end* of the first year. This term equals the prior, $\bar{\theta}_0$, plus the mean skill return. The parameters $\rho_g = 1/\sigma_g^2$ and $\rho_0 = 1/(\sigma(1 + \gamma))^2$ are the *precisions* of g_{i0} and $\bar{\theta}_0^{End}$ for unknown skill, θ_{i1}. Equation (4) is an application of Bayes' rule under normality (DeGroot, 2004, Section 9.5). Expected skill is a linear combination of the accumulated prior, $\bar{\theta}_0^{End}$, and the individual's grade, g_{i0}, where each term is weighted by its relative precision. Below we show that mean skill at the end of the first year, $\bar{\theta}_0^{End}$, differs from mean skill in the next class, $\bar{\theta}_1$, if students drop out.

Our model includes two other actors—professors and employers—and we assume both have the same information on individual ability as students. Specifically, professors do not know an individual's initial ability, θ_{i0}, but they know mean pre-college skill, $\bar{\theta}_0$, grades, g_{it}, and their distributions. Employers observe course choices and grades via student transcripts, and they know the distributions of these variables from their experience in the labor market.[7]

Equation (4) illustrates the *peer information externality*. Professors and employers expect that individuals with the same grade, g_{i0}, have higher skill when they take courses with more able peers, $\bar{\theta}_0^{End}$. We can shut down this externality by considering a case with perfectly precise grades, $\rho_g = \infty$ (equivalently, $\sigma_g^2 = 0$). In this case information frictions disappear after the first year because course grades reveal individual skill, $E[\theta_{i1}|\bar{\theta}_0, g_{i0}] = \theta_{i1}$.

1.3. **Wage determination.** We follow the benchmark employer learning model in specifying how individual skill affects earnings (Jovanovic, 1979). Students enter a perfectly competitive labor market after college where the only friction is that employers do not know individual skill. An individual's log wage is equal to her expected skill given the employer's information. In addition to courses and grades, employers observe a measure of a worker's on-the-job performance, y_{it}, at the end of each year, where

$$(5) \qquad y_{it} = \theta_{i,t+1} + \epsilon_{it}^y$$

with $\epsilon_{it}^y \overset{i.i.d.}{\sim} N(0, \sigma_y^2)$. We assume all employers observe y_{it}, and ignore any post-college skill accumulation due to on-the-job training (Farber and Gibbons, 1996).

---

[7] One could argue that students know their own ability better than other actors. Conversely, professors or employers may be better informed about a student's potential returns from their knowledge of advanced coursework or labor market institutions. We abstract from these possibilities for simplicity.

Under these assumptions, an individual's log wage in year t, w_{it}, is defined to be

$$(6) \qquad w_{it} = E[\theta_{it}|\bar{\theta}_0, \mathbf{g}_i, \mathbf{y}_{i,t-1}],$$

where $\mathbf{g}_i$ is a vector of individual i's grades, and $\mathbf{y}_{i,t-1}$ is her history of productivity signals. Wages as defined by equation (6) exhibit the standard properties of employer learning models. An individual's wage converges to her true skill over time as employer inferences improve. Characteristics that are observable at labor market entry—mean classroom ability, $\bar{\theta}_0$, and grades, $\mathbf{g}_i$—decline in importance over a worker's career (Altonji and Pierret, 2001).

1.4. Professor grading standards.

We now describe professors' incentives in setting grades when there are peer externalities. This links our model and empirics, as it shows why students may be more or less likely to pass courses with higher-ability classmates.

We model the professor by letting her choose a failure threshold F that determines which students pass the class. For simplicity, we focus on the first-year professor and assume college consists of only two courses, $t \in \{0, 1\}$. First-year students with raw grades above the failure threshold, $g_{i0} > F$, continue to the upper-level course, $t = 1$. Students with $g_{i0} < F$ drop out and enter the labor market. We abstract from the possibility of retaking courses.

The failure threshold allows for a simple translation of raw grades into normalized grades as they are typically reported. We let $\tilde{g}_{i0}$ denote a student's *normalized grade*, and define it as the difference between the individual's raw grade and the failure threshold:

$$(7) \qquad \tilde{g}_{i0} = g_{i0} - F.$$

With this normalization, a student passes the first-year course if $\tilde{g}_{i0} > 0$ and fails otherwise.[8]

To illustrate the professor's incentives in grading, we consider three possible choices of F:

(1) Professors choose the same F that an individual student would choose;
(2) Professors choose F to maximize the average return given learning externalities; and
(3) Professors choose F to maintain a fixed grading curve.

Below we present expressions for the failure threshold in each case and describe the main mechanisms. Appendix B.1 contains full derivations.

1.4.1. *Individual grading standards.*

A benchmark case is to let professors make the same schooling decisions as individuals would themselves. This reflects a benevolent professor who passes students who benefit from staying in college, and fails students who do not.

We follow Becker (1964) in assuming that individuals would continue in college if the expected return exceeds the cost. To focus on the role of classmates, we abstract from heterogeneity in costs by assuming that students stay in college if the expected wage return

---

[8] Grades are often normalized further to match a common scale (e.g., 0–4 in the U.S.). Our model abstracts from these additional normalizations since we assume all actors know the grade distribution.

exceeds an interest rate, r.[9] Equation (6) defines wages with and without the upper-level course, and the student's information on her skill includes her prior, $\bar{\theta}_0$, and her first-year grade, g_{i0}. Appendix B.1 shows that an individual would choose to stay in college if

(8)
$$E[v_{i1}|\bar{\theta}_0, g_{i0}] > r.$$

In other words, students persist in college if the expected skill return from the upper-level course, v_{i1}, exceeds the cost of an additional year, r.

Under the assumption that skill returns are increasing in individual ability, $\gamma > 0$, the decision rule (8) is equivalent to a benevolent professor's grade threshold that we denote by F^{Ind}. Students whose expected returns exceed r have first-year grades $g_{i0} > F^{Ind}$, and students with $g_{i0} < F^{Ind}$ choose to drop out. Using equations (2) and (4), this threshold is

(9)
$$F^{Ind} = \left(1 + \frac{\rho_0}{\rho_g}\right)\frac{r - \mu - \beta\bar{\theta}_1}{\gamma} - \frac{\rho_0}{\rho_g}\bar{\theta}_0^{End}.$$

The individual grade threshold, F^{Ind}, depends on information externalities because *expected* returns are increasing in the accumulated prior, $\bar{\theta}_0^{End}$. All else equal, imperfect information causes higher-ability classes to have lower failure thresholds, which makes students more likely to pass. This effect disappears if grades are perfectly precise, $\rho_g = \infty$.

Learning externalities also affect the individual threshold because expected returns vary with mean ability in the advanced course, $\bar{\theta}_1$. We characterize this term in the next section, but students naturally expect that $\bar{\theta}_1$ is increasing in mean pre-college ability, $\bar{\theta}_0$. An increase in $\bar{\theta}_0$ raises the return to college if learning externalities are positive, $\beta > 0$, which leads to a lower failure threshold. If $\beta < 0$, higher peer ability increases F^{Ind}.

1.4.2. *Optimal grading standards.* The individual grade threshold assumes that professors do not internalize the effect of student drop-out on mean skill in the upper-level class, $\bar{\theta}_1$. If there are learning externalities, however, the professor may have an incentive to "weed out" some students to raise average skill in the advanced classroom.

Our second rule allows for this by supposing that the professor sets F to maximize the average return among her students. The failure threshold that maximizes the average return balances two effects. First, a higher value of F increases mean skill in the upper-level classroom, $\bar{\theta}_1$, which affects the return to future coursework if there are learning externalities, $\beta \neq 0$.[10] Second, a higher F reduces the fraction of students who pass the class and thus

---

[9] Card (2001) derives a version of this simplified human capital model. See Appendix B.1 for details.

[10] Given the normality assumptions, mean skill in the upper-level class is given by

$$\bar{\theta}_1 = \bar{\theta}_0^{End} + \frac{\rho_g \sigma_{0g}}{\rho_0 + \rho_g}\lambda\left(\frac{F - \bar{\theta}_0^{End}}{\sigma_{0g}}\right),$$

where σ_{0g} is the standard deviation of first-year grades, g_{i0}, and $\lambda(\cdot)$ is the inverse Mills ratio for the standard normal distribution. $\bar{\theta}_1$ is increasing in both F and $\bar{\theta}_0$. See Appendix B.1 for details.

cannot accumulate further skill. We refer to this grading standard as the *optimal* threshold, and denote its value by F^*.[11] Appendix B.1 shows that the optimal threshold is given by

$$(10) \qquad F^* = \left(1 + \frac{\rho_0}{\rho_g}\right)\frac{r - \mu}{\gamma + \beta} - \frac{\rho_0}{\rho_g}\bar{\theta}_0^{End}.$$

One can show that $F^* > F^{Ind}$ if and only if $\beta > 0$ (see Appendix B.1). Thus if there are positive learning externalities, the professor has an incentive to "weed out" some students who would choose to stay in college because higher mean skill, $\bar{\theta}_1$, benefits other students in the upper-level class. If $\beta < 0$, the professor cannot influence upper-level class composition because the individual's schooling decision binds, $F^{Ind} > F^*$. If there are no learning externalities, $\beta = 0$, the individual and optimal failure thresholds coincide, $F^* = F^{Ind}$.

The key difference between the individual and optimal thresholds is that F^* does not depend on $\bar{\theta}_1$. This means that the optimal threshold is not increasing in mean classroom ability (i.e., $dF^*/d\bar{\theta}_0 \le 0$). Professors who maximize the average return may "weed out" some students, but they are not more likely to fail a *given* student in a higher-ability class. The intuition is that the benefit of failing a student—which raises her peers' learning if $\beta > 0$—is offset by the cost of preventing her from taking further coursework.

1.4.3. *Fixed grading standards.* Lastly, we consider the possibility that professors use a fixed grading curve. With a fixed curve, the same proportion of students fail regardless of mean classroom ability. We denote this threshold by F^{Fixed}, and let q be the fraction of students whom the professor wishes to fail. Appendix B.1 shows that the fixed failure threshold is

$$(11) \qquad F^{Fixed} = \bar{\theta}_0^{End} + \sigma_{0g}\Phi^{-1}(q),$$

where $\Phi^{-1}(\cdot)$ is the inverse standard normal CDF, and σ_{0g} is the standard deviation of first-year grades, g_{i0}. The fixed failure threshold increases one-for-one with mean end-of-year skill, $\bar{\theta}_0^{End}$. Importantly, a fixed curve increases the likelihood that a given student fails in a more able class, even if there are no peer externalities.

Anecdotally, many professors grade on a fixed curve, but this grading rule can be harmful if there is a large increase in mean classroom skill. If the returns to college are increasing in individual ability, the threshold F^{Fixed} fails more and more students who would benefit from further coursework as $\bar{\theta}_0$ increases. Thus professors may not choose to use the same curve if they experience a significant change in the average ability of their students.

Ultimately, the question of how professors grade is an empirical one, and professors likely differ in their standards. We now derive testable predictions that allow us to distinguish between peer externalities and professor grading choices in the data.

---

[11] Formally, we define the optimal threshold as $F^* = \underset{F}{\arg\max}\, E[w_{it}(G_i) - G_i r|\bar{\theta}_0, F]$, where G_i is an indicator for $g_{i0} > F$, $w_{it}(G_i)$ is the wage given G_i, and r is the cost of an additional year of college.

1.5. **Peer effects in first-year grades.** This section characterizes peer effects in first-year grades. We define this peer effect as the expected change in a student's normalized grade, $\tilde{g}_{i0}$, from an increase in mean classmate ability, $\bar{\theta}_0$, holding fixed individual ability, θ_{i0}. We let π_0 denote the peer effect in first-year grades, and define it formally as:

$$(12) \qquad \pi_0 = \frac{dE[\tilde{g}_{i0}|\theta_{i0}, \bar{\theta}_0]}{d\bar{\theta}_0} = \frac{dE[g_{i0}|\theta_{i0}, \bar{\theta}_0]}{d\bar{\theta}_0} - \frac{dF}{d\bar{\theta}_0}.$$

Grade peer effects can arise through either the student's raw grade, g_{i0}, or the professor's failure threshold, F. We develop two propositions that help separate these channels. Below we state the propositions and describe the intuition. Appendix B.2 contains details.

Proposition 1. *If peer effects in first-year grades are positive, $\pi_0 > 0$, then either:*

- *Higher-ability classmates increase individual learning, $\beta > 0$; or,*
- *Classmates have information externalities, $\rho_g < \infty$.*

If peer effects in first-year grades are negative, $\pi_0 < 0$, then either:

- *Higher-ability classmates reduce individual learning, $\beta < 0$; or,*
- *Professors use fixed grading standards, F^{Fixed}.*

Proposition 1 describes what one learns from the sign of grade peer effects, π_0. Positive peer effects can arise for two reasons. First, if learning externalities are positive, $\beta > 0$, an individual accumulates more skill in higher-ability classes and thus has a higher expected raw score, g_{i0}. Second, information externalities lead to *lower* failure thresholds in more able classes when professors use either individual or optimal grading standards, F^{Ind} or F^*.

Grade peer effects can be negative if learning externalities are negative, $\beta < 0$, because more able classmates reduce a student's expected raw score, g_{i0}. Higher-ability classmates can also reduce an individual's grade if professors use a fixed grading curve, F^{Fixed}, even if there are no other externalities. Since our empirical analysis finds evidence that $\pi_0 < 0$, we develop a second proposition that helps distinguish between these two possibilities.

Proposition 2. *If professors use fixed grading standards, F^{Fixed}, then peer effects in first-year grades, π_0, are decreasing in the skill return to individual ability, γ.*

If professors use optimal grading standards, F^*, then π_0 increases with γ.

Proposition 2 describes how grade peer effects, π_0, vary with the return to individual skill, γ. This helps distinguish between fixed and optimal grading standards.[12] If professors use a fixed curve, then the failure threshold, F^{Fixed}, increases one-for-one with mean end-of-class skill, $\bar{\theta}_0^{End}$. An increase in mean pre-college ability, $\bar{\theta}_0$, has a bigger effect on the failure

---

[12] The relationship between π_0 and γ cannot be signed in general if $F = F^{Ind}$. Thus Proposition 2 allows us to reject either fixed or optimal grading standards, but it does not rule out individual grading standards.

threshold when when γ is large because $\bar{\theta}_0^{End} = \mu + (1 + \gamma + \beta)\bar{\theta}_0$. In other words, more able peers induce more curving when grades are strongly related to individual ability.

If professors use optimal standards, F^*, then grade peer effects *increase* with γ. This effect arises from information externalities. Professors set lower failure thresholds in more able classes because students have higher expected ability. Information externalities are more pronounced when γ is large, leading to a more positive peer effect in grades.

1.6. **Peer effects in graduation and wages.** We conclude by describing peer effects in college graduation and post-college wages. As above, we define a peer effect as the expected change in each outcome from an increase in mean pre-college ability, $\bar{\theta}_0$, holding fixed individual skill, θ_{i0}. We briefly describe the mechanisms and summarize the predictions in Proposition 3. Appendix B.3 provides details.

Peer effects in graduation reflect the same mechanisms as peer effects in grades. In our model, graduation is equivalent to passing the first-year class since college is only two years long.[13] An individual passes the class if her normalized grade, $\tilde{g}_{i0}$, is greater than zero. Thus peer effects in graduation have the same sign as peer effects in first-year grades, π_0.

Peer effects in wages reflect a broader set of mechanisms. Wages depend on an individual's skill, which classmates can affect directly through learning externalities, β. Classmates can also affect skill accumulation indirectly by influencing whether individuals continue onto upper-level courses. Lastly, classmates can affect wages through information channels if employers use grades or priors on peer ability to infer individual skill.

Appendix B.3 shows that information mechanisms cause individuals to earn higher wages when they take classes with more able peers.[14] This holds because individuals have higher expected skill in more able classrooms. Peer effects in wages can only be negative if higher-ability classmates reduce individual skill. To distinguish between direct and indirect skill mechanisms, one can examine peer effects on wages or other measures of skill within the population of students who persist in college. Conditional on graduation, higher-ability classmates lead to lower wages only if there are negative learning externalities, $\beta < 0$.

Proposition 3 summarizes the predictions for graduation and wage peer effects.

Proposition 3.
   (a) Peer effects in graduation have the same sign as peer effects in first-year grades, π_0.
   (b) Peer effects in log wages can be negative if learning externalities are negative, $\beta < 0$,
      or if higher-ability classmates reduce the probability of graduation.
   (c) Conditional on graduation, peer effects in wages (or skill) are negative only if $\beta < 0$.

---

[13] Upper-level professors have no incentive to fail students in their last year of college in our model.

[14] One could test for the role of information externalities by looking at changes in wage peer effects over workers' careers, as in Altonji and Pierret (2001). We show this theoretically in Appendix B.3, but our data do not contain enough years of labor market experience for us to carry out this test.

We now turn to our empirical analysis, which exploits a natural experiment to shed light on the peer mechanisms in Propositions 1–3.

2. Data and identification

This section describes a tracking admission system at a Colombian flagship university, which allows us to identify how classmates affect an individual's returns to college. We give background on the flagship, our data, and the admissions system. We then present our empirical specification and show that tracking affected an individual's college peers.

2.1. **The flagship university.** The setting for our paper is Universidad del Valle, or *Univalle*, as it is often known. Univalle is a public flagship university in Cali, Colombia—the country's third largest city and the capital of the Valle del Cauca region. Like most Colombian flagships, Univalle offers a wide range of undergraduate and graduate programs and is much less expensive than comparable private colleges. These features make Univalle the largest and most selective university its region, and one of the more selective colleges in the country. In our sample, the median Univalle enrollee scored at the 88th percentile on the national college admission exam, and its yield among admitted students was 85 percent.

Each year Univalle offers admission to roughly 50 undergraduate majors that we refer to as "programs." As in many countries, prospective college students in Colombia apply to both an institution and a major. Admissions at Univalle are based solely on a student's performance on a national standardized exam called the ICFES.[15] The ICFES is similar to the U.S. SAT exam, but it is taken by nearly all high school graduates in the country. Admissions are determined by a program-specific weighted average of scores on different ICFES subject tests. The highest scoring applicants are admitted, with admission cutoffs determined by the number of available slots in each program.

Many programs at Univalle are offered twice per year, i.e., students can begin college in either January or August. Semi-annual admissions are the norm in Colombia because high schools also operate on two different academic calendars (Riehl, 2015). Univalle's test score admissions and semi-annual cohorts are the basis of our identification strategy below.

2.2. **Data sources.** For our analysis we use two datasets provided by the flagship university:

(1) Lists of applicants to Univalle's undergraduate programs from 2000–2003. These lists contain each applicant's admission score and admission outcome.
(2) Transcript records for all students in our sample of programs who enrolled in Univalle. The data contain course names, dates, and grades for all classes that each student took at the flagship through 2017, as well as their graduation outcome.

---

[15] The ICFES exam is now called Saber 11, but we use the name from the time period of our data.

TABLE 1. Tracking admissions example — Fall 2003 applicants to architecture

(A)	(B)	(C)
Admit rank	Admit score	Admission decision
1	404.16	
...	...	Admitted to Fall 2003 cohort
60	315.75	
61	315.05	
...	...	Admitted to Spring 2004 cohort
132	259.14	
133	258.94	
...	...	Rejected
426	14.01	

Notes: The admission score (column (B)) is a weighted average of an applicant's ranks on each ICFES subject score.

We combine the flagship data with three individual-level administrative datasets:

(1) Records from the ICFES national standardized college admission exam that include all students who took the exam in 1998–2003. These data contain students' test scores on the different exam subjects and their demographic characteristics. We also use ICFES records on a college *exit* exam (described below).

(2) Records from the Ministry of Education on students who enrolled in nearly all colleges in the country between 1998–2012.[16] These records contain each student's institution, program of study, date of entry and exit from college, and graduation outcome.

(3) Earnings records from the Ministry of Social Protection for the years 2008–2012. These data contain monthly earnings and days of employment for any college enrollee working in the formal sector.

We link the flagship and administrative data sources using individuals' names, birthdates, and ID numbers (see Appendix C.1). The resulting dataset allows us to observe admission scores, college choices, graduation outcomes, and earnings for all Univalle applicants, even if they attended another university. We observe college courses and grades only for applicants who enrolled in the flagship. We discuss this potential sample selection concern below.

2.3. **Tracking admissions.** Most programs at Univalle used separate admission cycles for the fall and spring cohorts. A few programs, however, used a "tracking" admission system that allows us to identify classmate peer effects on an individual's returns to college.

Table 1 gives an example of tracking for the flagship's architecture program. In 2003, 426 students applied to the architecture cohort that would begin in the fall of that year. The flagship computed admission scores that were weighted averages of applicants' ICFES

---

[16] College admissions in Colombia are decentralized; students apply to individual schools and each institution determines its own criteria. Nonetheless the Ministry tracks enrollment and graduation at almost all colleges.

TABLE 2. Programs with tracking admissions

		(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
			Apply pools	Total applied	Total admitted		Mean cohort size		Grad rate
Group	Program	Years		Fall	Fall	Spring	Fall	Spring	
Arch	Architecture	2000–03	7	1,488	231	234	50	50	0.53
Bus	Accounting (day)	2000–03	7	928	283	292	67	64	0.58
	Accounting (night)	2000–03	7	921	202	191	46	44	0.53
	Business Admin (day)	2000–03	7	1,171	295	289	68	62	0.63
	Business Admin (night)	2000–03	7	940	207	196	45	44	0.49
	Foreign Trade	2003	1	126	45	49	45	40	0.64
Eng	Chemical Engineering	2000	2	233	75	60	59	49	0.50
	Electrical Engineering	2000	2	129	66	60	60	52	0.41
	Electronic Engineering	2000	2	403	72	60	58	56	0.54
	Materials Engineering	2001	2	120	64	56	59	54	0.30
	Mechanical Engineering	2001	2	209	69	57	56	59	0.24
	Total	2000–03	46	6,668	1,609	1,544	56	53	0.53

Notes: Column (A) shows the years in our sample that each program used tracking. Column (B) shows the number of application pools in these programs and years. In 2000–2002, each program had separate admissions for students applying with pre- and post-2000 ICFES scores. In 2003, all applicants had to submit post-2000 exam scores.

Column (C) shows the total number of applicants (all applicants applied in the fall). Columns (D)–(E) show the total number of students admitted to the fall and spring cohorts. Columns (F)–(G) show the mean number of students who enrolled in each cohort. Column (H) shows each program's graduation rate across all cohorts.

subject scores. The top 60 students based on this score were admitted to the fall cohort (August 2003). The next 62 students were also admitted, but to an architecture cohort that began in the spring (January 2004). All other applicants were rejected. We call this "tracking" because students were tracked into fall and spring cohorts by admission scores.

Table 2 shows the programs and years in which the flagship used tracking. The architecture program had tracking admissions from 2000–2003. Accounting and business administration also used tracking in these four years, including separate day- and night-time programs.[17] Foreign trade and five engineering programs had tracking admissions for one year during this time period. All other Univalle programs did not use tracking; they were either offered only once per year or had separate admissions for fall and spring cohorts.

Throughout this paper, we use the term "application pool" to refer to the set of students who applied to the same program at the same time, and thus faced the same admission cutoff. From 2000–2002, programs had two separate application pools each fall because the ICFES exam underwent a major reform in 2000 (Riehl, 2019). The flagship allowed students to apply using either old or new exam scores, and had separate cutoffs for the two groups. After 2002, the flagship accepted only post-reform ICFES scores. Column (B) in Table 2

---

[17] Architecture and the five business programs used tracking admissions through 2004, but we do not have data on 2004 applicants. These programs switched back to semi-annual application cycles in 2005.

shows that our data contain 46 application pools across all programs and years. Roughly 6,700 students applied to programs with tracking admissions in these years (column (C)).

We use the term "cohort" to refer to the group of students who *began* in the same program at the same time. Nearly 3,200 applicants were admitted to any cohort, for an overall admission rate of 47 percent (columns (D)–(E)). Admission totals were similar for fall and spring cohorts, with about 50–70 admits per cohort in each program. Since the flagship is considered the top college in the area, 85 percent of admitted students accepted their offer. This led to fall and spring cohorts of 45–65 students per program (columns (F)–(G)).

Despite the flagship's selectivity, graduation rates are low in all programs (column (H)). The mean graduation rate across all programs and cohorts is 53 percent, with completion rates below one-third in several engineering programs. This shows that students have significant uncertainty about their suitability for college coursework at the time of enrollment.

Tracking created variation in an individual's classmates because enrollment cohorts usually take core courses together. For example, the 60th ranked student in Table 1 was admitted to a higher-ability fall cohort, and would typically take classes with more able peers if she enrolled. The 61st ranked applicant would often be in classrooms with less able peers from the flagship's spring cohort. Below we quantify how tracking affected both cohort and classroom peer composition, and also consider potential effects of differences in enrollment timing.

2.4. **Regression discontinuity specification.** The tracking admission system lends itself to a regression discontinuity (RD) design. A growing body of research uses RD designs to analyze the effects of attending a more selective university or field of study (e.g., Hastings et al., 2013; Kirkebøen et al., 2016). This is similar to analyzing the lower threshold in Table 1 because rejected applicants often attend less selective programs. In this paper, we focus instead on the first threshold in Table 1. This allows us to compare students admitted to the same college and major, but who took classes with peers of different ability.

Our benchmark empirical specification is the stacked RD regression

$$(13) \qquad y_{ip} = \pi D_i + \alpha^b x_i + \alpha^a D_i x_i + \gamma_p + \epsilon_{ip} \quad \text{if } |x_i| \leq h,$$

where y_{ip} is an outcome for individual i in application pool p. The running variable, x_i, is a student's admission rank (column (A) in Table 1) normalized so the last student above the tracking threshold in each pool has $x_i = 0$. D_i is an indicator for having an admission score above the tracking threshold. We allow for different coefficients on the running variable above and below the threshold by including both x_i and $D_i x_i$. The regression includes application pool fixed effects, γ_p, which are defined by an applicant's program, year, and whether she applied with old or new ICFES scores (column (B) in Table 2). We cluster standard errors at the individual level, as a few students applied to more than one flagship program.

Our sample includes applicants to the flagship programs and years in Table 2. We only include students admitted to either a fall or spring cohort (columns (D)–(E)). We also drop 94 students who were admitted through special disadvantaged group quotas. Our regressions focus on the subset of applicants whose admission ranks are within h positions of the tracking threshold. Our benchmark model uses $h = 30$, which is roughly the mean of the Calonico et al. (2014) bandwidths across all dependent variables. Appendix Tables A5–A8 show our main results using other bandwidths, and Appendix Figures A3–A6 show RD graphs of our main outcomes by program group. Appendix C.1 provides details on our sample.

The coefficient of interest, π, estimates the effect of admission to the high-ability cohorts at the tracking threshold. We present estimates that pool across all programs, as well as separate estimates for the architecture, business, and engineering groups defined in Table 2.

The main identification assumption is that individuals near the tracking threshold do not have perfect control of their admission score (Lee and Lemieux, 2010). Although students likely have an idea about the program's quota and standards, there is uncertainty in the final admission decision stemming from other applicants. A violation of the identification assumption would arise if, for example, students could petition the admission officer to move the tracking cutoff. This is unlikely given the formulaic nature of admissions.

As a test of this assumption, Figure 1 shows how individual characteristics vary across the tracking threshold. The y-axis in both panels is a student's *predicted* first-year GPA, which is estimated from a linear regression of first-year GPA on individual covariates including gender, age, national exam score, and socioeconomic traits. This dependent variable combines many pre-determined characteristics into a single index measured in GPA units. The x-axis is a student's rank in their application pool. We normalize these ranks to increase in admission scores, and so that the last student above the tracking threshold has rank zero. Dots are means of the dependent variable in five rank bins, and lines are predicted values from separate local linear regressions above and below the threshold.

Figure 1 shows no evidence of a discontinuous change in individual characteristics at the tracking threshold. The sample in Panel A includes all applicants, and the continuity of predicted GPA suggests that the admissions committee did not manipulate applicant ranks. In Panel B, the sample includes only students who chose to enroll in the flagship university. This balance test is important because our data only contain grades for Univalle enrollees. The characteristics of flagship enrollees do not change significantly at the tracking threshold, and the predicted GPAs of applicants and enrollees are similar overall. The results in Figure 1 are corroborated by McCrary (2008) density tests (Appendix Figure A1) and covariate balance tests (Appendix Tables A1–A2). Applicant and enrollee traits are similar because the vast majority of applicants choose to enroll in the flagship, as we show in the next section.
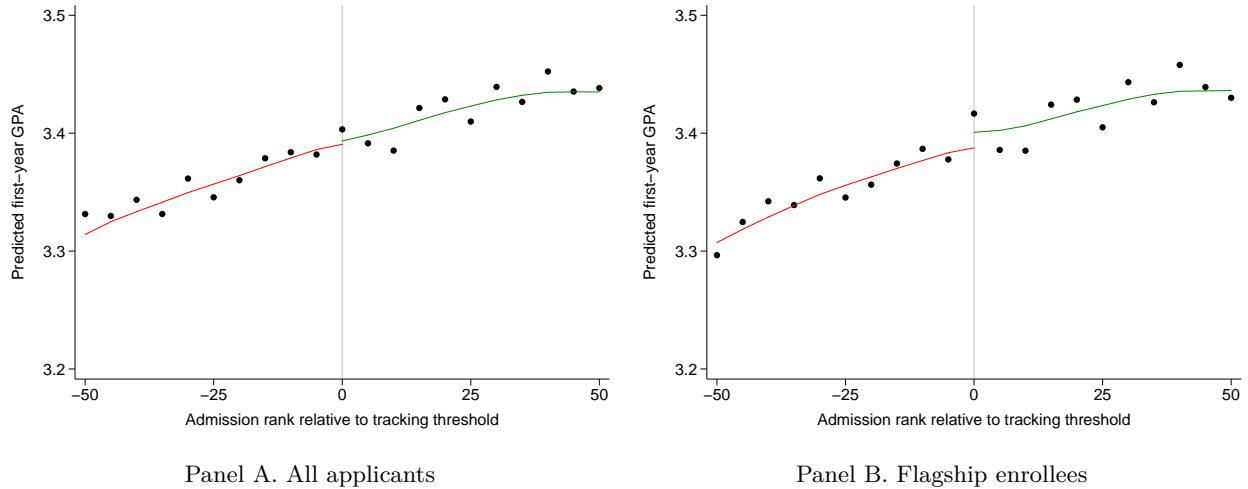
| |
| Panel A. All applicants   Panel B. Flagship enrollees |

FIGURE 1. Predicted first-year GPA (using individual characteristics)

Notes: The x-axis is a student's admission rank normalized so that zero is the first rank above the tracking threshold. The y-axis is an applicant's predicted first-year GPA from a regression with the covariates in Appendix Tables A1. Dots are means in five-rank bins. Lines are from separate above- and below-threshold local linear regressions using a triangular kernel with a 30 rank bandwidth. The sample for Panel A includes all applicants who were admitted to flagship programs with tracking admissions (columns (D)–(E) in Table 2). Panel B includes only flagship enrollees.

2.5. **Enrollment outcomes and cohort peer characteristics.** This section shows that the flagship's tracking admissions altered the cohort peers of applicants near the threshold.

Figure 2 shows that tracking affected students' starting cohorts, but not overall flagship enrollment. These graphs are similar to the RD diagrams in Figure 1, but each panel plots a different enrollment outcome on the y-axis. In Panel A, the dependent variable is an indicator for enrolling in one of the flagship's fall cohorts. The fall enrollment rate was roughly 85 percent for applicants above the tracking threshold, and almost zero percent below the threshold. In Panel B, the dependent variable is an indicator for enrolling in any flagship cohort. Below-threshold students also enrolled in the flagship at an 85 percent rate, and there is no discontinuity in overall enrollment at the threshold. Thus most of these students enrolled in a spring flagship cohort.

Tracking caused large differences in average peer ability in an individual's cohort. In Panel C, we measure cohort peer ability using percentile scores on the national ICFES admission exam. This variable ranges from zero to one and represents an individual's percentile in the national distribution of exam takers. Students below the tracking threshold enrolled in programs where the average student scored at the 77th percentile on the ICFES exam. Mean cohort ability jumps to the 87th percentile at the tracking threshold. Thus tracking induced a 10 percentile increase in mean cohort ability for marginal admits to the fall cohorts.

An increase in peer ability meant that applicants just above the tracking threshold were lower in their cohort's ability distribution. We illustrate this in Panel D by graphing an
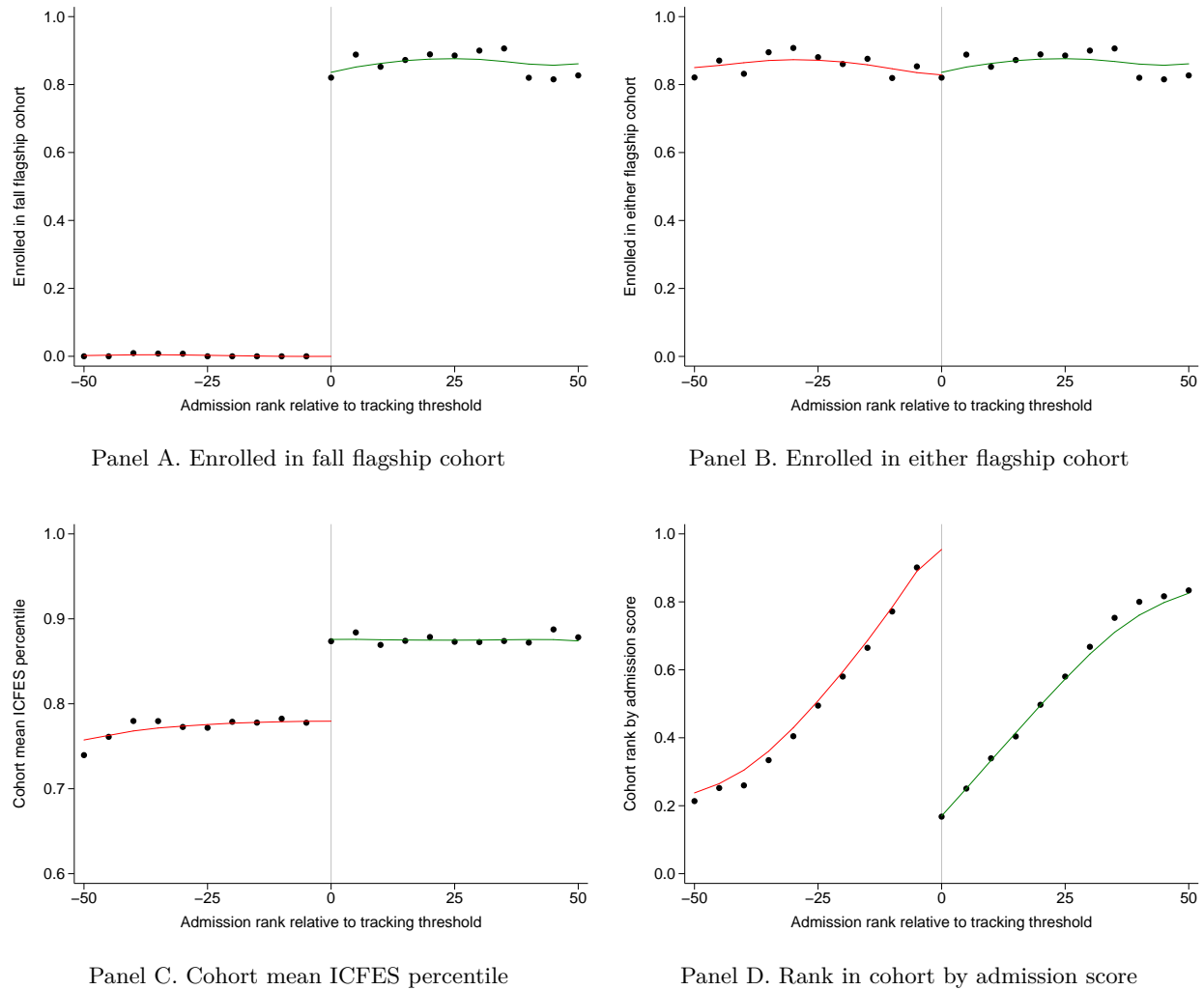
Panel A. Enrolled in fall flagship cohort

Panel B. Enrolled in either flagship cohort

Panel C. Cohort mean ICFES percentile

Panel D. Rank in cohort by admission score

FIGURE 2. Cohort enrollment and peer characteristics

Notes: This figure contains RD graphs (as in Figure 1) for the dependent variables in the panel titles. In Panel C, ICFES percentile is an individual's rank among all test takers in her cohort using the mean score across subjects. The sample for Panels A–C includes all flagship applicants. The sample for Panel D includes flagship enrollees.

individual's percentile rank in their cohort based on admission scores. Individuals just below the tracking threshold are near the top of their cohort's admission score distribution, while individuals just above the threshold are among the lowest ranked students in their cohort. This is a mechanical effect of tracking admissions.[18]

In sum, the vast majority of students in each application pool attended the same college program, but tracking induced significant differences in their cohort peers. The next section

---

[18] Cohort rank does not fall perfectly from one to zero at the threshold because a small number of students enroll in cohorts other than the one they were admitted to. Appendix Table A3 shows that the tracking effects in Figure 2 had similar magnitudes in architecture, business, and engineering programs.

TABLE 3. Examples of first-year required courses

Architecture			Business			Engineering		
Course	% took	% pass	Course	% took	% pass	Course	% took	% pass
Graphics I	0.95	0.62	Basic Mathematics	0.94	0.63	Calculus I	0.91	0.48
Projection Geometry	0.94	0.62	Calculus	0.89	0.72	Vector Geometry	0.94	0.56
Workshop Project I	0.97	0.68	Intro to Accounting	0.98	0.78	Linear Algebra	0.93	0.58
Intro to Technology	0.93	0.76	Computing I	0.96	0.92	Physics I	0.91	0.65
Theory I	0.94	0.81	Colombian Politics	0.93	0.94	Calculus II	0.90	0.70

Notes: This table shows the five most common first-year required courses for the three program groups in Table 2. We define first-year required courses as those that were taken by 75 percent or more of a cohort's graduates, and for which the modal graduate took the course in her first year. "% took" is the proportion of a cohort's graduates who took the course. "% pass" is the proportion of course enrollees who passed the class.

shows how cohort peer composition translated into *classroom* peer composition, and the resulting effects on academic performance.

3. Peer effects in first-year grades

This section shows how tracking admissions affected classmate ability and grades in first-year courses at the flagship. Our main result is that marginal admits to the flagship's fall cohorts received lower grades in classes with more able peers. We present evidence that this result is due to peer effects rather than differences in the timing of college enrollment. Lastly, we use testable implications from Propositions 1–2 to show that these peer effects reflect learning externalities more than grade curving.

3.1. **First-year courses and classrooms.** To estimate peer effects in grades, we focus on required first-year courses because students typically took these classes with their cohort peers. We use the term "course" to refer to classes that have the same name, and we identify first-year required courses from the data. Specifically, we restrict attention to courses that were taken by at least 75 percent of a cohort's graduates, and in which the modal graduate took the course in her first year. We exclude other elective courses because students had discretion on whether and when to take them.[19]

Table 3 shows the five most common first-year required courses in each program group. These include introductory courses related to the major such as Intro to Accounting and Physics I, as well as math courses like Calculus and Geometry. Most courses were taken by over 90 percent of the program's graduates. Passing rates ranged from below 60 percent in some math courses to above 90 percent in several business courses. Appendix Table A4 shows that students took nine first-year required courses on average.

[19] We use a data-driven method to define first-year required courses since program requirements change slightly over time. Our results are similar using other definitions, or including all first-year courses.

We use the terms "classroom" and "classmates" to refer to the set of students who took a course in the same room at the same time. In many cases, courses were offered in several sections, but the large majority of students in any classroom were from the same cohort. On average, first-year classrooms contained 36 students, and 77 percent of an individual's classmates were from her own program and cohort (Appendix Table A4).[20]

Our data do not contain information on professors, but Univalle faculty often teach the same courses in the fall and spring. It is also common for faculty to teach multiple sections of a course in a single semester. Some courses like Calculus and Physics were required in more than one program, and were thus taught by multiple professors. In sum, students in the fall and spring cohorts of a given program were likely taught by many but not all of the same professors. We discuss the role of professors in more detail below.

3.2. **Tracking effects on classmates and grades.** Table 4 shows how tracking admissions affected classroom peers and grades in first-year required courses. Each coefficient is an estimate of π from a separate RD regression (13). These regressions are at the individual/course level with an observation for each individual's first attempt at each course. The top row is the pooled estimate across all programs. Other rows present separate estimates for the three program groups in Table 2.

Columns (A)–(B) show that tracking affected the mean ability of individual's classmates. On average, individuals marginally admitted to the fall cohort took first-year courses with classmates who scored eight percentile points higher on the national ICFES exam (column (A)). As a result, these students were nearly 70 percentile points lower in the classroom distribution of admission scores (column (B)). These effects appear in all program groups, although there were slightly larger peer differences in business programs.

Column (C) shows that marginal admits to higher-ability cohorts received *lower* first-year grades. Crossing the tracking threshold reduced first-year course grades by 0.13 points on average. Colombian grades are on a 0–5 scale with 0.1 point increments (see Appendix Figure A2), and 0.2 points is roughly the difference between a B+ and a B on the U.S. scale. The negative effects occurred in architecture and business programs. In particular, admission to high-ability architecture cohorts reduced students' first-year performance by almost a full letter grade. The grade effect is positive but imprecisely estimated in engineering programs.

Figure 3 shows RD graphs of the results in column (C). Panel A displays the pooled effect for architecture and business programs. Average grades drop at the tracking threshold, and the full distribution of grades is similar in the fall and spring cohorts. Panel B shows the analogous graph for engineering programs. There is little evidence that tracking affected first-year engineering grades, and the mean grade is much lower in the spring cohorts.

---

[20] Students from other cohorts were often retaking the course after previously failing it.

TABLE 4. Tracking effects on classroom composition and grades in first-year courses

	(A)	(B)	(C)	(D)
	Classroom composition		Course grade	
	Mean ICFES percentile in class	Class rank by admit score	Numerical grade	Passing grade
All programs	0.079***	−0.686***	−0.126**	−0.045**
	(0.003)	(0.015)	(0.061)	(0.020)
Architecture	0.042***	−0.587***	−0.472**	−0.166**
	(0.006)	(0.049)	(0.206)	(0.076)
Business	0.090***	−0.731***	−0.164**	−0.044**
	(0.004)	(0.017)	(0.070)	(0.021)
Engineering	0.068***	−0.625***	0.161	0.017
	(0.005)	(0.028)	(0.136)	(0.049)
N	14,267	14,327	14,327	14,327
Mean, ranks [−5,−1]	0.791	0.844	3.476	0.821

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the column header. The sample includes flagship enrollees within 30 ranks of the tracking threshold. Regressions are at the individual/course level with an observation for each individual's first attempt at each first-year required course. The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column (D) in Table 4 shows that marginal admits to high-ability tracks were also less likely to pass first-year courses. The dependent variable is an indicator for receiving a passing grade, which is a value of 3 or above on the Colombian scale. Tracking decreased the course passing rate by 4.5 percentage points on average, with effects only in architecture and business programs. This effect is a 25 percent increase in the overall course failure rate, and a doubling of the failure rate in architecture. In practice, this meant that affected students would have to retake the course to continue in the program.

3.3. **Course timing.** The flagship's tracking admissions also affected the timing of when individuals took first-year courses. Applicants above the tracking threshold typically began the flagship program in August of the year they applied, while most below-threshold applicants started in the following January. This section provides evidence that differences in course timing are unlikely to explain the grade results in Table 4.

Delayed enrollment could have hindered spring cohort students in several ways. All students in our sample applied to a fall program and were presumably ready to begin at that time. The enrollment lag may have caused students to forget material, consistent with work on learning loss during academic breaks (Cooper et al., 1996). All else equal, any harmful effects of delayed enrollment would cause us to find *positive* effects of tracking.
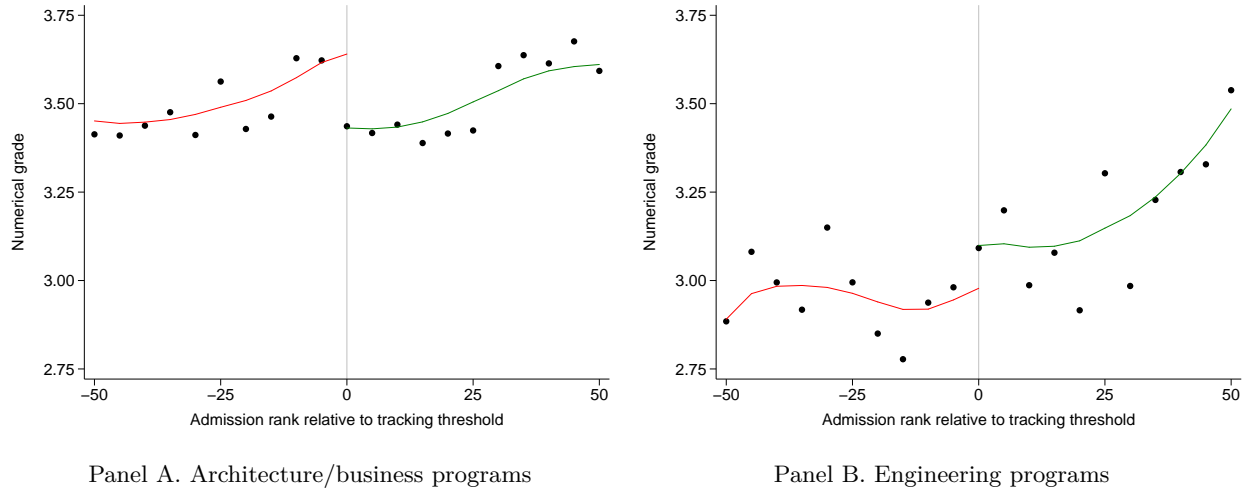
Panel A. Architecture/business programs   Panel B. Engineering programs

FIGURE 3. Grades in first-year courses

Notes: This figure contains RD graphs (as in Figure 1) with grades in first-year required courses as the dependent variable. The graphs include an observation for each individual's first attempt at each course. Panel A includes flagship enrollees in architecture and business programs. Panel B includes enrollees in engineering programs.

On the other hand, delayed enrollment may have helped spring cohort admits if there are benefits to starting college when one is older. Work on age-at-testing finds that students who are old for their grade perform better on standardized exams (Bedard and Dhuey, 2006), with some age benefits observed as late as age 18 (Black et al., 2011).[21] The hypothesized mechanisms in this work do not fully match our setting because students in the flagship's fall and spring cohorts had similar ages in their pre-college years. Further, it is unclear whether age benefits on standardized exams also exist for college coursework.

Table 5 explores the potential influence of age on our results. Columns (A)–(B) show how tracking affected the time at which students took first-year courses. In column (A), the dependent variable is the time in years between individuals' flagship applications and the beginning of each first-year course. On average, marginal admits to the fall cohort took these courses roughly half a year earlier than individuals just below the tracking threshold. Column (B) shows that marginal fall cohort students were also about five months younger at the start of their first-year courses, but the estimated age effects are much less precise.

The imprecise estimates in column (B) argue against age effects as a primary mechanism for the grade results. The age estimates are noisy because Colombian college students vary significantly in age. In Colombia, most students do not begin college right after high school, and the average time off is nearly three years (Riehl, 2015). In our sample, the average flagship applicant is 19.1 years old, and the standard deviation of applicant ages is 3.5 years.

---

[21] Black et al. (2011) find that a one year increase in age-at-test at around age 18 raises scores on a Norwegian IQ exam by one-tenth of a standard deviation. In their setting, older individuals also have more schooling.

TABLE 5. Tracking effects on first-year course timing and grades by age

	(A)	(B)	(C)	(D)	(E)
	Course timing		Course grade		
	Years from application to course	Age in yrs at start of course	Age at application < median	Age at application ≥ median	Controls for age in months
All programs	−0.433***	−0.398	−0.119	−0.151*	−0.158**
	(0.019)	(0.278)	(0.086)	(0.085)	(0.062)
Architecture	−0.425***	−0.195	−0.622**	−0.172	−0.470**
	(0.057)	(0.682)	(0.276)	(0.315)	(0.206)
Business	−0.465***	−0.312	−0.101	−0.253**	−0.184***
	(0.024)	(0.376)	(0.098)	(0.098)	(0.070)
Engineering	−0.350***	−0.789*	0.141	0.133	0.081
	(0.038)	(0.448)	(0.196)	(0.189)	(0.139)
N	14,327	14,311	7,017	7,294	14,311
Mean, ranks $[-5,-1]$	0.953	20.558	3.506	3.454	3.476

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the column header. The sample includes flagship enrollees within 30 ranks of the tracking threshold. Regressions are at the individual/course level with an observation for each individual's first attempt at each first-year required course. Columns (C)–(D) split the sample based on the median age in each application pool. Regressions in column (E) include fixed effects for age in months at the start of each course. The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The variation in course timing induced by tracking—six months—is small in comparison to the sample variation in ages, which makes age-related mechanisms less plausible.

Columns (C)–(D) in Table 5 corroborate this argument by showing that the grade results are similar across age groups. These columns estimate the regressions in column (C) of Table 4 separately for students who are below or above the median age in their application pool. The mean age at application is 17.2 for the below-median sample and 20.7 for the above-median sample. Crossing the tracking threshold reduced first-year grades for both younger and older students. The mean effect is slightly more negative for older students, but we cannot reject identical point estimates for any program group. Appendix Table A9 shows that we also find negative grade effects across age quartiles. The stability of these estimates argues against age-at-testing effects because there is evidence that age benefits decrease as students become older (Bedard and Dhuey, 2006; Cascio and Schanzenbach, 2016).

Column (E) in Table 5 shows that grade effects are similar even when we explicitly control for an individual's age in each course. These regressions are identical to those in column (C) of Table 4, but we include a full set of dummies for a student's age in months at the start of each course. This specification is potentially endogenous because identification comes from

23

students who applied to the flagship at different ages. Nonetheless, marginal fall cohort students had lower grades than spring cohort students who took classes at similar ages, and the point estimates are close to those in Table 4. Controlling for age makes little difference because age is a weak predictor of academic performance in our data. Appendix Table A10 shows that age has a small *negative* correlation with grades—both in the raw data and conditional on students' birth years and ICFES exam scores.

In sum, we find little evidence that grades and tracking effects are related to student age. These effects are thus likely the causal effect of differences in individuals' classmates.

3.4. **Grade curving.** The above results show that exposure to higher-ability flagship peers reduced students' grades and course passing rates. This section uses Propositions 1–2 from our model to show that these effects are not primarily driven by grade curving.

Proposition 1 describes two reasons why a given student could be more likely to fail in higher-ability classes. First, more able classmates may have externalities that reduce learning for students near the bottom of the class, which would cause these individuals to earn lower raw course grades. Second, professors may use fixed grading curves, which would increase the likelihood of failing in more able classes even if there are no learning externalities.

In our setting, the importance of learning and grading mechanisms depends on how flagship professors responded to tracking. Tracking may have affected learning if the variation in classroom ability caused professors to teach differently. To our knowledge, Univalle faculty also had full discretion in setting grades, and some professors may have been reluctant to adjust their curves. Learning and grading mechanisms are both components of the returns to attending a college with higher-ability peers. Nonetheless it is useful to examine whether our results are primarily due to learning or grading effects.

To explore the role of curving, we first ask if course grade distributions were the same in fall and spring cohorts. Panel A of Figure 4 plots the mean grade in each first-year required course, where we compute separate means for classrooms with mostly fall and mostly spring cohort students. The x-axis is the mean grade in classrooms where the majority of students were from the fall cohort of a given program. The y-axis is the mean grade in classrooms where most students were from the subsequent spring cohort of the same program. Black circles are architecture and business courses, and red triangles are engineering courses.

Panel A shows that mean grades varied significantly between fall and spring cohort classrooms. While there is a strong correlation between average grades in the two cohort groups (0.65), many courses lie off the 45 degree line that represents identical means. Hollow symbols depict courses where the mean grades are statistically different at $p < 0.05$. We reject identical means in 41 percent of courses, including many architecture and business courses.
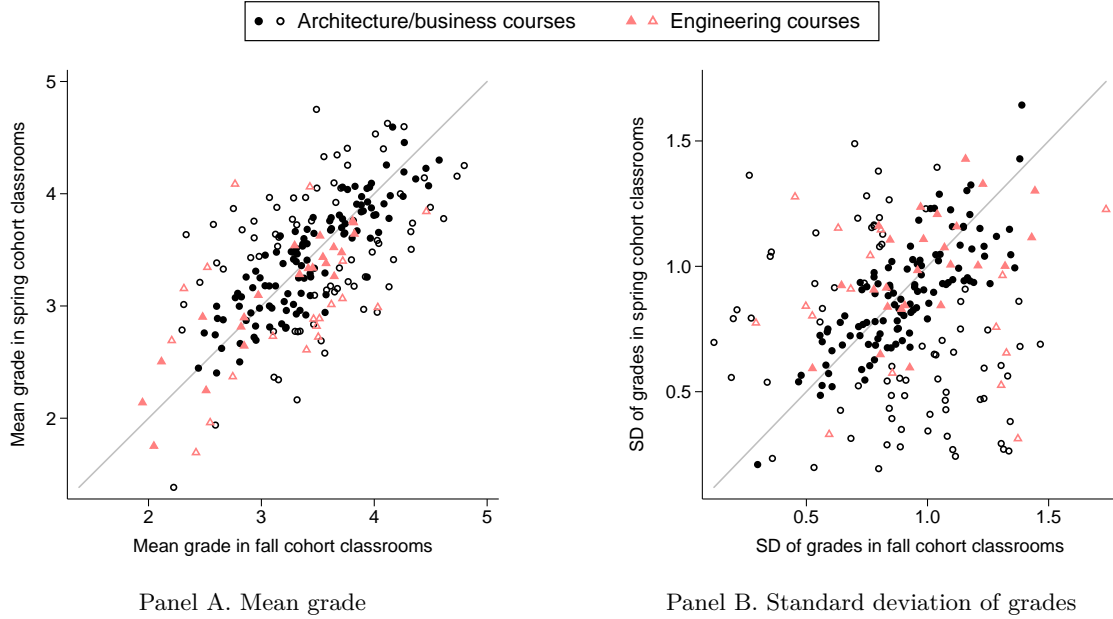
24

Panel A. Mean grade

Panel B. Standard deviation of grades

FIGURE 4. Grade distributions in first-year fall and spring cohort classrooms

Notes: This figure plots grade distributions in first-year required courses. We assign the classrooms of each course to a program and cohort, where cohorts are defined by enrollment year and semester. "Fall cohort classrooms" are those in which at least 50 percent of students were from the fall cohort of a given program. "Spring cohort classrooms" are those in which at least 50 percent of students were from the spring cohort of a given program. We omit classrooms with no cohort majority or with fewer than ten students. 591 classrooms fit these criteria.

In Panel A, dots depict the mean grade at the program/cohort/course level. The x-axis is the mean grade in fall cohort classrooms. The y-axis is the mean grade in the subsequent spring cohort's classrooms of the same program and course. There are 456 program/cohort/course cells, and thus 228 dots. Panel B is similar to Panel A, but dots depict the standard deviation of grades at the program/cohort/course level.

Black circles are architecture and business courses. Red triangles are engineering courses. Hollow symbols indicate that grade means/standard deviations are statistically different in fall and spring cohort classrooms at $p < 0.05$.

The mean absolute deviation between fall and spring cohort average grades is 0.38 points—more than half of a letter grade on the U.S. scale.

Panel B of Figure 4 shows an even weaker relationship between grade variances in first-year courses. This graph is similar to Panel A, but it plots the standard deviation of grades in fall (x-axis) and spring (y-axis) cohort classrooms. Grade standard deviations are only mildly correlated in the two cohort groups (0.17), and many courses lie far from the 45 degree line. We reject identical standard deviations in 40 percent of courses at $p < 0.05$. Thus many professors used different grade distributions in fall and spring cohort classrooms.

Appendix Table A11 shows that the grade effects in Table 4 are not systematically related to whether courses had similar or distinct grade distributions in the fall and spring cohorts. We estimate tracking effects on grades separately for courses with distinct grade distributions in Figure 4 (hollow symbols) and courses with similar distributions (solid symbols). If our results were primarily driven by curving, one would expect to find effects only in courses with similar grade distributions. Instead, tracking induced lower grades in both course groups.

Proposition 2 provides a testable implication of our model that sheds light on how flagship professors graded. The proposition describes how peer effects in first-year grades, π_0, relate to the skill return to individual ability, γ, under different grading methods. If professors use fixed curves, F^{Fixed}, then grade peer effects should be more negative in courses with larger values of γ. Higher-ability peers have a bigger effect on a fixed curve when the return to individual ability is large. If instead professors use optimal grading standards, F^*, then grade peer effects should be *less negative* in courses with higher values of γ. This is due to information externalities; more able classmates have a greater effect on an individual's *expected* ability when γ is large, which induces more lenient grading.

To test this proposition, Figure 5 displays estimates of grade peer effects and returns to individual skill in each first-year course. The y-axis depicts grade peer effects, π_0, which are the RD coefficients from separate grade regressions for each course. The x-axis depicts the course's return to individual skill, γ, which we define as the coefficient from a regression of grades on admission scores for each course. This measure of γ is based on the assumption that the return to individual skill is larger in courses where grades are more related to admission scores. These regressions include classroom fixed effects so that we use only within-classroom variation to identify γ.[22] We normalize our estimates of γ to have standard deviation one, and report the linear relationship between π_0 and γ.

The results in Figure 5 suggest grades are more related to students' expected returns than to fixed curves. A one standard deviation increase in the course's return to individual skill, γ, is associated with a 0.09 point increase in the estimated grade peer effect, π_0. This is consistent with an information benefit to more able classmates; tracking had a bigger effect on mean end-of-class skill in courses with large values of γ, which induced more lenient grading. For example, engineering courses (red triangles) had higher returns to individual ability and also more positive grade peer effects on average. Figure 5 suggests a less important role for fixed grading curves, which would induce a negative relationship between π_0 and γ.

In sum, Figures 4–5 suggest that our results are not driven solely by grade curving. While some professors may have used fixed curves, grade distributions vary significantly across flagship courses, and this variation does not explain the negative effects of tracking. This suggests that the grade results in Table 4 are primarily due to peer learning externalities.

3.5. **Learning externalities.** This section describes two potential sources of learning externalities from the flagship's tracking admissions. First, the variation in grade effects across programs is consistent with teaching responses. As shown in Table 2, most architecture and business programs used tracking admissions for all four years in our sample. By contrast,

---

[22] Appendix C.2 shows that the coefficient on admission scores in this regression is proportional to the γ term in our model. We use a leave-out estimator for γ—students in other admission pools who took the same course—so that γ is not mechanically related to the estimated peer effect, π_0. See Figure 5 for details.
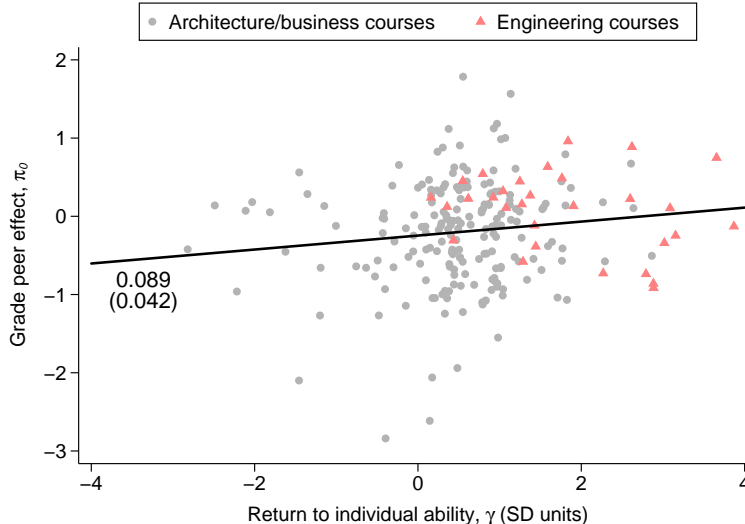
FIGURE 5. Return to individual ability, γ, and grade peer effects, π_0, in first-year courses

Notes: This figure plots estimates of returns to individual skill, γ, and grade peer effects, π_0, in first-year required courses. Dots are estimates for program/application year/course cells (225 in total).

The y-axis depicts grade peer effects, π_0, which are the RD coefficients from estimating equation (13) separately for each cell with grades as the dependent variable.

The x-axis depicts returns to individual skill, γ, which are coefficients from regressing course grades on admission scores in regressions with classroom fixed effects. For this we estimate a "leave-out" regression for each cell using students in other programs and/or application years who took the same course. Our data contain 254 total cells, but we exclude 29 cells for which this leave-out estimator is infeasible. We normalize γ to have standard deviation one.

Grey circles are architecture and business courses. Red triangles are engineering courses. The solid line is the linear relationship between π_0 and γ, and we report the slope coefficient and robust standard error (in parentheses).

engineering programs had tracking for a single year before returning to semi-annual admissions. Thus architecture and business professors had more time to adapt teaching to the variation in classroom ability induced by tracking. This may have improved learning for top students in lower-ability cohorts, leading them to earn higher first-year grades.[23]

Second, heterogenous effects by gender suggest that tracking affected student effort. Appendix Table A12 estimates tracking effects on first-year grades by student gender. The negative effects are overwhelmingly driven by males. For men, crossing the tracking threshold increased the course failure rate by six percentage points in business programs and 26 percentage points in architecture. These effects are close to zero for women. It is hard to explain gender-specific effects through changes in professor teaching or grading alone. One possible mechanism is that men in lower-ability cohorts increased their effort as a result of their higher relative rank, consistent with results in Murphy and Weinhardt (2018).

While our results are consistent with teaching and effort responses, other types of learning externalities are possible. Tracking may have affected social groups inside or outside of the classroom (Carrell et al., 2013). Learning and grading mechanisms could also interact; for

---

[23] Another hypothesis is that it may be easier to adapt curricula in architecture/business than in engineering.

example, students may have reduced effort after receiving lower grades than they anticipated (Wiswall and Zafar, 2015). Regardless of the mechanisms, our results provide clear evidence that exposure to more able classmates reduced individuals' academic performance. We now turn to the longer-term consequences of these initial grade effects.

4. The returns to college peers

This section shows that marginal admits to the high-ability flagship cohorts were less likely to graduate and had lower earnings one decade later. We present evidence that these students also had lower skill even conditional on completing the program. This suggests that the negative earnings return comes from peer learning externalities in addition to graduation effects, consistent with Proposition 3 in our model.

4.1. **Tracking effects on graduation.** The flagship's tracking admissions induced some students in high-ability cohorts to fail first-year courses. This section shows that these students also had lower graduation rates, consistent with Proposition 3(a).

Figure 6 shows the mean effect of tracking on flagship graduation rates across all programs and cohorts. Panel A plots the proportion of students who graduated from their flagship program (y-axis) by each year since the time of application (x-axis). The red circles depict the flagship graduation rate for students just below the tracking threshold—those with admission ranks between -5 and -1. Green triangles show this mean effect plus the RD estimate, which is the π coefficient from a separate estimation of equation (13) for each year. Vertical dashed lines are 95 percent confidence intervals for the RD coefficient.

Panel A shows that applicants just below the tracking threshold graduated at higher rates in the long run. Students began graduating 4–5 years after application, which is the nominal completion time for most programs.[24] The RD estimate is positive in these years, consistent with the six-month enrollment delay for applicants below the threshold. Over the next several years, however, below-threshold graduation rates caught up to and surpassed those for above-threshold students. By ten years after application, marginal admits to the high ability cohorts were eight percentage points less likely to have completed the program.

Panel B in Figure 6 shows the overall flagship graduation rate. This graph is similar to the RD graphs in Figures 1–3, but the dependent variable is an indicator for completing the flagship program by 2017. The tracking effect is noisily estimated as graduation rates are only moderately correlated with admission ranks. Nonetheless, there is evidence of a drop in flagship completion rates at the tracking threshold.

Columns (A)–(B) in Table 6 provide more detail on the flagship graduation effects. As in Table 4, this table displays RD coefficients for the dependent variables listed in the column

---

[24] On-time completion is five years in all programs except Foreign Trade (4.5 years) and nighttime business/accounting (six years).

Panel A. Flagship graduation effects by year   Panel B. Ever graduated from flagship
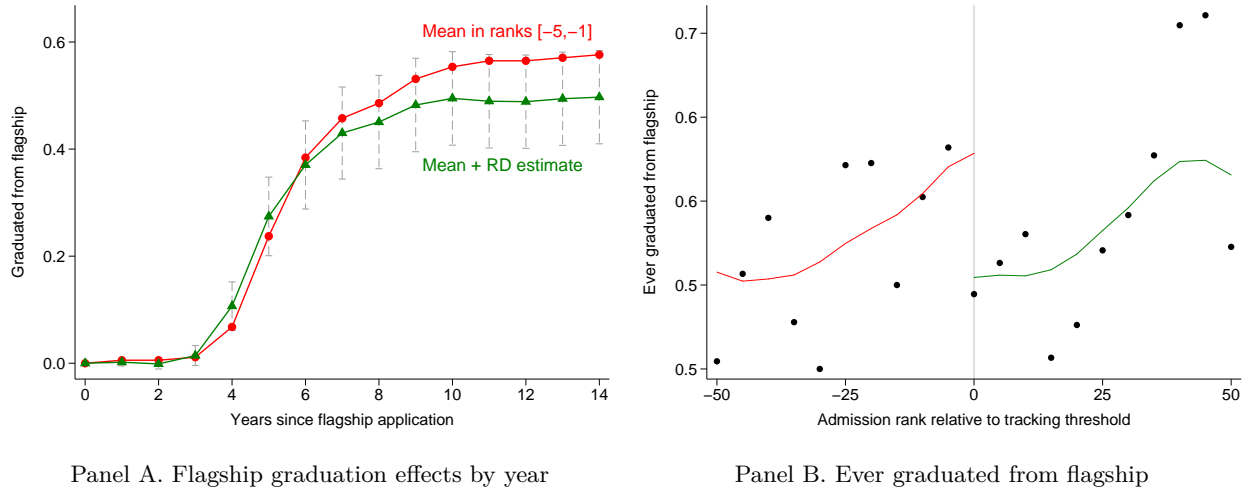
FIGURE 6. Tracking effects on flagship graduation

Notes: Panel A shows the proportion of flagship enrollees who graduated from the flagship (*y*-axis) by years since application (*x*-axis). Red dots depict the flagship graduation rate for enrollees with admission ranks between -5 and -1. Green triangles are equal to the red dots plus the π coefficients from the RD regression (13). Vertical dashed lines are 95 percent confidence intervals for the RD coefficient using standard errors clustered at the individual level.

    Panel B contains an RD graph (as in Figure 1) with an indicator for graduating from the flagship by 2017 as the dependent variable. The sample includes all flagship enrollees.

header. Column (A) shows that crossing the tracking threshold led to an average decrease of 0.64 years of full-time flagship enrollment. Column (B) shows the flagship graduation estimates that correspond to Figure 6. The negative persistence and graduation effects appear in architecture and business programs, consistent with the results for first-year grades. Tracking effects on flagship completion are positive but imprecisely estimated in engineering.

Column (C) shows that marginal admits to high-ability cohorts were also less likely to earn a degree from *any* college. These regressions use Ministry of Education data that allow us to observe graduation at nearly all Colombian colleges up to 12 years after application. The dependent variable is an indicator for finishing any college program, and the sample includes all applicants regardless of flagship enrollment. Crossing the tracking threshold reduced the degree attainment rate by nine percentage points, with effects concentrated in architecture and business. The point estimates are similar to those in column (B), which shows that most students who dropped out of the flagship did not complete another program.

The estimates in columns (A)–(C) of Table 6 are consistent with Proposition 3(a), which states that peer effects in grades and graduation rates have the same sign. Although our model abstracts from the possibility of retaking courses, our results show that the extra costs from failing first-year classes prevented some students from obtaining a college degree.[25]

---

[25] Appendix Table A13 shows that marginal admits to the high-ability cohorts were four percentage points more likely to retake first-year courses after failing, but few enrolled in another college program.

TABLE 6. Tracking effects on educational attainment and earnings

	(A)	(B)	(C)	(D)	(E)	(F)
				Earnings after 5–9 years	Earnings after 10–12 years	
	Educational attainment					
	Full-time years at flagship	Flagship graduate	Any college degree	Log daily earnings	Log daily earnings	Annual earnings in USD
All programs	−0.644***	−0.084*	−0.089**	−0.023	−0.115*	−913
	(0.185)	(0.044)	(0.042)	(0.042)	(0.067)	(734)
Architecture	−1.053*	−0.098	−0.185	−0.131	−0.358*	−3,617**
	(0.629)	(0.138)	(0.120)	(0.124)	(0.215)	(1,781)
Business	−0.857***	−0.137**	−0.113**	−0.042	−0.098	−1,121
	(0.212)	(0.055)	(0.052)	(0.049)	(0.082)	(971)
Engineering	0.167	0.075	0.016	0.126	−0.063	229
	(0.422)	(0.088)	(0.087)	(0.104)	(0.129)	(1,354)
N (all programs)	1,665	1,665	1,865	3,887	2,158	2,158
Mean, ranks $[-5,-1]$	3.537	0.582	0.634	10.282	10.592	7,987

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the column header. The dependent variable in column (A) is the number of academic years in which students took eight or more flagship courses. We observe flagship graduation through 2017 (column (B)) and graduation at other Colombian colleges through 2012 (column (C)). The earnings regressions include an observation for each year in which individuals have positive earnings in 5–9 years (column (D)) and 10–12 years (columns (E)–(F)) after application using 2008–2012 labor market data. Annual earnings in column (F) are expressed in 2012 U.S. dollars.

The sample for columns (A)–(B) includes flagship enrollees. The sample for columns (C)–(F) includes all flagship applicants. All regressions include only applicants within 30 ranks of the tracking threshold. The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.2. **Tracking effects on earnings.** This section shows that the decline in graduation rates from exposure to higher-ability classmates also led to lower post-college earnings. This is consistent with a negative peer effect on skill accumulation, as in Proposition 3(b).

Panel A in Figure 7 shows that below-threshold students began to earn more than above-threshold students one decade after applying. This graph is similar to Panel A in Figure 6, but the dependent variable is log daily earnings measured 5–12 years after application—the period we can observe in our data. The two groups had similar earnings 5–9 years after application, but Figure 6 shows that many students were still working on their degrees during these years. Ten years after application—when nearly all students' college careers were complete—earnings for below-threshold students began to rise more quickly. After 12 years, the earnings gap between the two groups was more than 0.1 log points.

Panel B in Figure 7 shows mean log daily earnings for the period of 10–12 years after application. The graph exhibits less variation than that in Panel B of Figure 6 because
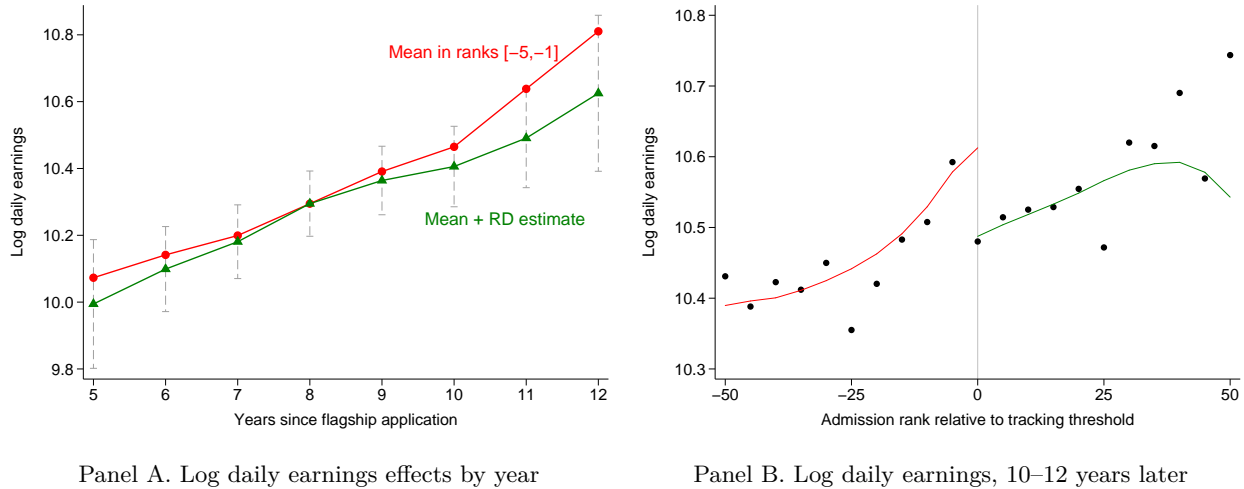
Panel A. Log daily earnings effects by year   Panel B. Log daily earnings, 10–12 years later

FIGURE 7. Tracking effects on earnings

Notes: Panel A shows log daily earnings for flagship applicants (*y*-axis) by years since application (*x*-axis). The graph is defined as in Panel A of Figure 6 with log daily earnings in that year as the dependent variable.

Panel B contains an RD graph (as in Figure 1) with log daily earnings as the dependent variable. The sample includes all flagship applicants and any earnings observations we observe 10–12 years after flagship application.

admission ranks are more related to daily earnings than to flagship graduation. Thus there is clear evidence of a decrease in earnings at the tracking threshold.

Columns (D)–(E) in Table 6 show the RD estimates that correspond to Figure 7. The pooled estimate for daily earnings measured 5–9 years later is negative but small in magnitude (D)). After 10–12 years, crossing the tracking threshold led to an 11 percent decrease in earnings on average, and this estimate is significant at the ten percent level (column (E)). The point estimates are largest in architecture and business programs, consistent with the grade and graduation effects. This includes an earnings decrease of more than 0.35 log points from admission to the higher-ability architecture cohorts.

Appendix Table A14 shows that tracking did not significantly affect the formal employment rate. We observe earnings only for individuals employed at firms that were registered with the government agency that provided our labor market data. Roughly 30 percent of applicants do not appear in these records; this includes both informal workers and individuals who are out of the labor force. However, the probability of formal employment does not change significantly at the tracking threshold for either of the time periods in Table 6. We find no discontinuity in part because most students worked during college, and graduation had little impact on employment rates. We also find no mean effect on the number of days of work, but in architecture programs, crossing the tracking threshold led to a decrease of roughly 70 days of employment per year. This is unlikely to change the sign of the earnings estimate because wages are typically much lower in the informal sector.

Column (F) in Table 6 shows effects on total annual earnings measured 10–12 years after flagship application. Applicants just below the tracking threshold earned $8,000 per year on average in 2012 U.S. dollars. Crossing the threshold decreased earnings by about $900. The magnitude of this effect is similar to the daily earnings effect in column (E), but the variation in hours worked makes the annual estimates less precise. Admission to higher-ability architecture cohorts reduced annual earnings by more than 40 percent on average, although this estimate has a large standard error.

From Proposition 3(b), the earnings results suggest a negative peer effect on skill accumulation. Our model allows classmates to have both information and skill externalities, but information mechanisms alone would lead to positive earnings estimates. Exposure to more able classmates can only reduce earnings if it causes individuals to gain less skill. The affected flagship students may have had lower skill because they learned less in higher-ability classrooms, or because they were less likely to take upper-level courses. The last section presents results that help distinguish between learning and persistence mechanisms.

4.3. **Tracking effects on skill for flagship graduates.** This section shows that exposure to higher-ability classmates reduced skill accumulation even among the population of flagship graduates. This suggests that the negative earnings return to peer ability was due in part to classmate learning externalities, as in Proposition 3(c).

We analyze skill peer effects by estimating tracking effects on upper-level course grades, college *exit* exam scores, and earnings conditional on graduation. Since tracking affected the composition of students for whom we observe these outcomes, we adjust our estimates for selection using a method similar to the Lee (2009) bounds for sample attrition.

We begin by examining upper-level course grades. Tracking may have affected upper-level grades by impacting an individual's learning in the first year, or by altering peer composition in upper-level courses. Appendix Table A15 shows that admission to a fall cohort also increased peer ability in upper-level courses, but this effect is roughly half of the magnitude as for first-year courses because the timing of advanced coursework is more flexible. To isolate the effect of first-year classmates, we exploit this variation in course timing by including fixed effects for upper-level classrooms in our regressions. These dummies absorb any effect of upper-level peers on learning or professor grading; the RD coefficients are identified only from students on either side of the tracking threshold who sat in the same classrooms. The estimates thus reflect only persistent effects of first-year classmates on learning.[26]

---

[26] A student's choice of when to take upper-level courses is often idiosyncratic, but it is not explicitly random. This creates a possible endogeneity when we include classroom fixed effects. Appendix Table A16 shows that our results for architecture and business programs are similar when we exclude these classroom dummies. This suggests that course timing and upper-level peer ability are not important drivers of our results.

TABLE 7. Tracking effects on skill for flagship graduates

	(A)	(B)	(C)	(D)	(E)	(F)
	Upper-level course grade		Exit exam score (SD units)		Log daily earnings after 10–12 years	
	Un-adjusted	Selection adjusted	Un-adjusted	Selection adjusted	Un-adjusted	Selection adjusted
All programs	−0.050	−0.132***	0.032	−0.153	−0.147*	−0.146
	(0.044)	(0.048)	(0.156)	(0.171)	(0.086)	(0.091)
Architecture	−0.101	−0.307**	0.141	−0.241	−1.053***	−1.338***
	(0.105)	(0.130)	(0.351)	(0.394)	(0.252)	(0.351)
Business	−0.048	−0.122**	−0.005	−0.161	−0.038	−0.062
	(0.054)	(0.057)	(0.189)	(0.207)	(0.102)	(0.107)
Engineering	0.022	−0.017	0.145	0.154	−0.131	−0.148
	(0.082)	(0.096)	(0.428)	(0.440)	(0.185)	(0.187)
N	34,637	34,244	484	462	1,210	1,144
Mean, ranks $[-5,-1]$	3.640	3.750	0.758	0.926	10.659	10.687
Classroom dummies	Y	Y				
Selection weights		Y		Y		Y

Notes: This table displays π coefficients from RD regressions (13). In columns (A)–(B), the dependent variable is the student's grade in upper-level required courses, with an observation for each course taken by each student. In columns (C)–(D), the dependent variable is the student's college exit exam score normalized to $(0,1)$ for each exam field and cohort. These regressions include only students whose exam field matches their program at the flagship. The dependent variable in columns (E)–(F) is log daily earnings measured 10–12 years after application. These regressions are similar to column (E) in Table 6, but the sample includes only flagship graduates.

All regressions include only applicants within 30 ranks of the tracking threshold. The first row shows the pooled estimate across all programs. Other rows show estimates by program group. Columns (A)–(B) include classroom fixed effects. Columns (B), (D), and (F) are estimated with selection weights, which we compute separately for each dependent variable and program group using the method as in Figure 8. See Appendix Tables A16–A18 for details.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column (A) in Table 7 shows how tracking affected upper-level grades. We focus on upper-level courses that were required for the program, which we identify using the same method as for first-year required courses.[27] Regressions are at the individual/course level with the student's course grade as the dependent variable. These regressions are similar to equation (13), but they include a fixed effect for each upper-level classroom. The results show that crossing the tracking threshold led to a small decrease in upper-level grades, with negative but insignificant point estimates in architecture and business programs.

The estimates in column (A) are likely biased for the true skill peer effects because tracking affected the composition of students who took upper-level courses. Crossing the tracking threshold increased drop-out rates and thus reduced the probability of taking upper-level

---

[27] Specifically, we define upper-level required courses as those that were taken by more than 75 percent of a program's graduates, and for which the modal graduate took the course after her first year.
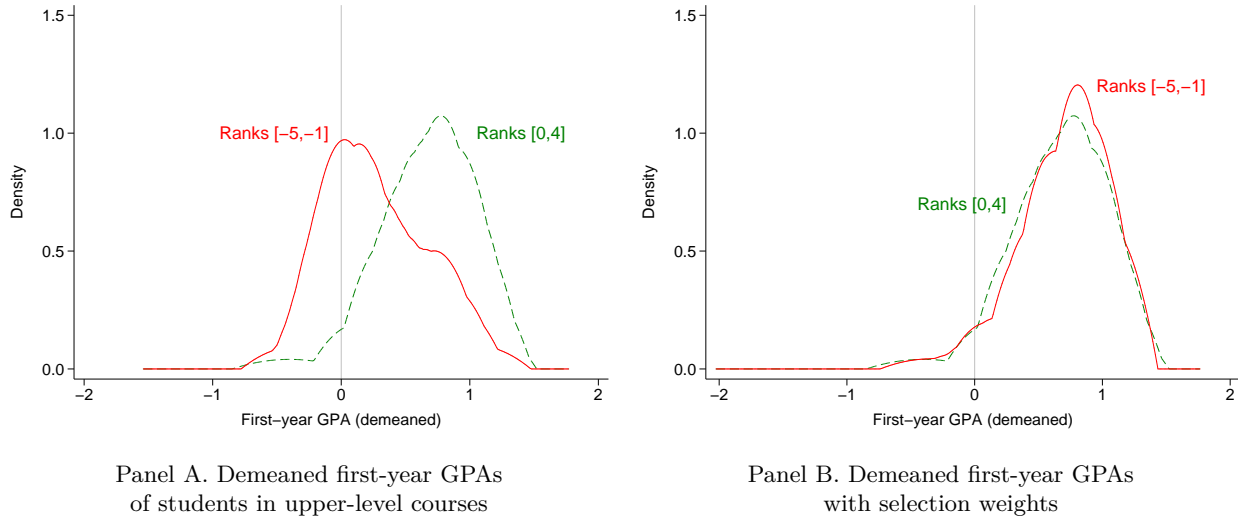
Panel A. Demeaned first-year GPAs
of students in upper-level courses

Panel B. Demeaned first-year GPAs
with selection weights

FIGURE 8. Selection correction in upper-level courses — Architecture programs

Notes: Panel A plots kernel densities of first-year GPA for architecture students in upper-level required courses. The solid line is the GPA distribution for students with admission ranks $[-5, -1]$. The dashed line is the GPA distribution for students with admission ranks $[0, 4]$. We include an observation for each upper-level required course taken by each student. We demean first-year GPA so that both distributions would be mean zero if all students took all courses.

In Panel B, the dashed line is replicated from Panel A. The solid line is the GPA distribution for admission ranks $[-5, -1]$ computed with weights that equal the ratio of the rank [0,4] and rank $[-5, -1]$ densities in Panel A.

courses. This can cause *upward* bias in the estimated grade effects because more able students are more likely to persist in the program. In other words, above-threshold students who took upper-level courses must have had especially high ability to persist despite tracking.

To address this bias, we use a selection-adjustment procedure in the spirit of Lee (2009). Lee derives bounds for a treatment effect when treatment induces data censoring. This matches our setting because tracking affected the likelihood that we observe upper-level grades. Following his insight, we correct for selection by re-weighting the distributions of ability in upper-level courses to be the same on both sides of the tracking threshold.[28]

Figure 8 illustrates our selection adjustment in architecture programs. Panel A plots distributions of ability for students who took upper-level courses, where we use first-year GPA as our measure of ability. The solid line is the ability distribution for students just below the tracking threshold (admission ranks -5 to -1) pooled across all upper-level courses. The dashed line is the ability distribution for above-threshold students (ranks 0 to 4). We demean first-year GPA so that it would be mean zero in both groups if all students stayed in the program. The rightward shift of the dashed line shows that above-threshold students who took upper-level courses were positively selected. In other words, the above-threshold architecture students who persisted in college had especially high first-year GPAs.

---

[28] Our approach differs from Lee's because our goal is to compute a point estimate for the treatment effect without selection, while Lee uses quantiles of the outcome distribution to bound the treatment effect.

Panel B illustrates our correction procedure. We adjust for selection by re-weighting the below-threshold ability distribution to match the above-threshold density. These weights are equal to the ratio of the above- and below-threshold densities in Panel A. This reduces weight on below-threshold students with low first-year GPAs, and increases weight on students with high first-year GPAs. Panel B replicates Panel A with the below-threshold density computed using these selection weights. The two ability distributions now coincide.

Column (B) of Table 7 shows selection-adjusted tracking effects on upper-level grades. The regressions are similar to those in column (A), but we weight below-threshold observations by the selection weights. We compute these weights separately for each program group using the method described for Figure 8. Appendix Table A16 provides details on this procedure, and shows that the selection weights eliminate a discontinuity in the mean ability of students who took upper-level courses, as intended.

The results in column (B) show that, after adjusting for selection, mean upper-level grades were 0.13 points lower for marginal admits to high-ability cohorts. The selection-adjusted estimates are similar to those for first-year grades (column (C) in Table 4), with effects appearing only in architecture and business programs. The selection weights have less impact in engineering because tracking did not significantly affect upper-level course taking.

The remaining columns in Table 7 show how tracking affected two other measures of end-of-college skill. In columns (C)–(D), the dependent variable is a field-specific college *exit* exam score. The Colombian exit exam is a national standardized test of skills that educators deem to be important in each major (MacLeod et al., 2017). These estimates are underpowered because the exam was optional at the time of our data, and only about one-third of students in our sample took it. In columns (E)–(F), the dependent variable is log daily earnings 10–12 years after application, but we restrict the sample to flagship graduates. Since tracking affected the probability that we observe both outcomes, we compute selection weights using the same method as for upper-level grades. Columns (C) and (E) present unadjusted tracking effects, and columns (D) and (F) present estimates with selection weights. Appendix Tables A17–A18 provide details.

The exit exam and earnings results provide further evidence that admission to the higher-ability cohorts reduced end-of-college skill. There is no discontinuity in unadjusted exit scores, but crossing the tracking threshold led to a 0.15 standard deviation decrease in exam performance after adjusting for selection (column (D)). We cannot draw strong conclusions from this estimate because of the small sample size. Yet we find similar results for earnings conditional on graduation. Crossing the tracking threshold reduced earnings by roughly 15 percent within the population of flagship graduates, and the results are similar with and without selection weights (column (E)–(F)). The negative earnings effects are especially large in architecture, consistent with the results for other outcomes.

The findings in Table 7 suggest that flagship classmates affected individual learning beyond their impact on college persistence. Proposition 3(c) states that exposure to more able classmates can only reduce earnings within the population of graduates if there are negative learning externalities. This corroborates the evidence from Section 3 that tracking effects on first-year grades reflect learning spillovers in addition to professor grading responses. In sum, our data suggest that the negative returns to peer ability arose at least in part because individuals learned less in the presence of more able classmates.

5. CONCLUSION

A common argument against affirmative action in college admissions is the "mismatch hypothesis," which states that disadvantaged students may actually be better off at less selective schools (Sander and Taylor Jr, 2012). In a review of research on this topic, Arcidiacono and Lovenheim (2016) note that economists should be naturally skeptical of this argument. Policies like affirmative action merely expand the choice set of disadvantaged applicants; they can always choose to attend other colleges. A necessary condition for mismatch, therefore, is that students have uncertainty about their returns at different colleges.

This paper showed evidence of mismatch in an environment with significant uncertainty on the prospects for college success. Our empirical setting was a selective flagship university in Colombia where nearly half of all students drop out. In this context, we find that students were more likely to fail courses when their own ability was below that of most of their classmates. This had long-term consequences for individuals' careers; the affected students were less likely to obtain a college degree, and had lower earnings one decade later.

Our model also highlighted the crucial role of uncertainty in an individual's college choice. In the model, all students have positive returns to the college program *ex ante*, but returns can be negative *ex post* for students who discover that they are relatively low ability. Mismatch arises if students learn less in the presence of more able peers, but this can only occur if students do not perfectly know their standing in the classroom at the time of enrollment.

Our results do not necessarily imply that lower-ability students can be worse off at selective colleges. Our research design nets out quality benefits of selective schools by comparing students who attended the same college programs. But our findings show that match effects are important, and they suggest that mismatch is more likely when success is less certain. If disadvantaged students are less informed about their suitability for college coursework, concerns about relative ability may partially explain why they are often less likely to apply to selective colleges (Hoxby and Avery, 2013; Dillon and Smith, 2017). Further research on the relative returns to peer ability and school quality is important for evaluating policies that aim to reduce inequality through selective college admissions.

References

Altonji, J. G. and C. R. Pierret (2001). Employer learning and statistical discrimination. *The Quarterly Journal of Economics 116*(1), 313–350.

Andrews, R. J., J. Li, and M. F. Lovenheim (2016). Quantile treatment effects of college quality on earnings. *Journal of Human Resources 51*(1), 200–238.

Arcidiacono, P., E. Aucejo, P. Coate, and V. J. Hotz (2014). Affirmative action and university fit: Evidence from Proposition 209. *IZA Journal of Labor Economics 3*(1), 1.

Arcidiacono, P., E. M. Aucejo, H. Fang, and K. I. Spenner (2011). Does affirmative action lead to mismatch? a new test and evidence. *Quantitative Economics 2*(3), 303–333.

Arcidiacono, P., E. M. Aucejo, and V. J. Hotz (2016). University differences in the graduation of minorities in STEM fields: Evidence from California. *American Economic Review 106*(3), 525–562.

Arcidiacono, P. and M. Lovenheim (2016). Affirmative action and the quality-fit tradeoff. *Journal of Economic Literature 54*(1), 3–51.

Backes, B. (2012). Do affirmative action bans lower minority college enrollment and attainment? evidence from statewide bans. *Journal of Human Resources 47*(2), 435–455.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education.* University of Chicago Press.

Bedard, K. and E. Dhuey (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics 121*(4), 1437–1472.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2011). Too young to leave the nest? the effects of school starting age. *The Review of Economics and Statistics 93*(2), 455–467.

Booij, A. S., E. Leuven, and H. Oosterbeek (2017). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies 84*(2), 547–578.

Bound, J., M. F. Lovenheim, and S. Turner (2010). Why have college completion rates declined? an analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics 2*(3), 129–57.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Canaan, S. and P. Mouganie (2018). Returns to education quality for low-skilled students: Evidence from a discontinuity. *Journal of Labor Economics 36*(2), 395–436.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica 69*(5), 1127–1160.

Carrell, S. E., R. L. Fullerton, and J. E. West (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics 27*(3), 439–464.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica 81*(3), 855–882.

Cascio, E. U. and D. W. Schanzenbach (2016). First in the class? age and the education production function. *Education Finance and Policy 11*(3), 225–250.

Chetty, R., J. N. Friedman, E. Saez, N. Turner, and D. Yagan (2017). Mobility report cards: The role of colleges in intergenerational mobility. Technical report, National Bureau of Economic Research.

Cohodes, S. R. and J. S. Goodman (2014). Merit aid, college quality, and college completion: Massachusetts' Adams scholarship as an in-kind subsidy. *American Economic Journal:*

Applied Economics 6(4), 251–285.

Cooper, H., B. Nye, K. Charlton, J. Lindsay, and S. Greathouse (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research 66*(3), 227–268.

Cortes, K. E. (2010). Do bans on affirmative action hurt minority students? evidence from the Texas Top 10% Plan. *Economics of Education Review 29*(6), 1110–1124.

Cortes, K. E. and J. A. Lincove (2019). Match or mismatch? automatic admissions and college preferences of low-and high-income students. *Educational Evaluation and Policy Analysis 41*(1), 98–123.

Dale, S. B. and A. B. Krueger (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics 117*(4), 1491–1527.

Dale, S. B. and A. B. Krueger (2014). Estimating the effects of college characteristics over the career using administrative earnings data. *Journal of Human Resources 49*(2), 323–358.

DeGroot, M. H. (2004). *Optimal Statistical Decisions.* McGraw-Hill.

Deming, D. J. and C. R. Walters (2018). The impact of state budget cuts on US postsecondary attainment. Harvard University working paper.

Dillon, E. W. and J. A. Smith (2017). Determinants of the match between student ability and college quality. *Journal of Labor Economics 35*(1), 45–66.

Dillon, E. W. and J. A. Smith (2018). The consequences of academic match between students and colleges. NBER Working Paper No. 25069.

Duflo, E., P. Dupas, and M. Kremer (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review 101*(5), 1739–1774.

Elsner, B., I. E. Isphording, and U. Zölitz (2018). Achievement rank affects performance and major choices in college. Working Paper.

Farber, H. S. and R. Gibbons (1996). Learning and wage dynamics. *The Quarterly Journal of Economics 111*(4), 1007–1047.

Foster, G. (2006). It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism. *Journal of Public Economics 90*(8-9), 1455–1475.

Goodman, J., M. Hurwitz, and J. Smith (2014). College access, initial college choice and degree completion. NBER Working Paper 20996.

Hastings, J. S., C. A. Neilson, and S. D. Zimmerman (2013). Are some degrees worth more than others? Evidence from college admission cutoffs in Chile. National Bureau of Economic Research Working Paper 19241.

Hinrichs, P. (2012). The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities. *Review of Economics and Statistics 94*(3), 712–722.

Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics 91*(4), 717–724.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. National Bureau of Economic Research.

Hoxby, C. M. (2018). The productivity of U.S. postsecondary institutions. In C. M. Hoxby and K. Stange (Eds.), *Productivity in Higher Education.* National Bureau of Economic

Research.

Hoxby, C. M. and C. Avery (2013). Missing one-offs: The hidden supply of high-achieving, low-income students. *Brookings Papers on Economic Activity*.

Hoxby, C. M. and G. Weingarth (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Working Paper.

Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy 87*(5), 972–90.

Kirkebøen, L., E. Leuven, and M. Mogstad (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics 131*(3), 1057–1111.

Lavy, V., O. Silva, and F. Weinhardt (2012). The good, the bad, and the average: Evidence on ability peer effects in schools. *Journal of Labor Economics 30*(2), 367–414.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies 76*(3), 1071–1102.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*, 281–355.

Lyle, D. S. (2007). Estimating and interpreting peer and role model effects from randomly assigned social groups at West Point. *The Review of Economics and Statistics 89*(2), 289–299.

MacLeod, W. B., E. Riehl, J. E. Saavedra, and M. Urquiola (2017, July). The big sort: College reputation and labor market outcomes. *American Economic Journal: Applied Economics 9*(3), 223–261.

MacLeod, W. B. and M. Urquiola (2015). Reputation and school competition. *American Economic Review 105*(11), 3471–3488.

MacLeod, W. B. and M. Urquiola (2018). Is education consumption or investment? implications for the effect of school competition. NBER Working Paper No. 25117.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies 60*(3), 531–542.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Murphy, R. and F. Weinhardt (2018). Top of the class: The importance of ordinal rank. NBER Working Paper No. 24958.

Ribas, R. P., B. Sampaio, and G. Trevisan (2018). Can better peers signal less success? the disruptive effect of perceived rank on career investment. Working Paper.

Riehl, E. (2015). Time gaps in academic careers. Columbia Center for Development Economics and Policy (CDEP) and Center on Global Economic Governance (CGEG) Working Paper No. 20.

Riehl, E. (2019). Fairness in college admission exams: From test score gaps to earnings inequality. Working Paper.

Saavedra, J. E. (2009). The returns to college quality: A regression discontinuity analysis. Harvard University.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics 116*(2), 681–704.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far. *Handbook of the Economics of Education 3*(3), 249–277.

Sander, R. and S. Taylor Jr (2012). *Mismatch: How Affirmative Action Hurts Students It's Intended to Help, and Why Universities Won't Admit It.* Basic Books.

Smith, J., M. Pender, and J. Howell (2013). The full extent of student-college academic undermatch. *Economics of Education Review 32*, 247–261.

Stinebrickner, R. and T. 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), 1435–1454.

Wiswall, M. and B. Zafar (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies 82*(2), 791–824.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics 85*(1), 9–23.

Zimmerman, S. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics 32*(4), 711–754.

Zimmerman, S. D. (2019). Elite colleges and upward mobility to top jobs and top incomes. *American Economic Review 109*(1), 1–47.

Appendix — For Online Publication

A. Appendix figures and tables

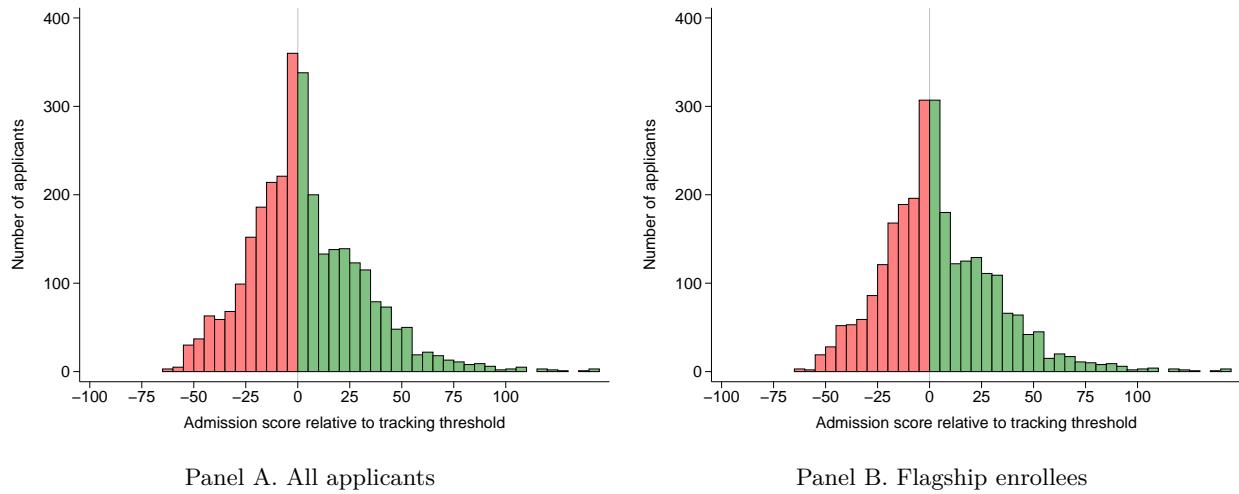Panel A. All applicants   Panel B. Flagship enrollees

FIGURE A1. Density of admission scores relative to the tracking threshold

Notes: The x-axis is a student's admission score (e.g., column (B) in Table 1) normalized so that zero is the first rank above the tracking threshold. The y-axis shows the number of applicants within five unit bins.

Panel A shows the distribution of admission scores for all flagship applicants in Table 2. Using the McCrary (2008) density test, the estimated discontinuity—i.e., the log difference in height at the threshold—is 0.011 with a standard error of 0.073. Panel B shows the distribution of admission scores for students who enrolled in the flagship. The estimated density discontinuity is 0.054 with a standard error of 0.078.
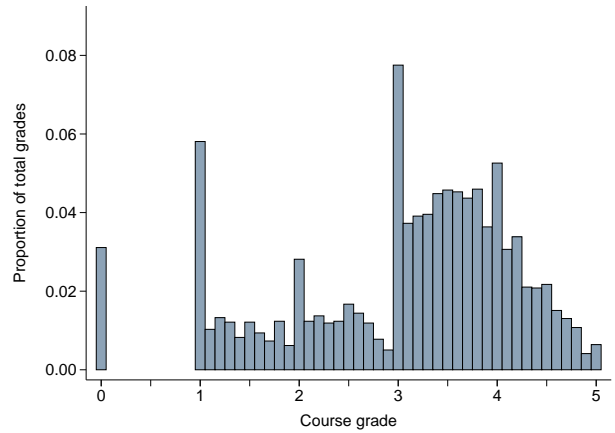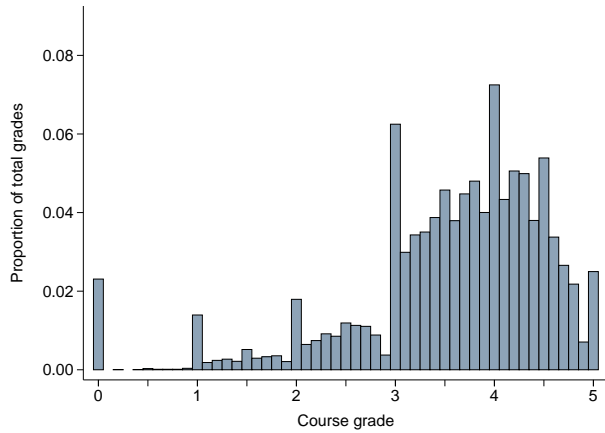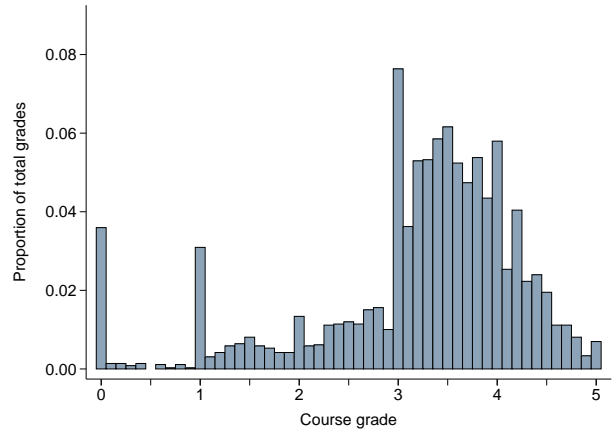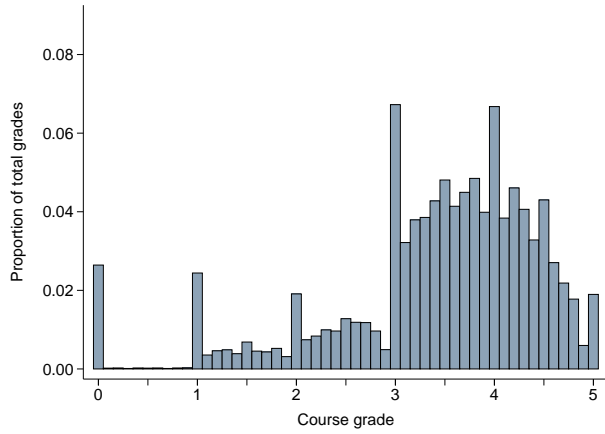
Panel A. All programs

Panel B. Architecture

Panel C. Business

Panel D. Engineering

FIGURE A2. Grade distributions in first-year required courses

Notes: This figure shows grade distributions in first-year required courses at the flagship. We define first-year required courses as described in the notes to Table 3. The graphs include an observation for each individual's first attempt at each course. Colombian college grades are on a 0–5 scale at 0.1 point increments, with 3 or above denoting a passing grade. The height of each bar is the number of grades for each 0.1 point increment as a proportion of all grades.
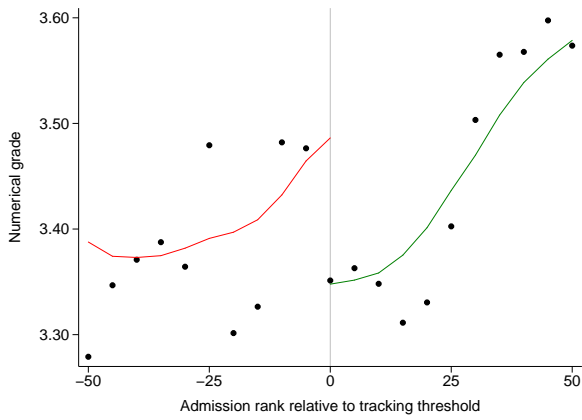
Panel A shows the grade distribution in all courses across all programs. Panels B–D show the grade distributions for courses in architecture, business, and engineering programs.
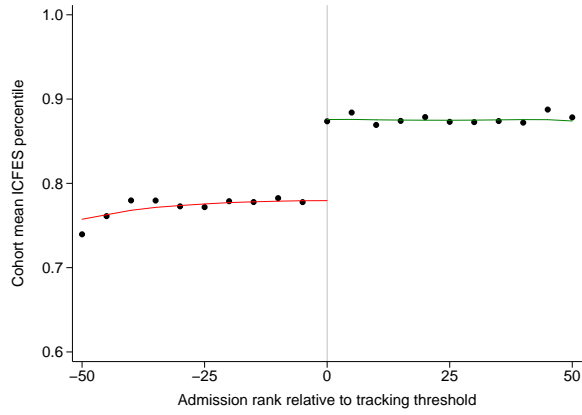
A. Cohort mean ICFES percentile
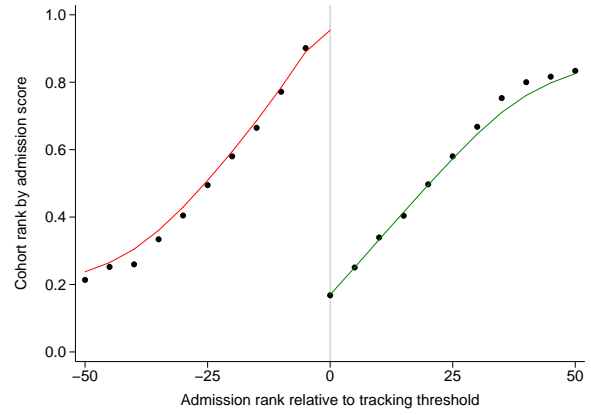
B. Rank in cohort by admission score

C. Numerical grade in first-year courses

D. Passing grade in first-year courses

E. Ever graduated from flagship

F. Log daily earnings after 10–12 years

Figure A3. Main outcomes — All programs

Notes: This figure contains RD graphs (as in Figure 1) for the dependent variables in the panel titles. See Figures 2, 3, 6, and 7 for variable definitions. The sample for Panels A and F includes flagship applicants to all programs. Panels B–E include the subset of these applicants who enrolled in the flagship.
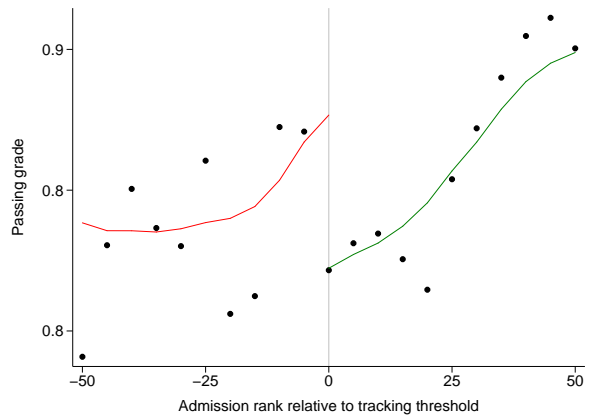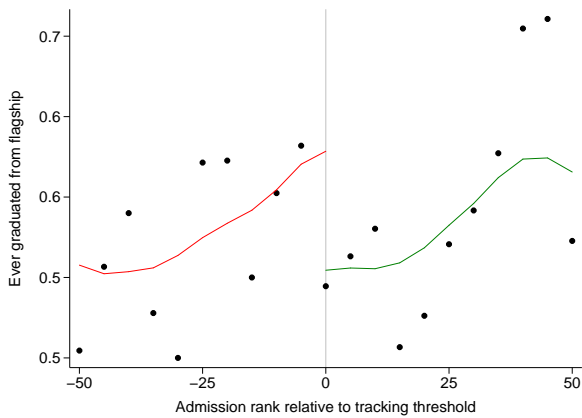
A. Cohort mean ICFES percentile
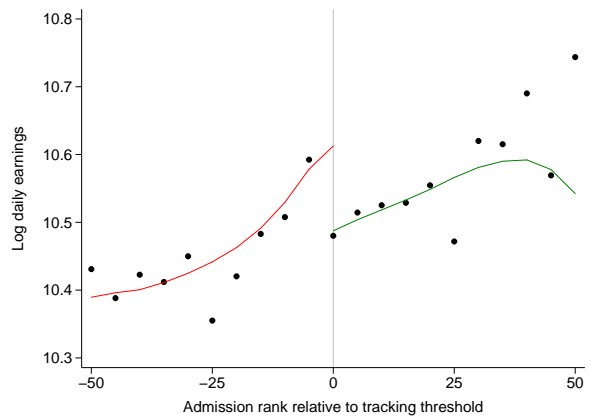
B. Rank in cohort by admission score

C. Numerical grade in first-year courses

D. Passing grade in first-year courses

E. Ever graduated from flagship
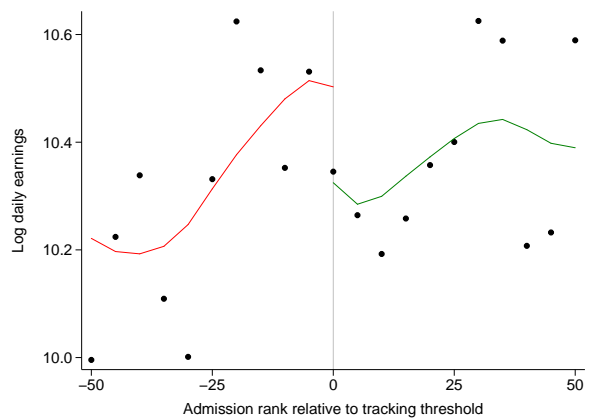
F. Log daily earnings after 10–12 years

Figure A4. Main outcomes — Architecture

Notes: This figure contains RD graphs (as in Figure 1) for the dependent variables in the panel titles. See Figures 2, 3, 6, and 7 for variable definitions. The sample for Panels A and F includes flagship applicants to architecture programs. Panels B–E include the subset of these applicants who enrolled in the flagship.
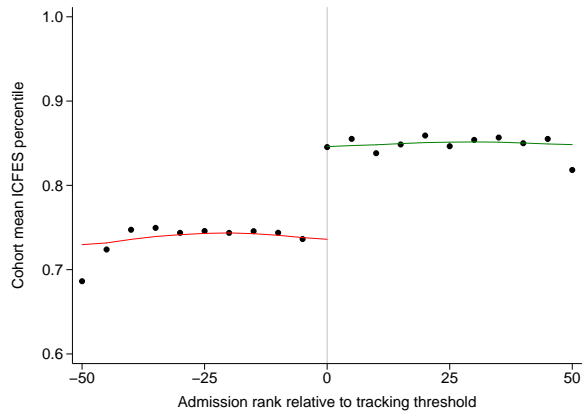
A. Cohort mean ICFES percentile
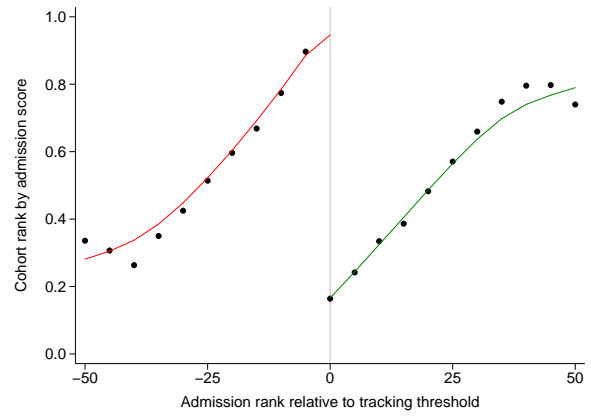
B. Rank in cohort by admission score
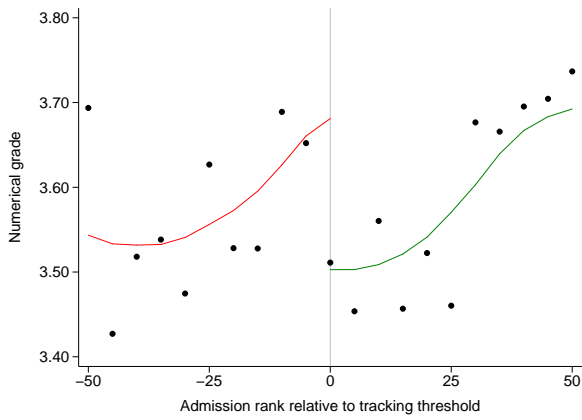
C. Numerical grade in first-year courses
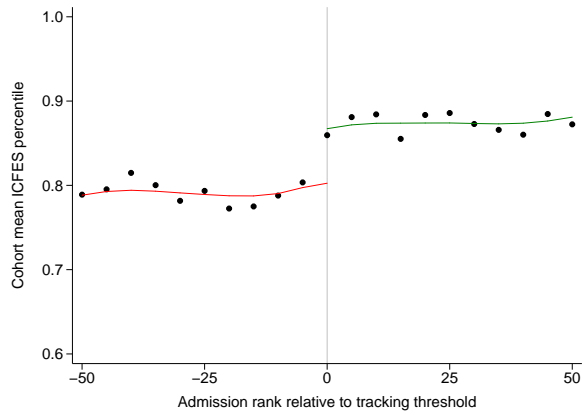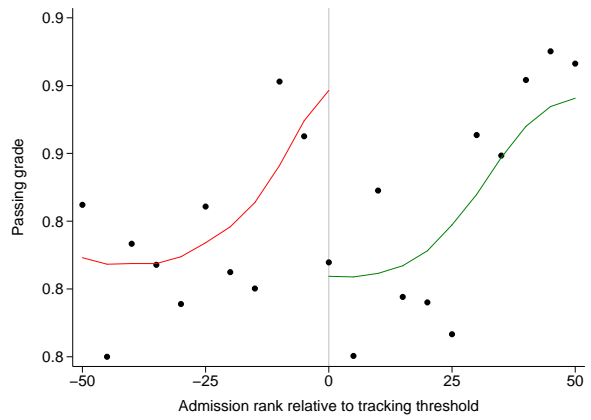
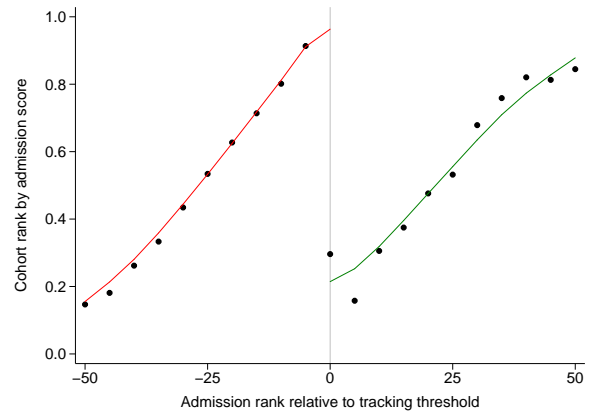D. Passing grade in first-year courses
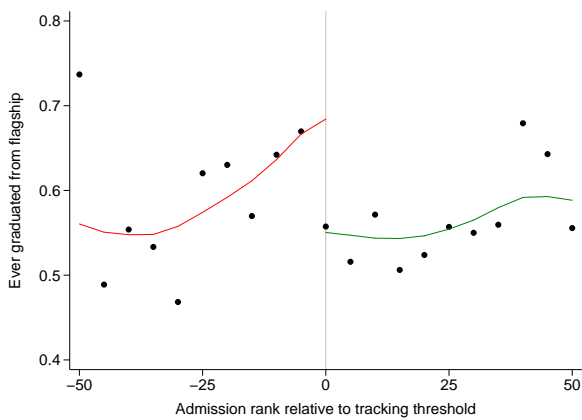
E. Ever graduated from flagship
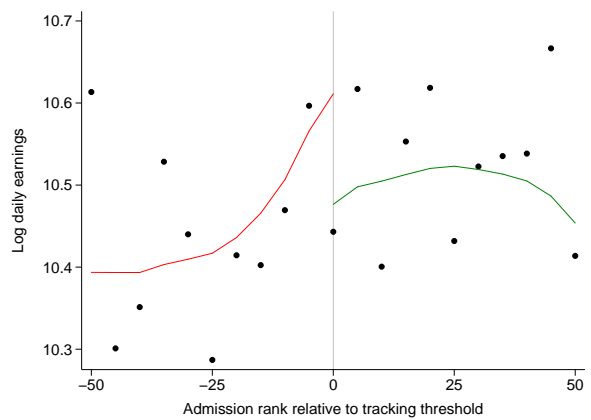
F. Log daily earnings after 10–12 years

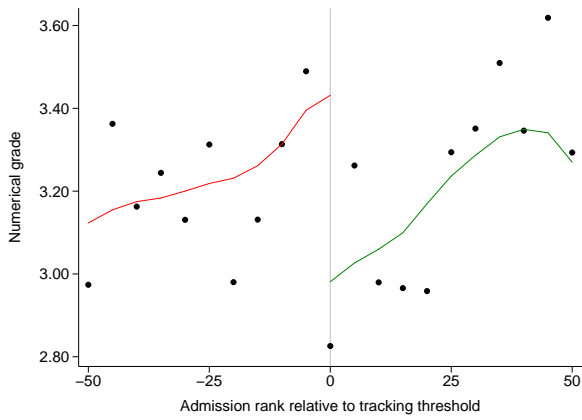FIGURE A5. Main outcomes — Business

Notes: This figure contains RD graphs (as in Figure 1) for the dependent variables in the panel titles. See Figures 2, 3, 6, and 7 for variable definitions. The sample for Panels A and F includes flagship applicants to business programs. Panels B–E include the subset of these applicants who enrolled in the flagship.
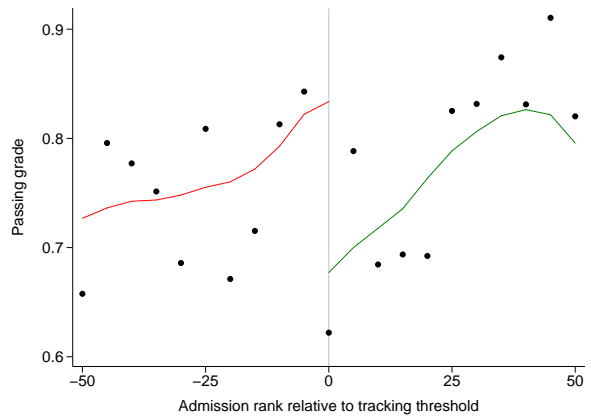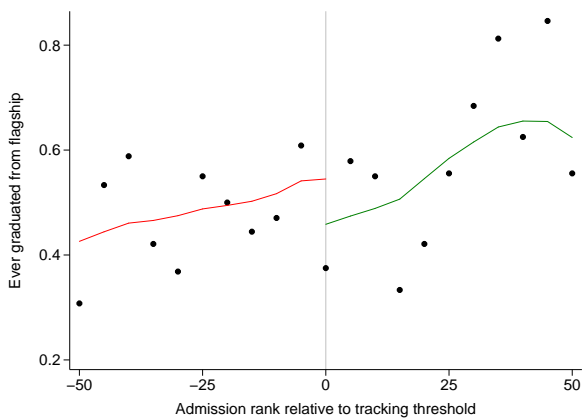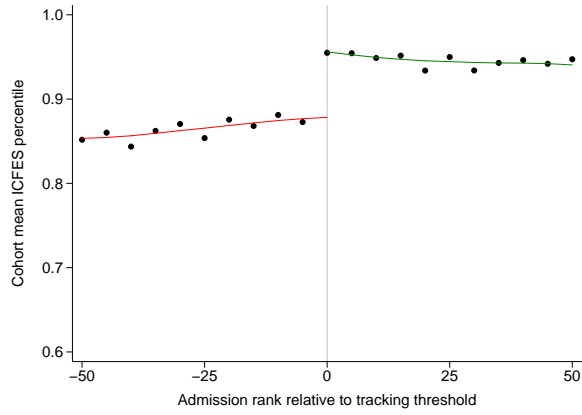
45

A. Cohort mean ICFES percentile
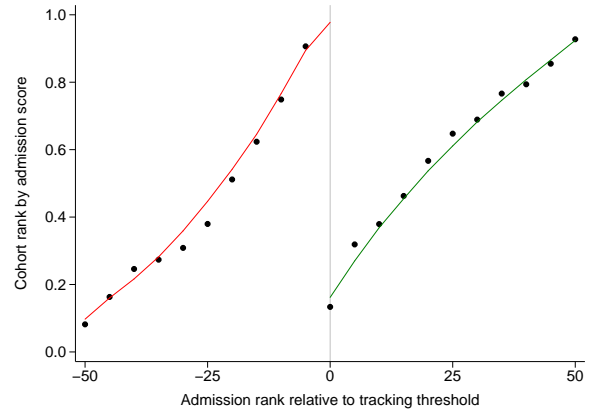
B. Rank in cohort by admission score
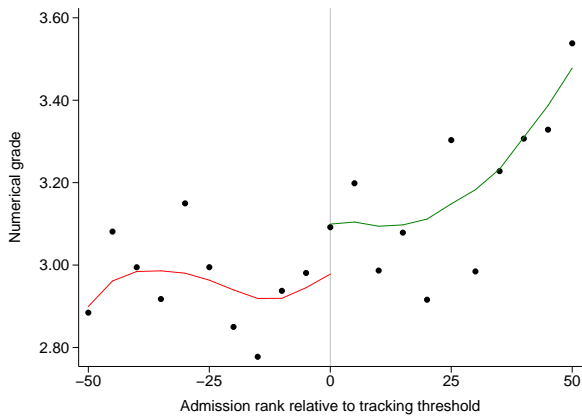
C. Numerical grade in first-year courses
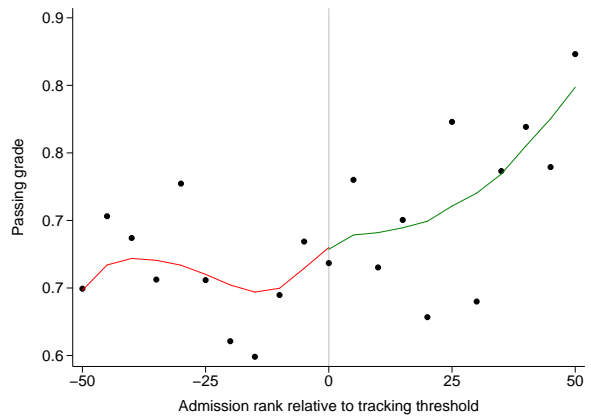
D. Passing grade in first-year courses
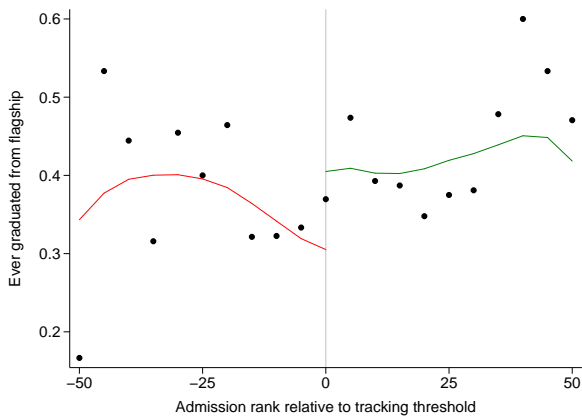
E. Ever graduated from flagship
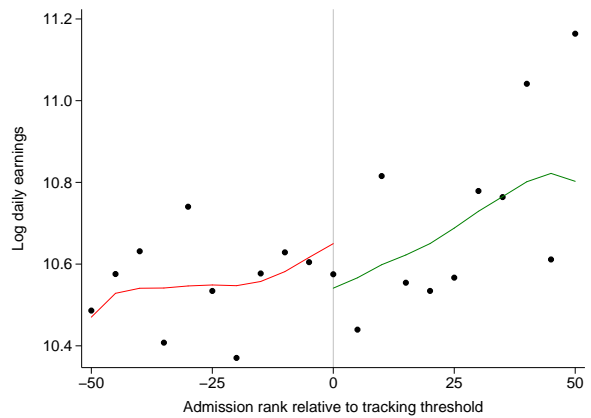
F. Log daily earnings after 10–12 years

FIGURE A6. Main outcomes — Engineering

Notes: This figure contains RD graphs (as in Figure 1) for the dependent variables in the panel titles. See Figures 2, 3, 6, and 7 for variable definitions. The sample for Panels A and F includes flagship applicants to engineering programs. Panels B–E include the subset of these applicants who enrolled in the flagship.

TABLE A1. Balance tests — All applicants

	(A)	(B)	(C)	(D)	(E)	(F)
		Mean, ranks	RD coefficients by program group			
	N	$[-5, -1]$	All	Arc	Bus	Eng
Predicted GPA	1,338	3.382	0.001	0.001	−0.011	0.030
			(0.015)	(0.043)	(0.021)	(0.021)
Female	1,865	0.395	0.048	0.117	0.032	0.052
			(0.039)	(0.112)	(0.052)	(0.061)
Age	1,856	20.106	0.038	0.348	0.141	−0.405
			(0.258)	(0.567)	(0.361)	(0.402)
Years since HS graduation	1,694	1.720	0.066	0.334	0.022	0.052
			(0.170)	(0.476)	(0.216)	(0.342)
ICFES percentile	1,759	0.862	−0.003	0.039	−0.017	0.013
			(0.010)	(0.031)	(0.014)	(0.013)
Secondary educated mother	1,733	0.647	−0.032	−0.065	0.003	−0.115
			(0.041)	(0.099)	(0.054)	(0.082)
College educated mother	1,733	0.267	−0.005	0.215*	−0.038	−0.041
			(0.038)	(0.124)	(0.043)	(0.089)
Secondary educated father	1,474	0.675	−0.059	−0.005	−0.044	−0.125
			(0.047)	(0.118)	(0.062)	(0.090)
College educated father	1,474	0.344	−0.009	0.169	−0.031	−0.064
			(0.046)	(0.138)	(0.055)	(0.096)
Family income > 2x min wage	1,739	0.316	−0.041	0.001	−0.031	−0.084
			(0.040)	(0.125)	(0.048)	(0.092)
p value: Jointly zero			0.859	0.159	0.936	0.683

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the row header. Predicted GPA is the fitted value from a regression of first-year GPA on all other covariates in the table. The sample includes flagship applicants within 30 ranks of the tracking threshold.

Column (A) shows the number of non-missing observations across all programs. Column (B) shows the mean of the dependent variable for students with admission ranks $[-5, -1]$. Column (C) shows the pooled RD coefficient across all programs. Columns (D)–(F) rows show RD coefficients by program group.

The last row reports p values from F tests that the coefficients on all covariates (except predicted GPA) are equal to zero.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	(A)	(B)	(C)	(D)	(E)	(F)
		Mean, ranks	RD coefficients by program group			
	N	$[-5, -1]$	All	Arc	Bus	Eng
Predicted GPA	1,227	3.378	0.011 (0.016)	0.049 (0.045)	−0.007 (0.021)	0.038 (0.023)
Female	1,665	0.395	0.067 (0.042)	0.213* (0.127)	0.034 (0.055)	0.080 (0.066)
Age	1,663	19.900	0.074 (0.275)	0.176 (0.619)	0.246 (0.378)	−0.483 (0.431)
Years since HS graduation	1,563	1.743	−0.021 (0.175)	−0.252 (0.324)	−0.016 (0.225)	0.109 (0.394)
ICFES percentile	1,623	0.859	0.001 (0.011)	0.050 (0.031)	−0.013 (0.015)	0.014 (0.015)
Secondary educated mother	1,594	0.641	−0.017 (0.044)	−0.111 (0.111)	0.026 (0.056)	−0.094 (0.088)
College educated mother	1,594	0.269	−0.005 (0.040)	0.171 (0.137)	−0.030 (0.044)	−0.034 (0.096)
Secondary educated father	1,313	0.652	−0.036 (0.050)	0.034 (0.135)	−0.024 (0.065)	−0.106 (0.098)
College educated father	1,313	0.312	0.020 (0.048)	0.278* (0.151)	−0.023 (0.058)	−0.016 (0.104)
Family income > 2x min wage	1,602	0.317	−0.021 (0.042)	0.010 (0.134)	−0.006 (0.049)	−0.071 (0.097)
p value: Jointly zero			0.878	0.110	0.963	0.690

Notes: This table is similar to Appendix Table A1, but the sample includes only students who enrolled in the flagship. The table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the row header. Predicted GPA is the fitted value from a regression of first-year GPA on all other covariates in the table. The sample includes flagship enrollees within 30 ranks of the tracking threshold.

Column (A) shows the number of non-missing observations across all programs. Column (B) shows the mean of the dependent variable for students with admission ranks $[-5, -1]$. Column (C) shows the pooled RD coefficient across all programs. Columns (D)–(F) rows show RD coefficients by program group.

The last row reports p values from F tests that the coefficients on all covariates (except predicted GPA) are equal to zero.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A3. Tracking effects on cohort enrollment and peer characteristics

	(A)	(B)	(C)	(D)
	Enrolled in fall cohort	Enrolled in either cohort	Cohort mean ICFES percentile	Cohort rank by admission score
All programs	0.843***	0.012	0.095***	−0.808***
	(0.020)	(0.030)	(0.004)	(0.016)
Architecture	0.730***	−0.062	0.073***	−0.797***
	(0.070)	(0.097)	(0.010)	(0.056)
Business	0.863***	0.003	0.107***	−0.806***
	(0.024)	(0.035)	(0.005)	(0.020)
Engineering	0.853***	0.073	0.076***	−0.831***
	(0.041)	(0.064)	(0.008)	(0.024)
N	1,865	1,865	1,803	1,665
Mean, ranks $[-5,-1]$	0.000	0.854	0.778	0.901

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the column header. In column (C), ICFES percentile is an individual's rank among all test takers in her cohort using the mean score across subjects. The sample for columns (A)–(C) includes all flagship applicants. The sample for column (D) includes flagship enrollees.

The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A4. Summary statistics for first-year required courses

Group	Program	(A) Total # students	(B) # courses per student	(C) # students in classroom	(D) % of classmates from own cohort	(E) Mean grade	(F) % pass
Arch	Architecture	394	9.1	37.1	0.64	3.15	0.75
Bus	Accounting (day)	519	9.8	41.6	0.81	3.48	0.81
	Accounting (night)	356	8.6	37.1	0.77	3.34	0.77
	Business Admin (day)	513	9.8	37.4	0.80	3.68	0.87
	Business Admin (night)	349	6.9	34.4	0.80	3.54	0.84
	Foreign Trade	84	8.1	33.1	0.80	3.59	0.85
Eng	Chemical Engineering	107	8.1	32.6	0.80	3.16	0.73
	Electrical Engineering	109	8.4	30.0	0.74	2.99	0.67
	Electronic Engineering	98	9.1	31.7	0.68	3.27	0.78
	Materials Engineering	109	7.3	26.2	0.84	2.90	0.60
	Mechanical Engineering	110	8.2	26.6	0.85	2.87	0.61
	Total	2,748	9.2	36.3	0.77	3.38	0.79

Notes: This table displays summary statistics on first-year required courses for each program in our sample. We define first-year required courses as described in the notes to Table 3.

Column (A) is the total number of students in our sample who took any first-year required course. Column (B) is the average number of first-year required courses taken by each student. Column (C) is the average number of classmates in individual's first-year courses. Column (D) is the average proportion of these students who are from the individual's own cohort, where cohort is defined by program and starting semester. Columns (E) and (F) are the mean grade and mean course passing rate across all first-year required courses.

TABLE A5. Tracking effects by RD bandwidth — All programs

	(A)	(B)	(C)	(D)	(E)	(F)
		Mean, ranks	RD coefficients by bandwidth (in ranks)			
	N	[−5, −1]	45	30	15	CCT
First-year required courses						
Mean ICFES pctile in class	14,267	0.791	0.080*** (0.002)	0.079*** (0.003)	0.080*** (0.004)	0.079*** (0.004)
Class rank by admit score	14,327	0.844	−0.652*** (0.013)	−0.686*** (0.015)	−0.709*** (0.020)	−0.708*** (0.019)
Numerical grade	14,327	3.476	−0.152*** (0.051)	−0.126** (0.061)	−0.221*** (0.084)	−0.089 (0.065)
Passing grade	14,327	0.821	−0.047*** (0.017)	−0.045** (0.020)	−0.086*** (0.027)	−0.037* (0.021)
Educational attainment						
Full-time years at flagship	1,665	3.537	−0.693*** (0.157)	−0.644*** (0.185)	−0.847*** (0.252)	−0.576*** (0.191)
Flagship graduate	1,665	0.582	−0.097** (0.038)	−0.084* (0.044)	−0.130** (0.060)	−0.093** (0.044)
Any college degree	1,865	0.634	−0.085** (0.036)	−0.089** (0.042)	−0.101* (0.057)	−0.049 (0.045)
Earnings						
Log daily earnings after 5–9 years	3,887	10.282	−0.028 (0.036)	−0.023 (0.042)	−0.006 (0.056)	−0.018 (0.051)
Log daily earnings after 10–12 years	2,158	10.592	−0.093 (0.058)	−0.115* (0.067)	−0.132 (0.086)	−0.135* (0.071)
Annual earnings in USD after 10–12 years	2,158	7,987	−610 (630)	−913 (734)	−1,744* (932)	−707 (760)
Skill of flagship graduates (with selection weights)						
Upper-level course grade	34,244	3.750	−0.146*** (0.040)	−0.132*** (0.048)	−0.135** (0.066)	−0.086 (0.055)
Exit exam score (SD units)	462	0.926	−0.284** (0.136)	−0.153 (0.171)	0.031 (0.203)	0.006 (0.187)
Log daily earnings after 10–12 years	1,144	10.687	−0.174** (0.080)	−0.146 (0.091)	−0.085 (0.122)	−0.138 (0.099)

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the row header. See Tables 4, 6, and 7 for variable and sample definitions. Regression specifications for the last three rows match those in columns (B), (D), and (F) of Table 7.

The sample includes applicants to all programs who are within h ranks of the tracking threshold, where h is given by the header in columns (C)–(F). CCT is the bandwidth recommended by Calonico et al. (2014).

Column (A) shows the number of non-missing observations for the $h = 30$ sample. Column (B) shows the mean of the dependent variable for students with admission ranks [−5, −1]. Columns (C)–(F) report RD coefficients.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A6. Tracking effects by RD bandwidth — Architecture programs

	(A)	(B)	(C)	(D)	(E)	(F)
		Mean, ranks	RD coefficients by bandwidth (in ranks)			
	N	$[-5, -1]$	45	30	15	CCT
First-year required courses						
Mean ICFES pctile in class	1,994	0.808	0.046*** (0.005)	0.042*** (0.006)	0.035*** (0.009)	0.041*** (0.007)
Class rank by admit score	2,002	0.755	−0.609*** (0.038)	−0.587*** (0.050)	−0.513*** (0.085)	−0.482*** (0.097)
Numerical grade	2,002	3.490	−0.398** (0.166)	−0.472** (0.207)	−0.647** (0.296)	−0.647** (0.296)
Passing grade	2,002	0.843	−0.127** (0.061)	−0.166** (0.076)	−0.240** (0.115)	−0.220** (0.108)
Educational attainment						
Full-time years at flagship	219	3.696	−0.903* (0.516)	−1.053* (0.633)	−1.790* (0.927)	−1.485* (0.861)
Flagship graduate	219	0.609	−0.116 (0.113)	−0.098 (0.139)	−0.216 (0.194)	−0.129 (0.180)
Any college degree	255	0.615	−0.198** (0.100)	−0.185 (0.121)	−0.153 (0.161)	−0.167 (0.154)
Earnings						
Log daily earnings after 5–9 years	432	10.194	−0.141 (0.101)	−0.131 (0.125)	0.081 (0.165)	0.081 (0.148)
Log daily earnings after 10–12 years	230	10.531	−0.276 (0.179)	−0.358 (0.217)	−0.069 (0.278)	−0.117 (0.247)
Annual earnings in USD after 10–12 years	230	7,333	−3,232** (1,482)	−3,617** (1,800)	−2,189 (2,509)	−2,773 (1,930)
Skill of flagship graduates (with selection weights)						
Upper-level course grade	3,071	4.065	−0.376*** (0.108)	−0.307** (0.128)	−0.391*** (0.137)	−0.351** (0.138)
Exit exam score (SD units)	88	0.843	−0.331 (0.331)	−0.241 (0.390)	−0.002 (0.515)	−0.366 (0.405)
Log daily earnings after 10–12 years	91	10.785	−1.094*** (0.290)	−1.338*** (0.362)	−1.184** (0.493)	−0.866 (0.789)

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the row header. See Tables 4, 6, and 7 for variable and sample definitions. Regression specifications for the last three rows match those in columns (B), (D), and (F) of Table 7.

The sample includes applicants to architecture programs who are within h ranks of the tracking threshold, where h is given by the header in columns (C)–(F). CCT is the bandwidth recommended by Calonico et al. (2014).

Column (A) shows the number of non-missing observations for the $h = 30$ sample. Column (B) shows the mean of the dependent variable for students with admission ranks $[-5, -1]$. Columns (C)–(F) report RD coefficients.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7. Tracking effects by RD bandwidth — Business programs

	(A)	(B)	(C)	(D)	(E)	(F)
		Mean, ranks	RD coefficients by bandwidth (in ranks)			
	N	$[-5, -1]$	45	30	15	CCT
First-year required courses						
Mean ICFES pctile in class	9,306	0.756	0.091*** (0.003)	0.090*** (0.004)	0.093*** (0.005)	0.091*** (0.004)
Class rank by admit score	9,331	0.868	−0.695*** (0.016)	−0.731*** (0.017)	−0.752*** (0.022)	−0.752*** (0.022)
Numerical grade	9,331	3.652	−0.206*** (0.059)	−0.164** (0.070)	−0.236** (0.098)	−0.186*** (0.067)
Passing grade	9,331	0.865	−0.054*** (0.018)	−0.044** (0.021)	−0.075*** (0.028)	−0.044** (0.021)
Educational attainment						
Full-time years at flagship	1,081	3.857	−0.886*** (0.180)	−0.857*** (0.211)	−0.983*** (0.283)	−0.749*** (0.223)
Flagship graduate	1,081	0.670	−0.140*** (0.047)	−0.137** (0.055)	−0.176** (0.074)	−0.145*** (0.056)
Any college degree	1,179	0.651	−0.108** (0.045)	−0.113** (0.052)	−0.130* (0.070)	−0.079 (0.057)
Earnings						
Log daily earnings after 5–9 years	2,863	10.324	−0.028 (0.042)	−0.042 (0.049)	−0.032 (0.065)	−0.040 (0.059)
Log daily earnings after 10–12 years	1,189	10.597	−0.045 (0.069)	−0.098 (0.082)	−0.137 (0.103)	−0.141 (0.088)
Annual earnings in USD after 10–12 years	1,189	8,250	−288 (801)	−1,121 (971)	−1,689 (1,228)	−1,247 (1,036)
Skill of flagship graduates (with selection weights)						
Upper-level course grade	26,207	3.817	−0.126*** (0.048)	−0.122** (0.056)	−0.140* (0.079)	−0.108* (0.062)
Exit exam score (SD units)	310	0.967	−0.250 (0.164)	−0.161 (0.206)	−0.017 (0.229)	−0.053 (0.226)
Log daily earnings after 10–12 years	727	10.653	−0.075 (0.089)	−0.062 (0.107)	−0.082 (0.144)	−0.037 (0.103)

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the row header. See Tables 4, 6, and 7 for variable and sample definitions. Regression specifications for the last three rows match those in columns (B), (D), and (F) of Table 7.

The sample includes applicants to business programs who are within h ranks of the tracking threshold, where h is given by the header in columns (C)–(F). CCT is the bandwidth recommended by Calonico et al. (2014).

Column (A) shows the number of non-missing observations for the $h = 30$ sample. Column (B) shows the mean of the dependent variable for students with admission ranks $[-5, -1]$. Columns (C)–(F) report RD coefficients.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A8. Tracking effects by RD bandwidth — Engineering programs

	(A)	(B)	(C)	(D)	(E)	(F)
		Mean, ranks	RD coefficients by bandwidth (in ranks)			
	N	$[-5, -1]$	45	30	15	CCT
First-year required courses						
Mean ICFES pctile in class	2,967	0.882	0.068***	0.068***	0.066***	0.066***
			(0.005)	(0.005)	(0.007)	(0.007)
Class rank by admit score	2,994	0.830	-0.563***	-0.625***	-0.676***	-0.676***
			(0.024)	(0.028)	(0.035)	(0.035)
Numerical grade	2,994	2.981	0.142	0.161	-0.001	-0.056
			(0.117)	(0.136)	(0.180)	(0.190)
Passing grade	2,994	0.684	0.023	0.017	-0.051	-0.083
			(0.042)	(0.049)	(0.064)	(0.069)
Educational attainment						
Full-time years at flagship	365	2.595	-0.015	0.167	-0.044	-0.215
			(0.362)	(0.423)	(0.577)	(0.610)
Flagship graduate	365	0.333	0.040	0.075	0.024	0.024
			(0.076)	(0.088)	(0.117)	(0.117)
Any college degree	431	0.600	0.036	0.016	-0.002	0.000
			(0.074)	(0.087)	(0.118)	(0.106)
Earnings						
Log daily earnings after 5–9 years	592	10.166	0.043	0.126	0.028	0.008
			(0.088)	(0.105)	(0.137)	(0.143)
Log daily earnings after 10–12 years	739	10.605	-0.105	-0.063	-0.153	-0.225
			(0.114)	(0.129)	(0.165)	(0.155)
Annual earnings in USD after 10–12 years	739	7,778	-282	229	$-1,806$	$-2,516$
			(1,204)	(1,352)	(1,683)	(1,833)
Skill of flagship graduates (with selection weights)						
Upper-level course grade	4,966	3.179	-0.086	-0.017	0.106	0.082
			(0.085)	(0.104)	(0.145)	(0.116)
Exit exam score (SD units)	64	0.807	-0.107	0.154	0.258	-0.221
			(0.330)	(0.457)	(0.730)	(0.970)
Log daily earnings after 10–12 years	326	10.781	-0.231	-0.148	0.025	-0.018
			(0.180)	(0.187)	(0.228)	(0.227)

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the row header. See Tables 4, 6, and 7 for variable and sample definitions. Regression specifications for the last three rows match those in columns (B), (D), and (F) of Table 7.

The sample includes applicants to engineering programs who are within h ranks of the tracking threshold, where h is given by the header in columns (C)–(F). CCT is the bandwidth recommended by Calonico et al. (2014).

Column (A) shows the number of non-missing observations for the $h = 30$ sample. Column (B) shows the mean of the dependent variable for students with admission ranks $[-5, -1]$. Columns (C)–(F) report RD coefficients.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A9. Tracking effects on first-year grades by age at application

Dependent variable: Numerical grade

	(A)	(B)	(C)	(D)	(E)
	RD coefficient by age at time of application (quartiles)				
	Youngest	2nd quartile	3rd quartile	Oldest	p value: All equal
All programs	-0.076	-0.216^*	-0.032	-0.215^*	0.614
	(0.123)	(0.125)	(0.126)	(0.115)	
Architecture	-0.864^{***}	-0.504	-0.264	0.136	0.136
	(0.261)	(0.450)	(0.482)	(0.341)	
Business	-0.074	-0.253^*	-0.031	-0.460^{***}	0.122
	(0.149)	(0.145)	(0.129)	(0.145)	
Engineering	0.465	0.019	0.136	0.221	0.740
	(0.300)	(0.276)	(0.330)	(0.212)	
N	3,403	3,614	3,629	3,665	
Mean grade, ranks $[-5,-1]$	3.513	3.499	3.297	3.594	
Average age at application	16.7	17.6	18.8	22.5	

Notes: This table displays tracking effects on first-year course grades—as in column (C) of Table 4—by quartile of student age at the time of application. Columns (A)–(D) display π coefficients from separate RD regressions (13) with grades in first-year required courses as the dependent variable. The columns separate the sample into quartiles defined by age at the time of application. Column (E) reports p values from F tests that the coefficients are equal for all four quartiles.

The sample includes flagship enrollees within 30 ranks of the tracking threshold. Regressions are at the individual/course level with an observation for each individual's first attempt at each first-year required course. The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A10. Effect of age on grades in first-year required courses

Dependent variable: Numerical grade

| | (A) | (B) | (C) | (D) | (E) |
	No controls	Within cohort	Within class	Within birth yr	ICFES controls
Age in years × 2	−0.001	−0.003	−0.005**	−0.021	−0.036*
	(0.002)	(0.002)	(0.002)	(0.021)	(0.021)
N	22,164	22,164	22,164	22,164	22,164
R^2	0.000	0.074	0.350	0.358	0.374
Cohort dummies		Y	Y	Y	Y
Classroom dummies			Y	Y	Y
Birth year dummies				Y	Y
ICFES subject scores					Y

Notes: This table displays coefficients from a regression of student grades in first-year required courses on their ages at the start of the course. We use age in years × 2 so that one unit represents six months in age.

The sample includes flagship enrollees in all programs. Regressions are at the individual/course level with an observation for each individual's first attempt at each first-year required course.

Column (A) reports the coefficient on age with no other controls. Column (B) adds cohort fixed effects, which are defined by program and starting semester. Column (C) adds classroom fixed effects. Column (D) adds fixed effects for year of birth. Column (E) adds linear terms for scores on the ICFES national college entrance exam in six subjects (math, language arts, biology, chemistry, physics, and social science).

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A11. Tracking effects on first-year grades by course grade distribution

Dependent variable: Numerical grade

	(A)	(B)	(C)	(D)
	Mean grade in fall & spring cohorts		St. deviation of grades in fall & spring cohorts	
	Different means	Similar means	Different SDs	Similar SDs
All programs	−0.096	−0.179***	−0.198***	−0.106
	(0.075)	(0.069)	(0.069)	(0.073)
Architecture	−0.725***	−0.357	−0.855***	−0.357
	(0.265)	(0.224)	(0.210)	(0.233)
Business	−0.153*	−0.220***	−0.236***	−0.152*
	(0.085)	(0.080)	(0.079)	(0.082)
Engineering	0.265	0.078	0.077	0.214
	(0.167)	(0.151)	(0.163)	(0.151)
N	5,322	7,757	5,329	7,750
Mean, ranks $[-5,-1]$	3.419	3.533	3.660	3.368

Notes: This table displays tracking effects on first-year course grades—as in column (C) of Table 4—by the features of the course's grade distribution defined in Figure 4. The table shows π coefficients from separate RD regressions (13) with the student's grades in first-year required courses as the dependent variable.

Columns (A)–(B) separate first-year required courses based on their mean grade in fall and spring cohort classrooms, as shown in Panel A of Figure 4. Column (A) includes courses where fall and spring cohort classrooms have statistically different mean grades at $p < 0.05$ (hollow symbols). Column (B) includes courses where fall and spring cohort classrooms have statistically indistinguishable mean grades at $p < 0.05$ (solid symbols).

Columns (C)–(D) are similar to columns (A)–(B), but we separate courses based on the standard deviation of grades in fall and spring cohort classrooms, as shown in Panel B of Figure 4.

The sample includes flagship enrollees within 30 ranks of the tracking threshold. Regressions are at the individual/course level with an observation for each individual's first attempt at each first-year required course. The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A12. Tracking effects on first-year course grades by student gender

| | (A) | (B) | (C) | (D) |
| | Numerical grade | | Passing grade | |
	Women	Men	Women	Men
All programs	−0.079	−0.148*	−0.025	−0.054*
	(0.083)	(0.087)	(0.026)	(0.029)
Architecture	−0.191	−0.605**	0.019	−0.260***
	(0.337)	(0.271)	(0.127)	(0.099)
Business	−0.049	−0.297***	−0.021	−0.062*
	(0.090)	(0.110)	(0.026)	(0.035)
Engineering	−0.105	0.217	−0.020	0.032
	(0.255)	(0.151)	(0.120)	(0.053)
N	6,865	7,462	6,865	7,462
Mean, ranks $[−5,−1]$	3.585	3.402	0.842	0.806

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the column header. The sample includes flagship enrollees within 30 ranks of the tracking threshold. Columns (A) and (C) include female students. Columns (B) and (D) include male students.

Regressions are at the individual/course level with an observation for each individual's first attempt at each first-year required course. The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A13. Tracking effects on other schooling outcomes

	(A)	(B)	(C)	(D)
	Flagship program		Other programs	
	Retook first-year course	Took the advanced course	Started another program	Finished another program
All programs	0.041***	−0.062*	0.029	−0.009
	(0.011)	(0.036)	(0.033)	(0.019)
Architecture	0.121***	−0.167	−0.042	−0.053
	(0.042)	(0.126)	(0.095)	(0.066)
Business	0.030***	−0.082*	0.062*	0.002
	(0.011)	(0.042)	(0.038)	(0.015)
Engineering	0.031	0.057	−0.015	−0.023
	(0.029)	(0.080)	(0.082)	(0.058)
N (all programs)	14,327	51,528	1,865	1,865
Mean, ranks $[-5,-1]$	0.074	0.656	0.190	0.063

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the column header. In column (A), the dependent variable is an indicator equal to one if the student took the first-year required course more than once. In column (B), the dependent variable is an indicator equal to one if the student took the upper-level required course at all. The sample for columns (A)–(B) includes flagship enrollees, and regressions include an observation for each potential individual/course pair.

In column (C), the dependent variable is an indicator equal to one if the student enrolled in the flagship program she was admitted to and also enrolled in at least one other college program. In column (D), the dependent variable is an indicator equal to one if the student enrolled in the flagship program she was admitted to and completed another college program. We observe enrollment and graduation and other Colombian colleges through 2012. The sample for columns (C)–(D) includes all flagship applicants.

All regressions include only applicants within 30 ranks of the tracking threshold. The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A14. Tracking effects on formal employment

	(A)	(B)	(C)	(D)
	5–9 years after application		10–12 years after application	
	Has any formal earnings	# days of earnings per year	Has any formal earnings	# days of earnings per year
All programs	−0.024 (0.036)	−1.050 (9.074)	0.038 (0.044)	3.809 (9.985)
Architecture	−0.096 (0.101)	−73.258** (34.178)	−0.031 (0.133)	−62.056** (30.988)
Business	−0.022 (0.044)	9.312 (10.668)	0.022 (0.058)	1.750 (13.378)
Engineering	0.022 (0.075)	−0.767 (17.607)	0.080 (0.080)	25.444 (17.145)
N (all programs)	5,987	3,887	3,338	2,158
Mean, ranks $[-5,-1]$	0.707	267.833	0.691	283.441

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the column header. Regressions include an observation for each year in 5–9 years (columns (A)–(B)) and 10–12 years (columns (C)–(D)) after application using 2008–2012 labor market data. The dependent variable in columns (A) and (C) is an indicator equal to one if the student appears in the dataset in that year. The dependent variable in columns (B) and (D) is the number of days with reported earnings in that year conditional on appearing in the dataset.

The regression sample includes all flagship applicants within 30 ranks of the tracking threshold. The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A15. Tracking effects on peer composition in first-year and upper-level required courses

	(A)	(B)	(C)	(D)
	Mean ICFES percentile in class		Class rank by admission score	
	First-year course	Upper-level course	First-year course	Upper-level course
All programs	0.079*** (0.003)	0.039*** (0.003)	−0.686*** (0.015)	−0.382*** (0.019)
Architecture	0.042*** (0.006)	0.013* (0.007)	−0.587*** (0.049)	−0.246*** (0.059)
Business	0.090*** (0.004)	0.045*** (0.004)	−0.731*** (0.017)	−0.435*** (0.021)
Engineering	0.068*** (0.005)	0.024*** (0.004)	−0.625*** (0.028)	−0.214*** (0.036)
N	14,267	34,047	14,327	34,637
Mean, ranks $[-5,-1]$	0.791	0.817	0.844	0.745

Notes: This table displays π coefficients from separate RD regressions (13) using the dependent variable listed in the column header. The sample includes flagship enrollees within 30 ranks of the tracking threshold. Columns (A) and (C) include first-year required courses. Columns (B) and (D) include upper-level required courses. Regressions are at the individual/course level with an observation for each individual's first attempt at each required course.

The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	(A)	(B)	(C)	(D)	(E)	(F)
	Demeaned first-year GPA of students in upper-level courses			Upper-level course grade		
	Un-adjusted	Selection adjusted	Within class	Un-adjusted	Selection adjusted	Within class
All programs	0.124***	0.023	−0.070	0.043	−0.040	−0.132***
	(0.040)	(0.054)	(0.051)	(0.042)	(0.049)	(0.048)
Architecture	0.457***	0.059	−0.064	−0.012	−0.261*	−0.307**
	(0.122)	(0.139)	(0.137)	(0.112)	(0.143)	(0.130)
Business	0.098**	0.003	−0.018	0.006	−0.079	−0.122**
	(0.046)	(0.062)	(0.060)	(0.049)	(0.055)	(0.057)
Engineering	0.127	0.149	−0.194**	0.254**	0.256**	−0.017
	(0.099)	(0.129)	(0.091)	(0.110)	(0.121)	(0.096)
N	34,637	34,244	34,244	34,637	34,244	34,244
Mean, ranks $[-5,-1]$	0.216	0.338	0.338	3.640	3.750	3.750
Selection weights		Y	Y		Y	Y
Classroom dummies			Y			Y

Notes: This table displays tracking effects on grades in upper-level required courses with regressions weighted to adjust for selection into observing this outcome. We compute selection weights separately for each program group (architecture, business, and engineering) using the following procedure:

(1) Compute first-year GPA, g_i, for each individual i, defined as the mean grade in all first-year required courses.
(2) Compute mean first-year GPA separately for students with admission ranks $[0, 4]$ and $[-5, -1]$. Let $\bar{g}_a = E[g_i | 0 \le x_i \le 4]$ and $\bar{g}_b = E[g_i | -5 \le x_i \le -1]$, where x_i is individual i's admission rank.
(3) Demean first-year GPA for all students. Let $\tilde{g}_i = g_i - \bar{g}_a$ if $x_i \ge 0$, and $\tilde{g}_i = g_i - \bar{g}_b$ if $x_i < 0$.
(4) Estimate the kernel density of $\tilde{g}_i$ for students who took upper-level required courses. We use Stata's `kdensity` with the default options and include an observation for each course taken by each individual. We estimate this density separately for ranks $[0, 4]$ and $[-5, -1]$, which yields the distributions $f_a(\tilde{g}_i)$ and $f_b(\tilde{g}_i)$.
(5) Compute selection weights as $w_i = 1$ if $x_i \ge 0$, and $w_i = f_a(\tilde{g}_i)/f_b(\tilde{g}_i)$ if $x_i < 0$, where $w_i = 0$ if $f_b(\tilde{g}_i) = 0$.

The table displays π coefficients from separate RD regressions (13). In columns (A)–(C), the dependent variable is demeaned first-year GPA, $\tilde{g}_i$, for students who took upper-level required courses. This is a measure of selection into taking upper-level courses. In columns (D)–(F), the dependent variable is the student's upper-level course grade. All regressions include an observation for each individual/course pair.

Columns (A) and (D) present unweighted estimates from equation (13). Columns (B) and (E) estimate this specification with selection weights, w_i. Columns (C) and (F) add classroom fixed effects to columns (B) and (E).

All regressions include only flagship enrollees within 30 ranks of the tracking threshold. The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A17. Tracking effects on exit exam scores

	(A)	(B)	(C)	(D)
	Demeaned first-year GPA of exit exam takers		Exit exam score (SD units)	
	Un-adjusted	Selection adjusted	Un-adjusted	Selection adjusted
All programs	0.212***	0.085	0.032	−0.153
	(0.068)	(0.061)	(0.156)	(0.171)
Architecture	0.501***	0.239	0.141	−0.241
	(0.162)	(0.197)	(0.351)	(0.394)
Business	0.205**	0.090	−0.005	−0.161
	(0.080)	(0.071)	(0.189)	(0.207)
Engineering	0.090	0.056	0.145	0.154
	(0.169)	(0.168)	(0.428)	(0.440)
N	484	462	484	462
Mean, ranks $[-5,-1]$	0.442	0.603	0.758	0.926
Selection weights		Y		Y

Notes: This table displays tracking effects on college exit exam scores with regressions weighted to adjust for selection into observing this outcome. We compute selection weights separately for each program group (architecture, business, and engineering) using the following procedure:

(1) Compute first-year GPA, g_i, for each individual i, defined as the mean grade in all first-year required courses.
(2) Compute mean first-year GPA separately for students with admission ranks $[0,4]$ and $[-5,-1]$. Let $\bar{g}_a = E[g_i|0 \leq x_i \leq 4]$ and $\bar{g}_b = E[g_i|-5 \leq x_i \leq -1]$, where x_i is individual i's admission rank.
(3) Demean first-year GPA for all students. Let $\tilde{g}_i = g_i - \bar{g}_a$ if $x_i \geq 0$, and $\tilde{g}_i = g_i - \bar{g}_b$ if $x_i < 0$.
(4) Estimate the kernel density of $\tilde{g}_i$ for students who took the exit exam. We use Stata's `kdensity` with the default options. We estimate this density separately for ranks $[0,4]$ and $[-5,-1]$, which yields the distributions $f_a(\tilde{g}_i)$ and $f_b(\tilde{g}_i)$.
(5) Compute selection weights as $w_i = 1$ if $x_i \geq 0$, and $w_i = f_a(\tilde{g}_i)/f_b(\tilde{g}_i)$ if $x_i < 0$, where $w_i = 0$ if $f_b(\tilde{g}_i) = 0$.

The table displays π coefficients from separate RD regressions (13). In columns (A)–(B), the dependent variable is demeaned first-year GPA, $\tilde{g}_i$, for students who took the exit exam. This is a measure of selection into taking the exam. In columns (C)–(D), the dependent variable is the student's college exit exam score normalized to $(0,1)$ for each exam field and cohort. In computing selection weights and in these regressions, we count students as taking the exit exam only if their exam field matches their program at the flagship (e.g., applicants to the flagship's accounting program who took the accounting exit exam).

Columns (A) and (C) present unweighted estimates from equation (13). Columns (B) and (D) estimate this specification with selection weights, w_i.

All regressions include only flagship enrollees within 30 ranks of the tracking threshold. The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A18. Tracking effects on earnings for flagship graduates

	(A)	(B)	(C)	(D)
	Demeaned first-year GPA of flagship graduates		Log daily earnings after 10–12 years	
	Un-adjusted	Selection adjusted	Un-adjusted	Selection adjusted
All programs	0.095	0.026	−0.147*	−0.146
	(0.059)	(0.057)	(0.086)	(0.091)
Architecture	0.756***	0.417***	−1.053***	−1.338***
	(0.163)	(0.094)	(0.252)	(0.351)
Business	0.102	0.043	−0.038	−0.062
	(0.064)	(0.059)	(0.102)	(0.107)
Engineering	−0.136	−0.062	−0.131	−0.148
	(0.134)	(0.148)	(0.185)	(0.187)
N	1,210	1,144	1,210	1,144
Mean, ranks $[-5, -1]$	0.375	0.471	10.659	10.687
Selection weights		Y		Y

Notes: This table displays tracking effects on earnings conditional on flagship graduation, with regressions weighted to adjust for selection into observing this outcome. We compute selection weights separately for each program group (architecture, business, and engineering) using the following procedure:

(1) Compute first-year GPA, g_i, for each individual i, defined as the mean grade in all first-year required courses.
(2) Compute mean first-year GPA separately for students with admission ranks $[0, 4]$ and $[-5, -1]$. Let $\bar{g}_a = E[g_i | 0 \leq x_i \leq 4]$ and $\bar{g}_b = E[g_i | -5 \leq x_i \leq -1]$, where x_i is individual i's admission rank.
(3) Demean first-year GPA for all students. Let $\tilde{g}_i = g_i - \bar{g}_a$ if $x_i \geq 0$, and $\tilde{g}_i = g_i - \bar{g}_b$ if $x_i < 0$.
(4) Estimate the kernel density of $\tilde{g}_i$ for students who graduated from the flagship. We use Stata's `kdensity` with the default options. We estimate this density separately for ranks $[0, 4]$ and $[-5, -1]$, which yields the distributions $f_a(\tilde{g}_i)$ and $f_b(\tilde{g}_i)$.
(5) Compute selection weights as $w_i = 1$ if $x_i \geq 0$, and $w_i = f_a(\tilde{g}_i)/f_b(\tilde{g}_i)$ if $x_i < 0$, where $w_i = 0$ if $f_b(\tilde{g}_i) = 0$.

The table displays π coefficients from separate RD regressions (13). In columns (A)–(B), the dependent variable is demeaned first-year GPA, $\tilde{g}_i$, for students who graduated from the flagship. This is a measure of selection into graduation. In columns (C)–(D), the dependent variable is log daily earnings measured 10–12 years after application. All regressions include an observation for each year in which individuals have positive earnings using 2008–2012 labor market data.

Columns (A) and (C) present unweighted estimates from equation (13). Columns (B) and (D) estimate this specification with selection weights, w_i.

All regressions include only flagship enrollees within 30 ranks of the tracking threshold. The first row shows the pooled estimate across all programs. Other rows show estimates by program group.

Parentheses contain standard errors clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.1. Derivation of professor failure thresholds. This section derives the three professor grading standards in Section 1.4:

- The individual failure threshold, F^{Ind}, defined in equation (9)
- The optimal failure threshold, F^*, defined in equation (10)
- The fixed failure threshold, F^{Fixed}, defined in equation (11)

Individual failure threshold. We begin with the schooling decision that an individual student would make herself in a standard human capital model. We use a simplified version of the Becker (1964) model in which students choose to stay in college if the expected wage return exceeds an interest rate, r. This schooling model can be derived following the assumptions described in Card (2001). In short, it assumes students can borrow without constraint at rate r, and it ignores tuition costs, disutility of college relative to work, earnings while in school, and heterogeneity in hours worked.[29]

Given these assumptions, consider an individual's schooling decision at the end of the first year of college. The individual's information on her unknown skill, θ_{i1}, includes her prior, $\bar{\theta}_0$, and her first-year grade, g_{i0}. The individual remains in college if

(B1) $$E[w_{it}(1) - w_{it}(0)|\bar{\theta}_0, g_{i0}] > r,$$

where $w_{it}(0)$ is the individual's log wage if she leaves after year $t = 0$, and $w_{it}(1)$ the log wage if the individual stays in college for year $t = 1$.

Using the wage equation (6), expression (B1) simplifies to:

$$E\left[E[\theta_{i2}|\bar{\theta}_0, \mathbf{g}_i, \mathbf{y}_{i,t-1}] - E[\theta_{i1}|\bar{\theta}_0, \mathbf{g}_i, \mathbf{y}_{i,t-1}]\Big|\bar{\theta}_0, g_{i0}\right] > r$$

$$E[\theta_{i2} - \theta_{i1}|\bar{\theta}_0, g_{i0}] > r$$

(B2) $$E[v_{i1}|\bar{\theta}_0, g_{i0}] > r.$$

The second line follows from the law of iterated expectations. The third line follows from the definition of skill accumulation (equation (1)). Equation (B2) reproduces the decision rule (8).

From equation (B2), the individual failure threshold, F^{Ind}, is the value of g_{i0} such that:

$$E[v_{i1}|\bar{\theta}_0, g_{i0}] = r.$$

---

[29] With this simplification we abstract from the possibility that classmates affect the costs of college (e.g., disutility). We make this choice because we believe that the main outcome in our model—course grades—are likely to be more related to a student's returns to college than to her costs.

Using the definitions of value added (equation (2)) and expected skill (equation (4)), this threshold is given by

$$\mu + \gamma\left(\frac{\rho_0}{\rho_0 + \rho_g}\bar{\theta}_0^{End} + \frac{\rho_g}{\rho_0 + \rho_g}g_{i0}\right) + \beta\bar{\theta}_1 = r$$

$$\frac{\rho_g}{\rho_0 + \rho_g}g_{i0} = \frac{r - \mu - \beta\bar{\theta}_1}{\gamma} - \frac{\rho_0}{\rho_0 + \rho_g}\bar{\theta}_0^{End}$$

$$F^{Ind} = \left(1 + \frac{\rho_0}{\rho_g}\right)\frac{r - \mu - \beta\bar{\theta}_1}{\gamma} - \frac{\rho_0}{\rho_g}\bar{\theta}_0^{End},$$

where the last line replicates equation (9).

Optimal failure threshold. The optimal failure threshold, F^*, is the value of F that maximizes the average return in the classroom. In this case, the professor considers the effect of her failure threshold on mean skill in the upper-level classroom, $\bar{\theta}_1$. Thus we denote $\bar{\theta}_1(F)$ as a function of F from the professor's perspective. Given the normality assumptions, this term is equal to

(B3)
$$\bar{\theta}_1(F) = \bar{\theta}_0^{End} + \frac{\rho_g\sigma_{0g}}{\rho_0 + \rho_g}\lambda\left(\frac{F - \bar{\theta}_0^{End}}{\sigma_{0g}}\right),$$

where σ_{0g} is the standard deviation of first-year grades, g_{i0}, and $\lambda(\cdot)$ is the inverse Mills ratio for the standard normal distribution.[30] Mean skill in the advanced course, $\bar{\theta}_1(F)$, is equal to mean skill at the end of the first year, $\bar{\theta}_0^{End}$, plus a nonnegative selection term that is increasing in the failure threshold, F.

We define the optimal failure threshold, F^*, as

(B4)
$$F^* = \operatorname*{argmax}_F E[w_{it}(G_i) - G_i r|\bar{\theta}_0, F],$$

where G_i is an indicator for passing the first-year class, or, equivalently in this model, graduating from college. $w(G_i)$ is individual i's log wage given the graduation outcome, and r is the cost of an additional year of college.

The mean return to college in equation (B4) is a weighted average of returns for two groups: students who continue onto the advanced course, and students who drop out. If the first-year professor sets a threshold, F, that determines an individual's graduation outcome,

---

[30] Specifically, $\sigma_{0g}^2 = \sigma^2(1 + \gamma)^2 + \sigma_g^2$, and $\lambda(\cdot) = \phi(\cdot)/(1 - \Phi(\cdot))$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and cumulative distribution functions. Equation (B3) can be derived by $E[\theta_{i1}|\bar{\theta}_0, g_{i0} > F] = E\big[E[\theta_{i1}|\bar{\theta}_0, g_{i0}]|\bar{\theta}_0, g_{i0} > F\big] = (\rho_0\bar{\theta}_0^{End} + \rho_g E[g_{i0}|g_{i0} > F])/(\rho_0 + \rho_g)$, where $E[g_{i0}|g_{i0} > F] = \bar{\theta}_0^{End} + \sigma_{0g}\lambda(\tilde{F})$ by the formula for a truncated normal variable g_{i0} with mean $\bar{\theta}_0^{End}$ and standard deviation σ_{0g}.

G_i, the mean return can be expressed as

$$E[w_{it}(G_i) - G_i r | \bar{\theta}_0, F] = Pr[g_{i0} < F | \bar{\theta}_0]\Big(E[w_{it}(0)|\bar{\theta}_0, g_{i0} < F]\Big) +$$
$$Pr[g_{i0} > F | \bar{\theta}_0]\Big(E[w_{it}(1)|\bar{\theta}_0, g_{i0} > F] - r\Big).$$

Students with initial course grades $g_{i0} < F$ fail out of college and have expected earnings defined by the non-graduate wage, $w_{it}(0)$. Students with grades $g_{i0} > F$ earn the college graduate wage, $w_{it}(1)$, but pay the cost of an additional year of school, r. Equivalently, the mean return can be written as:

$$E[w_{it}(G_i) - G_i r | \bar{\theta}_0, F] = E[w_{it}(0)|\bar{\theta}_0] + Pr[g_{i0} > F | \bar{\theta}_0]\Big(E[w_{it}(1) - w_{it}(0)|\bar{\theta}_0, g_{i0} > F] - r\Big).$$

To evaluate this expression, note that $Pr[g_{i0} < F | \bar{\theta}_0]$ is equal to $\Phi(\tilde{F})$, where $\Phi(\cdot)$ is the standard normal CDF, and we define $\tilde{F} = (F - \bar{\theta}_0^{End})/\sigma_{0g}$ to be the normalized grading threshold. Using this fact and the expressions for log wage (equation (6)) and value added (equation (2)), the mean return is

(B5) $\qquad E[w_{it}(G_i) - G_i r | \bar{\theta}_0, F] = \bar{\theta}_0^{End} + \Big(1 - \Phi(\tilde{F})\Big)\Big(\mu + (\gamma + \beta)\bar{\theta}_1(F) - r\Big).$

The professor's optimal grade standard, F^*, is the value of F that maximizes equation (B5), which is given by:[31]

$$\Big(1 - \Phi(\tilde{F})\Big)(\gamma + \beta)\frac{\rho_g}{\rho_0 + \rho_g}\lambda(\tilde{F})\Big(\lambda(\tilde{F}) - \tilde{F}\Big) = \frac{1}{\sigma_{0g}}\phi(\tilde{F})\Big(\mu + (\gamma + \beta)\bar{\theta}_1(F) - r\Big)$$

$$(\gamma + \beta)\frac{\rho_g}{\rho_0 + \rho_g}\sigma_{0g}\Big(\lambda(\tilde{F}) - \tilde{F}\Big) = \mu + (\gamma + \beta)\Big(\bar{\theta}_0^{End} + \frac{\rho_g}{\rho_0 + \rho_g}\sigma_{0g}\lambda(\tilde{F})\Big) - r$$

$$(\gamma + \beta)\frac{\rho_g}{\rho_0 + \rho_g}(F - \bar{\theta}_0^{End}) = r - \mu - (\gamma + \beta)\bar{\theta}_0^{End}$$

$$F^* = \Big(1 + \frac{\rho_0}{\rho_g}\Big)\frac{r - \mu}{\gamma + \beta} - \frac{\rho_0}{\rho_g}\bar{\theta}_0^{End},$$

where the last line replicates equation (10).

We note that $F^* > F^{Ind}$ if and only if learning externalities are positive, $\beta > 0$. This follows from

$$\frac{r - \mu}{\gamma + \beta} > \frac{r - \mu - \beta\bar{\theta}_1}{\gamma}$$

$$0 > \beta(r - \mu) - (\gamma + \beta)\beta\bar{\theta}_1$$

$$\beta\Big(\bar{\theta}_1 - \frac{r - \mu}{\gamma + \beta}\Big) > 0.$$

---

[31] The derivation of (10) uses the result that the derivative of the inverse Mills ratio is $\lambda'(x) = \lambda(x)\big(\lambda(x) - x\big)$. Thus using (B3) it follows that $d\bar{\theta}_1(F)/dF = \rho_g/(\rho_0 + \rho_g)\lambda'(\tilde{F}) = \rho_g/(\rho_0 + \rho_g)\lambda(\tilde{F})\big(\lambda(\tilde{F}) - \tilde{F}\big)$.

One can show that the term in parentheses, $\bar{\theta}_1 - (r-\mu)/(\gamma+\beta)$, must be positive for students to have chosen to enroll in the college in the first place.[32] Thus it follows that $F^* > F^{Ind}$ if and only if $\beta > 0$. If $\beta < 0$, the student's schooling decision binds, i.e., $F^* < F^{Ind}$.

Fixed failure threshold. The fixed failure threshold, F^{Fixed}, is the threshold such that a constant proportion of students, q, fail regardless of the skill composition of the class. Letting $\Phi(\cdot)$ denote the standard normal cumulative distribution function, this threshold can be derived by:

$$Pr[g_{i0} > F|\bar{\theta}_0] = q$$

$$\Phi\left(\frac{F - \bar{\theta}_0^{End}}{\sigma_{0g}}\right) = q$$

$$F^{Fixed} = \bar{\theta}_0^{End} + \sigma_{0g}\Phi^{-1}(q)$$

where the last line replicates equation (11).

B.2. **Details on Propositions 1–2.** This section derives expressions for peer effects in first-year grades, and uses these to prove Propositions 1 and 2 from Section 1.5.

We denote peer effects in first-year grades by π_0. From equation (12), π_0 is given by:

(B6)
$$\pi_0 = \frac{dE[g_{i0}|\theta_{i0}, \bar{\theta}_0]}{d\bar{\theta}_0} - \frac{dF}{d\bar{\theta}_0}.$$

The first term on the righthand side of equation (B6) is the derivative of the individual's expected raw score, g_{i0}, with respect to mean classroom skill, $\bar{\theta}_0$. From equations (1)–(3), the raw score is

$$g_{i0} = \mu + (1+\gamma)\theta_{i0} + \beta\bar{\theta}_0 + \epsilon_{it}^g.$$

Thus the derivative of the expected raw grade with respect to $\bar{\theta}_0$ is:

(B7)
$$\frac{dE[g_{i0}|\theta_{i0}, \bar{\theta}_0]}{d\bar{\theta}_0} = \beta.$$

The second term on the righthand side of equation (B6) is the derivative of the professor's failure threshold, F, with respect to mean classroom skill, $\bar{\theta}_0$. The term depends on whether the professor uses individual grading standards, F^{Ind}, optimal grading standards, F^*, or fixed grading standards, F^{Fixed}. We derive each of these three derivatives separately.

Individual grading standards, F^{Ind}, are defined by equation (9). In this case, the derivative with respect to $\bar{\theta}_0$ is complicated to compute because mean peer ability in advanced courses,

---

[32] In this model all students have expected pre-college skill $\bar{\theta}_0$. A student will choose to enroll in college only if the expected return to doing so is greater than the return to no college, i.e., if $E[w_{it}(G_i) - G_i r|\bar{\theta}_0, F] > \bar{\theta}_0 + r$. Using (B5), this evaluates to $\mu + (\gamma + \beta)\bar{\theta}_0 - r + (1 - \Phi(\tilde{F}))(\mu + (\gamma + \beta)\bar{\theta}_1 - r) > 0$. Since $\bar{\theta}_1 > \bar{\theta}_0$, for this inequality to hold we must have $\mu + (\gamma + \beta)\bar{\theta}_1 - r > 0$, which is equivalent to $\bar{\theta}_1 - (r - \mu)/(\gamma + \beta) > 0$.

$\bar{\theta}_1$, depends in part on F^{Ind} (equation (B3)). We derive $dF/d\bar{\theta}_0$ using the implicit function theorem on equation (9) with $F = F^{Ind}$.[33]

$$dF = -\frac{\beta}{\gamma}\left(1 + \frac{\rho_0}{\rho_g}\right)\left(\frac{\partial\bar{\theta}_1(F)}{\partial F}dF + \frac{\partial\bar{\theta}_1(F)}{\partial\bar{\theta}_0}d\bar{\theta}_0\right) - \frac{\rho_0}{\rho_g}(1 + \gamma + \beta)d\bar{\theta}_0,$$

$$dF + \frac{\beta}{\gamma}\lambda'(\tilde{F})dF = -\frac{\beta}{\gamma}\left(1 + \frac{\rho_0}{\rho_g}\right)(1 + \gamma + \beta)\left(1 - \frac{\rho_g}{\rho_0 + \rho_g}\lambda'(\tilde{F})\right)d\bar{\theta}_0 - \frac{\rho_0}{\rho_g}(1 + \gamma + \beta)d\bar{\theta}_0,$$

$$\left\{1 + \frac{\beta}{\gamma}\lambda'(\tilde{F})\right\}dF = (1 + \gamma + \beta)\left\{-\frac{\beta}{\gamma}\left(1 + \frac{\rho_0}{\rho_g}\right) + \frac{\beta}{\gamma}\lambda'(\tilde{F}) - \frac{\rho_0}{\rho_g}\right\}d\bar{\theta}_0,$$

$$\left\{\gamma + \beta\lambda'(\tilde{F})\right\}dF = (1 + \gamma + \beta)\left\{-\left(1 + \frac{\rho_0}{\rho_g}\right)(\gamma + \beta) + \gamma + \beta\lambda'(\tilde{F})\right\}d\bar{\theta}_0,$$

(B8) $$\qquad \frac{dF^{Ind}}{d\bar{\theta}_0} = -(1 + \gamma + \beta)\left\{\left(1 + \frac{\rho_0}{\rho_g}\right)\frac{\gamma + \beta}{\gamma + \beta\lambda'(\tilde{F}^{Ind})} - 1\right\}.$$

This derivation defines $\lambda'(\tilde{F}^{Ind})$ to be the derivative of the inverse Mills ratio evaluated at the normalized grading threshold, $\tilde{F}^{Ind} = (F^{Ind} - \bar{\theta}_0^{End})/\sigma_{0g}$.

Optimal grading standards, F^*, are defined by equation (10). The derivative with respect to $\bar{\theta}_0$ is

(B9) $$\frac{dF^*}{d\bar{\theta}_0} = -\frac{\rho_0}{\rho_g}(1 + \gamma + \beta)$$

Finally, fixed grading standards, F^{Fixed}, are defined by equation (11). The derivative with respect to $\bar{\theta}_0$ is

(B10) $$\frac{dF^{Fixed}}{d\bar{\theta}_0} = (1 + \gamma + \beta)$$

Combining equations (B6)–(B10) yields the expression for peer effects in first-year grades, π_0, given the professor's grading behavior:

(B11) $$\pi_0 = \begin{cases} \beta + \left(1 + \gamma + \beta\right)\left\{\left(1 + \frac{\rho_0}{\rho_g}\right)\frac{\gamma + \beta}{\gamma + \beta\lambda'(\tilde{F}^{Ind})} - 1\right\} & \text{if } F = F^{Ind} \\[2mm] \beta + \left(1 + \gamma + \beta\right)\frac{\rho_0}{\rho_g} & \text{if } F = F^* \\[2mm] -\left(1 + \gamma\right) & \text{if } F = F^{Fixed} \end{cases}$$

We use this expression for the derivations of Propositions 1 and 2 below.

Proposition 1. *If peer effects in first-year grades are positive, $\pi_0 > 0$, then either:*

- *Higher-ability classmates increase individual learning, $\beta > 0$; or,*

---

[33] Recall that mean classroom skill at the end of the first year is $\bar{\theta}_0^{End} = \mu + (1 + \gamma + \beta)\bar{\theta}_0$.

- *Classmates have information externalities, $\rho_g < \infty$.*

If peer effects in first-year grades are negative, $\pi_0 < 0$, then either:

- *Higher-ability classmates reduce individual learning, $\beta < 0$; or,*
- *Professors use fixed grading standards, F^{Fixed}.*

We demonstrate the two statements in Proposition 1 by contradiction.

For the first statement, suppose conversely that $\beta \leq 0$ and $\rho_g = \infty$. Consider equation (B11) for each possible failure threshold, F:

- If $F = F^{Ind}$, we have $(\gamma + \beta)/(\gamma + \beta\lambda'(\tilde{F}^{Ind})) \leq 1$ because $\beta > -\gamma$ by assumption (Section 1.1), and because the derivative of the inverse Mills ratio, $\lambda'(\tilde{F}^{Ind})$, is bounded between zero and one. Thus the term in brackets, $\{\cdot\}$, is non-positive when $\rho_g = \infty$. Since $\beta \leq 0$ and $(1 + \gamma + \beta) > 0$, it follows that $\pi_0 \leq 0$.
- If $F = F^*$, then $\pi_0 = \beta \leq 0$.
- If $F = F^{Fixed}$, then $\pi_0 = -(1 + \gamma) < 0$.

Thus if $\pi_0 > 0$, it must be that either $\beta > 0$ or $\rho_g < \infty$.

For the second statement, suppose conversely that $\beta \geq 0$ and $F \neq F^{Fixed}$.

- If $F = F^{Ind}$, we have $(\gamma+\beta)/(\gamma+\beta\lambda'(\tilde{F}^{Ind})) \geq 1$ by a similar argument to that above. Thus the term in brackets, $\{\cdot\}$, is non-negative. Since $\beta \geq 0$ and $(1 + \gamma + \beta) > 0$, it follows that $\pi_0 \geq 0$.
- If $F = F^*$, then $\pi_0 = \beta + (1 + \gamma + \beta)\rho_0/\rho_g \geq 0$.

Thus if $\pi_0 < 0$, it must be that either $\beta < 0$ or $F = F^{Fixed}$.

Proposition 2. *If professors use fixed grading standards, F^{Fixed}, then peer effects in first-year grades, π_0, are decreasing in the skill return to individual ability, γ.*

If professors use optimal grading standards, F^*, then π_0 increases with γ.

From equation (B11), we have:

$$\frac{d\pi_0}{d\gamma} = \begin{cases} \dfrac{\rho_0}{\rho_g} & \text{if } F = F^* \\[2ex] -1 & \text{if } F = F^{Fixed}, \end{cases}$$

where $\rho_0, \rho_g > 0$. We note that $d\pi_0/d\gamma$ cannot be signed in general if $F = F^{Ind}$.

B.3. **Details on Proposition 3.** This section provides details on Proposition 3 from Section 1.6, which characterizes peer effects in graduation and wages.

Proposition 3.

(a) *Peer effects in graduation have the same sign as peer effects in first-year grades, π_0.*

(b) Peer effects in log wages can be negative if learning externalities are negative, $\beta < 0$, or if higher-ability classmates reduce the probability of graduation.

(c) Conditional on graduation, peer effects in wages (or skill) are negative only if $\beta < 0$.

We show parts (a), (b), and (c) in turn.

Proposition 3(a). Peer effects in graduation. In the model, graduation occurs if the individual remains in college after the first year, which is determined by the condition $g_{i0} > F$. Letting G_i denote graduation, the probability of graduating conditional on individual pre-college ability, θ_{i0}, is

$$E[G_i|\theta_{i0}, \bar{\theta}_0] = Pr[g_{i0} > F|\theta_{i0}, \bar{\theta}_0] = Pr\left[\frac{\epsilon_{i0}^g}{\sigma_g} > \frac{F - \theta_{i1}}{\sigma_g}\bigg|\theta_{i0}, \bar{\theta}_0\right] = 1 - \Phi(\tilde{F}_i),$$

where we define $\tilde{F}_i = (F - \theta_{i1})/\sigma_g$ to be the failure threshold F normalized for an individual with skill $\theta_{i1} = \mu + (1+\gamma)\theta_{i0} + \beta\bar{\theta}_0$ at the end of the first year. The peer effect in graduation is therefore

(B12) $$\frac{dE[G_i|\theta_{i0}, \bar{\theta}_0]}{d\bar{\theta}_0} = -\phi(\tilde{F}_i)\frac{d\tilde{F}_i}{d\bar{\theta}_0} = \frac{\phi(\tilde{F}_i)}{\sigma_g}\left(\beta - \frac{dF}{d\bar{\theta}_0}\right) = \frac{\phi(\tilde{F}_i)\pi_0}{\sigma_g},$$

where $\phi(\cdot)$ is the standard normal density and π_0 is the grade peer effect defined by equation (B6). Since $\phi(\cdot)$ and σ_g are both positive, graduation peer effects have the same sign as grade peer effects, π_0. This demonstrates Proposition 3(a).

We note that graduation peer effects are large in magnitude only for students with $\tilde{F}_i \approx 0$. These are the students with end-of-year skill near the failure threshold, i.e., $\theta_{i1} \approx F$.

Proposition 3(b). Peer effects in wages. Proposition 3(b) states that peer effects in log wages can be negative only if one of two conditions are met: 1) learning externalities are negative, $\beta < 0$, or 2) higher-ability classmates reduce the likelihood of graduation. From Proposition 3(a), a negative graduation peer effect is equivalent to $\pi_0 < 0$. Thus to demonstrate Proposition 3(b), we derive the expression for peer effects in log wages, and then show it is non-negative whenever both $\beta \geq 0$ and $\pi_0 \geq 0$.

We denote the peer effect in log wages at year t by $dE[w_{it}|\theta_{i0}, \bar{\theta}_0]/d\bar{\theta}_0$. Using potential outcome notation for wages given the graduation decision, $w_{it}(G_i)$, we can write the expected wage of an individual with pre-college ability θ_{i0} and mean classroom skill $\bar{\theta}_0$ as

(B13) $$E[w_{it}|\theta_{i0}, \bar{\theta}_0] = E[w_{it}(0)|\theta_{i0}, \bar{\theta}_0] + Pr[g_{i0} > F|\theta_{i0}, \bar{\theta}_0]E[w_{it}(1) - w_{it}(0)|\theta_{i0}, \bar{\theta}_0, g_{i0} > F],$$

where $Pr[g_{i0} > F|\theta_{i0}, \bar{\theta}_0] = 1 - \Phi(\tilde{F}_i)$ is the probability of graduating, and $\tilde{F}_i = (F - \theta_{i1})/\sigma_g$ as above.

To evaluate equation (B13), we use the definitions of individual skill at the time of labor market entry both with and without college graduation. From equations (1) and (2), skill is

θ_{i1} for individuals who dropped out of college, which is given by

$$\theta_{i1} = \theta_{i0} + v_{i0}$$
$$= \theta_{i0} + \mu + \gamma\theta_{i0} + \beta\bar{\theta}_0.$$

Skill for college graduates, θ_{i2}, is given by:

$$\theta_{i2} = \theta_{i1} + v_{i1}$$
$$= \theta_{i1} + \mu + \gamma\theta_{i1} + \beta\bar{\theta}_1$$

(B14)
$$= \mu(2 + \gamma) + (1 + \gamma)^2\theta_{i0} + \beta(1 + \gamma)\bar{\theta}_0 + \beta\bar{\theta}_1.$$

Using these definitions, we evaluate each of the potential wage terms in equation (B13):

- We derive the term $E[w_{it}(0)|\theta_{i0}, \bar{\theta}_0, g_{i0} > F]$ using the definition of log wage (equation (6)) and applying Bayes' Rule (similar to equation (4)):

$$E[w_{it}(0)|\theta_{i0}, \bar{\theta}_0, g_{i0} > F] = E\left\{ E[\theta_{i1}|\bar{\theta}_0, g_{i0}, \mathbf{y}_{i,t-1}] \Big| \theta_{i0}, \bar{\theta}_0, g_{i0} > F \right\}$$
$$= E\left\{ \frac{\rho_0\bar{\theta}_0^{End} + \rho_g g_{i0} + (t-1)\rho_y\bar{y}_{i,t-1}}{\rho_0 + \rho_g + (t-1)\rho_y} \Big| \theta_{i0}, \bar{\theta}_0, g_{i0} > F \right\}$$
$$= \theta_{i1} + \frac{\rho_0}{\rho_t}(1 + \gamma)(\bar{\theta}_0 - \theta_{i0}) + \frac{\rho_g}{\rho_t}E[\epsilon_{i0}^g|\epsilon_{i0}^g > F - \theta_{i1}]$$
$$= \theta_{i1} + \frac{\rho_0}{\rho_t}(1 + \gamma)(\bar{\theta}_0 - \theta_{i0}) + \frac{\rho_g}{\rho_t}\sigma_g\lambda(\tilde{F}_i).$$

In this expression we define $\rho_t = \rho_0 + \rho_g + (t-1)\rho_y$ to be the employer's total precision for θ_{i1} in year t.

- Using a similar derivation, the term $E[w_{it}(0)|\theta_{i0}, \bar{\theta}_0]$ is given by:

$$E[w_{it}(0)|\theta_{i0}, \bar{\theta}_0] = \theta_{i1} + \frac{\rho_0}{\rho_t}(1 + \gamma)(\bar{\theta}_0 - \theta_{i0}).$$

- Lastly we compute $E[w_{it}(1)|\theta_{i0}, \bar{\theta}_0, g_{i0} > F]$. This expression is more complicated because individuals now have two years of skill accumulation. Specifically, individual skill after two years of college is θ_{i2}, as defined in equation (B14). We also define $\bar{\theta}_1^E$ to be mean (potential) skill at the end of college, which is equal to

$$\bar{\theta}_1^E = (2 + \gamma)\mu + (1 + \gamma)^2\bar{\theta}_0 + (1 + \gamma)\beta\bar{\theta}_0 + \beta\bar{\theta}_1.$$

Thus it follows that $\theta_{i2} \sim N(\bar{\theta}_1^E, 1/\rho_1)$ where $\rho_1 = 1/(\sigma^2(1 + \gamma)^4)$ is the precision of $\bar{\theta}_1^E$ for θ_{i2}. Note that $\rho_1 = \rho_0/(1 + \gamma)^2$ since $\rho_0 = 1/(\sigma^2(1 + \gamma)^2)$.

The expression for $w_{it}(1)$ is also complicated because employers observe two grades for college graduates: an initial course grade, g_{i0}, and an upper-level course grade, g_{i1}. The upper level grade has conditional mean θ_{i2} since $g_{i1} = \theta_{i2} + \epsilon_{i1}^g$ by equation (3). The first-year course grade, however, has conditional mean θ_{i1} since $g_{i0} = \theta_{i1} + \epsilon_{i0}^g$.

Thus we define $\hat{g}_{i0}$ to be:

$$
\begin{aligned}
\hat{g}_{i0} &= \mu + (1+\gamma)g_{i0} + \beta\bar{\theta}_1 \\
&= \mu + (1+\gamma)(\theta_{i1} + \epsilon_{i0}^g) + \beta\bar{\theta}_1 \\
&= \theta_{i2} + (1+\gamma)\epsilon_{i0}^g.
\end{aligned}
$$

The distribution of $\hat{g}_{i0}$ conditional on θ_{i0} and $\bar{\theta}_0$ is $N(\theta_{i2}, 1/\hat{\rho}_g)$, where $\hat{\rho}_g = \rho_g/(1+\gamma)^2$. Similarly, we define $\hat{\mathbf{y}}_{i,t-1} = \{y_{i2}, \ldots, y_{i,t-1}\}$ to be the vector of productivity signals for college graduates, whose first period of employment is $t = 2$. Let $\hat{\bar{y}}_{i,t-1}$ be the mean of $\hat{\mathbf{y}}_{i,t-1}$.

To simplify the derivation of equation (B13), we make the assumption that advanced course grades and on-the-job productivity signals are less precise for college graduates. Specifically, let ρ_{g1} and ρ_{y1} denote the precisions of advanced course grades, g_{i1}, and productivity signals, y_{it}, for college graduates. We assume $\rho_{g1} = \rho_{y1} = \rho_y/(1+\gamma)^2$, where ρ_y is the precision of productivity signals for college drop-outs. This simplifies the derivation of equation (B13) significantly because the graduation/drop-out choice does not affect the relative precisions of different productivity signals.[34]

With this assumption, we now can write $E[w_{it}(1)|\theta_{i0}, \bar{\theta}_0, g_{i0} > F]$ as:

$$
\begin{aligned}
E[w_{it}(1)|\theta_{i0}, \bar{\theta}_0, g_{i0} > F] &= E\left\{ E[\theta_{i2}|\bar{\theta}_0, g_{i0}, g_{i1}, \hat{\mathbf{y}}_{i,t-1}] \Big| \theta_{i0}, \bar{\theta}_0, g_{i0} > F \right\} \\
&= E\left\{ \frac{\rho_1\bar{\theta}_1^E + \hat{\rho}_g\hat{g}_{i0} + \rho_{g1}g_{i1} + (t-2)\rho_{y1}\hat{\bar{y}}_{i,t-1}}{\rho_1 + \hat{\rho}_g + \rho_{g1} + (t-2)\rho_{y1}} \Big| \theta_{i0}, \bar{\theta}_0, g_{i0} > F \right\} \\
&= \theta_{i2} + \frac{\rho_0}{\rho_t}(1+\gamma)^2(\bar{\theta}_0 - \theta_{i0}) + (1+\gamma)\frac{\rho_g}{\rho_t}E[\epsilon_{i0}^g|\epsilon_{i0}^g > F - \theta_{i1}] \\
&= \theta_{i2} + (1+\gamma)\left\{ \frac{\rho_0}{\rho_t}(1+\gamma)(\bar{\theta}_0 - \theta_{i0}) + \frac{\rho_g}{\rho_t}\sigma_g\lambda(\tilde{F}_i) \right\}.
\end{aligned}
$$

(B15)

Combining these three potential wage expressions, equation (B13) simplifies to:

$$
E[w_{it}|\theta_{i0}, \bar{\theta}_0] = \theta_{i1} - \frac{\rho_0}{\rho_t}(1+\gamma)(\theta_{i0} - \bar{\theta}_0) + \left(1 - \Phi(\tilde{F}_i)\right)\left\{ v_{i1} - \frac{\rho_0}{\rho_t}\gamma(1+\gamma)(\theta_{i0} - \bar{\theta}_0) \right\} + \gamma\frac{\rho_g}{\rho_t}\sigma_g\phi(\tilde{F}_i),
$$

where $v_{i1} = \theta_{i2} - \theta_{i1}$ by equation (1).

---

[34] Since skill accumulation is recursive (equation (1)), the variance of skill increases when students remain in school for a second year. If the precision of advanced course grades and productivity signals were the same for college graduates and drop-outs, employers would place more weight on these signals for college graduates. We abstract from this possibility here, as this is likely to be a second-order consideration in educational choices and wage dynamics.

The peer effect in log wages at year t is thus:[35]

$$\frac{dE[w_{it}|\theta_{i0}, \bar{\theta}_0]}{d\bar{\theta}_0} = \beta + \frac{\rho_0}{\rho_t}(1+\gamma) + \left(1 - \Phi(\tilde{F}_i)\right)\left\{\beta\left(\frac{d\bar{\theta}_1}{d\bar{\theta}_0} + \gamma\right) + \frac{\rho_0}{\rho_t}\gamma(1+\gamma)\right\}$$

(B16)
$$+ \frac{\phi(\tilde{F}_i)\pi_0}{\sigma_g}\left\{v_{i1} - \frac{\rho_0}{\rho_t}\gamma(1+\gamma)(\theta_{i0} - \bar{\theta}_0) - \gamma\frac{\rho_g}{\rho_t}(\theta_{i1} - F)\right\},$$

where $\phi(\cdot)$ is the standard normal probability density function.

Equation (B16) is hard to sign in general because both skill and information mechanisms depend on an individual's own skill, θ_{i0}. We evaluate the sign of wage peer effects in the case that matches our empirical setting. Specifically, we consider an individual with pre-college ability above the classroom mean, $\theta_{i0} > \bar{\theta}_0$. This matches our context because our RD coefficients estimate the effects of a large increase in $\bar{\theta}_0$ for individuals who otherwise would have been near the top of the classroom ability distribution.

To demonstrate Proposition 3(b), we show that if $\pi_0 \geq 0$ and $\beta \geq 0$, then equation (B16) is non-negative. We consider the terms of equation (B16) in turn.

- If $\beta \geq 0$, the first three terms in (B16) are non-negative because γ, ρ_0, and ρ_t are all positive, $1 - \Phi(\tilde{F}_i) \in (0, 1)$, and mean skill in the advanced class is increasing in mean skill in the first-year class $(d\bar{\theta}_1/d\bar{\theta}_0 > 0)$.
- To sign the fourth term, first note that $\phi(\tilde{F}_i)\pi_0/\sigma_g$ is non-negative when $\pi_0 \geq 0$. For the expression in brackets,

(B17)
$$v_{i1} - \frac{\rho_0}{\rho_t}\gamma(1+\gamma)(\theta_{i0} - \bar{\theta}_0) - \gamma\frac{\rho_g}{\rho_t}(\theta_{i1} - F),$$

begin by considering an individual i' whose ability, $\theta_{i'0}$, is such that her skill at the end of the first year is exactly equal to the failure threshold, i.e., $\theta_{i'1} = F$. When $F = F^{Ind}$ or $F = F^*$, the skill return to the advanced course is positive for this individual, $v_{i'1} > 0$, when there is a positive cost to an addition year of college, $r > 0$.[36] If the majority of students pass the first-year class, as is true in our empirical setting, then this individual has ability $\theta_{i'0} < \bar{\theta}_0$. Thus expression (B17) is positive for the individual with $\theta_{i'1} = F$.

Finally, note that expression (B17) is increasing in θ_{i0}, since the derivative of this term with respect to θ_{i0} is given by:

$$\gamma(1+\gamma)\left(1 - \frac{\rho_0 + \rho_g}{\rho_t}\right).$$

Thus expression (B17) is positive for any individual with $\theta_{i0} > \bar{\theta}_0$.

---

[35] The derivation of equation (B16) uses the fact that $\phi'(\tilde{F}_i) = -\phi(\tilde{F}_i)\tilde{F}_i$ and the definition $\tilde{F}_i = (F - \theta_{i1})/\sigma_g$.
[36] Recall from equation (B11) that π_0 can only be positive if $F = F^{Ind}$ or $F = F^*$.

In sum, we have shown that $dE[w_{it}|\theta_{i0}, \bar{\theta}_0]/d\bar{\theta}_0 \geq 0$ when $\beta \geq 0$ and $\pi_0 \geq 0$ for an individual with skill above the classroom mean. This demonstrates Proposition 3(b).

We note that given the right data, one could use changes in wage peer effects over time to examine the role of information mechanisms. This is similar to the empirical analysis in Altonji and Pierret (2001). In our model, the change in wage peer effects over time is given by the derivative of equation (B16) with respect to t:[37]

$$\frac{dE[w_{it}|\theta_{i0}, \bar{\theta}_0]}{d\bar{\theta}_0 dt} = -\frac{\rho_0 \rho_y}{\rho_t^2}(1+\gamma)\left\{1 + \gamma\left(1 - \Phi(\tilde{F}_i) - \frac{\phi(\tilde{F}_i)\pi_0}{\sigma_g}(\theta_{i0} - \bar{\theta}_0)\right)\right\} + \frac{\phi(\tilde{F}_i)\pi_0}{\sigma_g}\frac{\rho_g \rho_y}{\rho_t^2}\gamma(\theta_{i1} - F)$$

One can show that this expression is negative in the case of $\theta_{i1} > F$, which implies that wage peer effects decrease over a worker's career if there are information externalities.

Proposition 3(c). Peer effects in wages (or skill) for graduates. Proposition 3(c) states that peer effects in wages/skill for college graduates can only be negative if learning externalities are negative, $\beta < 0$.

We derived the expression for mean wage conditional on graduation in equation (B15) above:

$$E[w_{it}(1)|\theta_{i0}, \bar{\theta}_0, g_{i0} > F] = \theta_{i2} + (1+\gamma)\left\{\frac{\rho_0}{\rho_t}(1+\gamma)(\bar{\theta}_0 - \theta_{i0}) + \frac{\rho_g}{\rho_t}\sigma_g \lambda(\tilde{F}_i)\right\},$$

where $\tilde{F}_i = (F - \theta_{i1})/\sigma_g$ and $\theta_{i2} = \mu(2+\gamma) + (1+\gamma)^2\theta_{i0} + \beta(1+\gamma)\bar{\theta}_0 + \beta\bar{\theta}_1$.

Peer effect in wages conditional on graduation are thus given by:

(B18) $$\frac{dE[w_{it}(1)|\theta_{i0}, \bar{\theta}_0, g_{i0} > F]}{d\bar{\theta}_0} = \beta\left(1 + \gamma + \frac{d\bar{\theta}_1}{d\bar{\theta}_0}\right) + \frac{\rho_0}{\rho_t}(1+\gamma)^2 - \frac{\rho_g}{\rho_t}(1+\gamma)\lambda'(\tilde{F}_i)\pi_0.$$

To sign the expression in (B18), note that γ, $d\bar{\theta}_1/d\bar{\theta}_0$, and $\lambda'(\tilde{F}_i)$ are all positive. If $\pi_0 < 0$, then (B18) can only be negative if $\beta < 0$.

In this model—as is standard in the employer learning literature—wages converge to an individual's true skill over time. That is, $\rho_t \to \infty$ as $t \to \infty$, and thus $w_{it}(1) \to \theta_{i2}$. We can therefore derive the expression for peer effects in skill for college graduates by setting $\rho_t = \infty$ in equation (B18):

$$\frac{dE[\theta_{i2}|\theta_{i0}, \bar{\theta}_0]}{d\bar{\theta}_0} = \beta\left(1 + \gamma + \frac{d\bar{\theta}_1}{d\bar{\theta}_0}\right).$$

As with equation (B18), skill peer effects for graduates can only be negative if learning externalities are negative, $\beta < 0$. This demonstrates Proposition 3(c).

---

[37] For this derivation, note that $d\rho_t/dt = \rho_y$ since $\rho_t = \rho_0 + \rho_g + (t-1)\rho_y$.

B.4. **Classroom ability distribution.** In Section 1.2, we assume that ability in the first-year course is normally distributed, i.e.,

$$\theta_{i0} \sim N(\bar{\theta}_0, \sigma^2).$$

One can microfound this ability distribution following MacLeod and Urquiola (2015). Suppose ability in the full population of potential college students is distributed

$$\theta_{i0} \sim N(0, 1/\rho_a).$$

Suppose also that students take a college admission exam and receive a test score

$$\tau_i = \theta_{i0} + \epsilon_i^\tau,$$

where $\epsilon_i^\tau \overset{i.i.d.}{\sim} N(0, 1/\rho_\tau)$. If a college program admits only students with scores $\tau_i = \tau_0$, and admitted students take courses together, then ability in a first-year classroom is distributed $N(\bar{\theta}_0, \sigma^2)$, with

$$\bar{\theta}_0 = \frac{\tau_0 \rho_\tau}{\rho_a + \rho_\tau}$$

$$\sigma^2 = \frac{1}{\rho_a + \rho_\tau}.$$

C.1. **Data and sample.** This section provides details on our data merging and analysis sample.

We use two datasets provided by the flagship university:

(1) Lists of applicants to Univalle's undergraduate programs from 2000–2003.
(2) Transcript records for all students in our sample of programs who enrolled in Univalle.

Our sample includes flagship applicants to the programs and years with tracking admissions, as listed in Table 2. Column (A) of Table C1 shows the total number of applicants to the programs and cohorts in our sample. Our analysis includes only students admitted to either a fall or spring cohort, as depicted in column (B).

We combine the flagship data with three individual-level administrative datasets:

(1) Records from the ICFES national standardized college admission exam that include all students who took the exam in 1998–2003. This agency also provided records on a college *exit* exam that we use for some analyses.
(2) Records from the Ministry of Education on students who enrolled in nearly all colleges in the country between 1998–2012.
(3) Earnings records from the Ministry of Social Protection for the years 2008–2012.

We merge the three administrative datasets using national ID numbers, birth dates, and full names. The appendix in Riehl (2019) provides details on the merge process and coverage of these administrative datasets.

We merge the flagship application and transcript data into the administrative data using applicants' full names. Since roughly 85 percent of applicants enrolled in the flagship program they were admitted to, most individuals match uniquely on name, program, and cohort. Most applicants who enrolled in other programs also match uniquely on full name. In cases with duplicate names, we use information from the administrative records on individuals' exam cohorts and high school location to identify the correct match; most Colombian students stay in region for college and apply shortly after taking the ICFES entrance exam. Through this process we are able to match over 99 percent of individuals in the flagship datasets to our administrative records, as shown in column (C) of Table C1.

Column (D) in Table C1 shows that there were 94 applicants who were admitted through special quotas for disadvantaged groups (e.g., indigenous students or students with disabilities). Our full analysis sample is depicted in column (E), which is the difference between columns (B) and (D).

Most of our regressions include the subset of applicants whose admission scores are within h ranks of the tracking threshold. Our benchmark model uses $h = 30$, which is roughly the mean of the Calonico et al. (2014) bandwidths across all dependent variables. Column

TABLE C1. Analysis sample

		(A)	(B)	(C)	(D)	(E)	(F)	(G)
			All admission ranks				Within 30 ranks of tracking threshold	
Group	Program	Total applied	Total admitted	Merge rate	Quota admits	Full sample	Benchmark sample	Flagship enrollees
Arch	Architecture	1,488	465	0.989	19	446	255	219
Bus	Accounting (day)	928	575	0.997	16	559	260	241
	Accounting (night)	921	393	0.987	10	383	287	268
	Business Admin (day)	1,171	584	0.995	12	572	267	240
	Business Admin (night)	940	403	0.995	16	387	304	274
	Foreign Trade	126	94	0.989	3	91	61	58
Eng	Chemical Engineering	233	135	1.000	5	130	95	77
	Electrical Engineering	129	126	0.992	1	125	111	97
	Electronic Engineering	403	132	1.000	4	128	78	60
	Materials Engineering	120	120	0.992	0	120	70	61
	Mechanical Engineering	209	126	0.992	8	118	77	70
	Total	6,668	3,153	0.993	94	3,059	1,865	1,665

Notes: Column (A) shows the number of flagship applicants to the programs and years with tracking admissions (see Table 2). Column (B) shows the subset of these applicants who were admitted to either a fall or spring cohort. Column (C) shows the proportion of students who were matched to any of our administrative datasets using the method described in the text. Column (D) shows the number of students who were admitted through special quotas for disadvantaged groups. Column (E) shows our full analysis sample, which is the difference between columns (B) and (D). Column (F) shows the subset of applicants from column (E) who are within 30 admission ranks of the tracking threshold. Column (G) shows the subset of applicants from column (F) who enrolled in the flagship.

(F) shows the number of students in our analysis sample whose admission ranks are within 30 positions of the tracking threshold. In addition, when we analyze outcomes from the transcript data, we restrict our sample to the subset of these applicants who enrolled in the flagship, as shown in column (G).

C.2. **Measure of return to individual ability.** This section provides details on the measure of the return to individual ability that we use for Figure 5. Specifically, we show that our measure is proportional to the γ parameter in our model.

Let $\hat{\gamma}$ denote the measure of the return to individual ability that we use for Figure 5. We define $\hat{\gamma}$ as the coefficient from a within-classroom regression of first-year course grades on individuals' admission scores. Let τ_i denote individual i's admission score as computed by the flagship university for use in tracking admissions. As above, let $\tilde{g}_{i0}$ denote individual's normalized first-year course grade. From equations (3) and (7), $\tilde{g}_{i0}$ is given by

$$\tilde{g}_{i0} = \theta_{i1} + \epsilon_{i0}^g - F$$
$$= \mu + (1+\gamma)\theta_{i0} + \beta\bar{\theta}_0 + \epsilon_{i0}^g - F$$

where $\epsilon_{i0}^g \overset{i.i.d.}{\sim} N(0, \sigma_g^2)$. A within-classroom regression of $\tilde{g}_{i0}$ on τ_i holds fixed mean classroom ability, $\bar{\theta}_0$. Thus we can define $\hat{\gamma}$ as

$$\hat{\gamma} = \frac{\text{cov}(\tilde{g}_{i0}, \tau_i | \bar{\theta}_0)}{\text{var}(\tau_i | \bar{\theta}_0)}$$

$$= \frac{\text{cov}\left(\mu + (1 + \gamma)\theta_{i0} + \beta\bar{\theta}_0 + \epsilon_{i0}^g - F, \tau_i \middle| \bar{\theta}_0\right)}{\text{var}(\tau_i | \bar{\theta}_0)}$$

$$= (1 + \gamma)\left(\frac{\text{cov}(\theta_{i0}, \tau_i | \bar{\theta}_0)}{\text{var}(\tau_i | \bar{\theta}_0)}\right).$$

The term $\text{cov}(\theta_{i0}, \tau_i | \bar{\theta}_0)/\text{var}(\tau_i | \bar{\theta}_0)$ is the coefficient from a within-classroom linear projection of true skill θ_{i0}, on admission scores, τ_i. As long as true skill and admission scores are related, $\text{cov}(\theta_{i0}, \tau_i | \bar{\theta}_0) \neq 0$, then $\hat{\gamma}$ is proportional to γ.